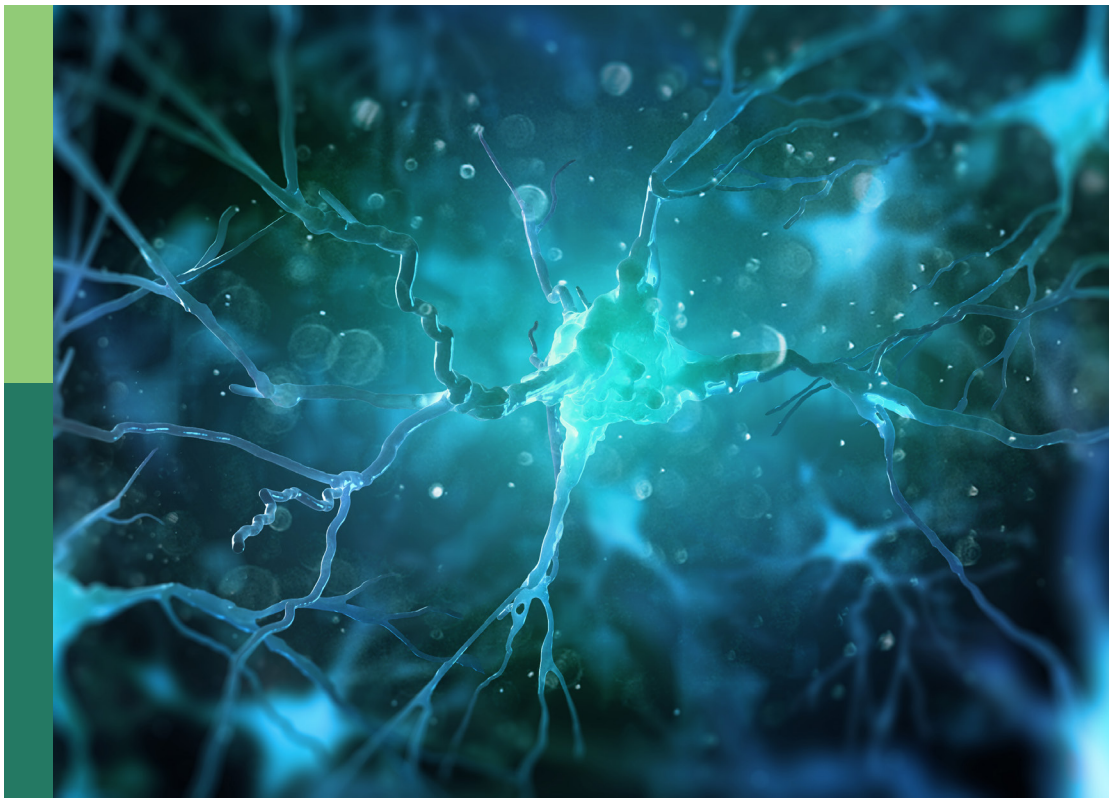
frontiers | Research Topics

New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention

Edited by
Joan E. Van Horn, Josanne van Dongen, Yvonne H. A. Bouman, Märta Wallinius and Patrice Renaud

Published in
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-5531-6
DOI 10.3389/978-2-8325-5531-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention

Topic editors
Joan E. Van Horn — De Forensische Zorgspecialisten, Netherlands
Josanne van Dongen — Erasmus University Rotterdam, Netherlands
Yvonne H. A. Bouman — Stichting Transfore, Netherlands
Märta Wallinius — Lund University, Sweden
Patrice Renaud — University of Quebec in Outaouais, Canada

Citation
Van Horn, J. E., van Dongen, J., Bouman, Y. H. A., Wallinius, M., Renaud, P., eds. (2024). *New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5531-6

Prof. Patrice Renaud is involved with BehaVR solutions, in a private business related to the use of new technologies in mental health.

Table of contents

VR QUANTITATIVE STUDIES

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 24 September 2024
DOI 10.3389/fpsyg.2024.1479498

OPEN ACCESS

EDITED AND REVIEWED BY
Stefano Ferracuti,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Joan E. van Horn
✉ JvanHorn@dewaagnederland.nl

RECEIVED 12 August 2024
ACCEPTED 09 September 2024
PUBLISHED 24 September 2024

CITATION
van Horn JE, Wallinius M, Bouman YHA, Renaud P and van Dongen JDM (2024) Editorial: New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention. *Front. Psychol.* 15:1479498. doi: 10.3389/fpsyg.2024.1479498

Editorial: New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention

Joan E. van Horn[1]*, Märta Wallinius[2,3,4], Yvonne H. A. Bouman[5], Patrice Renaud[6] and Josanne D. M. van Dongen[7]

[1]Research Department de Waag, Center for Outpatient Treatment, Utrecht, Netherlands, [2]Evidence-Based Forensic Psychiatry, Department of Clinical Sciences Lund, Psychiatry, Lund University, Lund, Sweden, [3]Center of Ethics, Law and Mental Health, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [4]Department of Research, Regional Forensic Psychiatric Clinic, Växjö, Sweden, [5]Department of Research, Transfore, Institute for Forensic Mental Health Treatment, Deventer, Netherlands, [6]Université du Québec en Outaouais and Institut National de Psychiatrie Légale Philippe-Pinel, Montréal, QC, Canada, [7]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, Netherlands

KEYWORDS

forensic, virtual reality, wearables, neuropsychology, biomarkers

Editorial on the Research Topic
New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention

General information

The forensic field has recently witnessed a growing interest in neuropsychology, wearables, and VR technology. These emerging areas promise to enhance the diagnosis and treatment of various forensic subpopulations. Our Research Topic, "*New directions in forensic psychology: applying neuropsychology, biomarkers and technology in assessment & intervention*," encompasses 15 papers detailing the latest advancements in these domains. Through these contributions, we aim to encourage further research.

Most papers focus on the application, viability, and impact of virtual reality (VR) across various forensic groups. While submissions on wearables and neuropsychology were less frequent within this Research Topic, the included studies are of significant value. Eighty-nine contributing authors hail from diverse global regions, including Italy, Canada, and Colombia, with a notable concentration from the Netherlands and Sweden. We value the international diversity of the submissions, which mirrors the widespread global interest in cutting-edge diagnostic and therapeutic options in forensic mental health care.

The Research Topic is organized into four thematic sections: neuropsychology and neurofeedback, wearables, qualitative VR studies, and quantitative VR studies. The following sections provide a concise overview of the respective papers included.

Neuropsychology and neurofeedback

Four papers cover forensic neuropsychology, neurofeedback, and a critical reflection of this use via the possible implementation of neurorights. The paper by Hutten et al. encompasses a Delphi consensus study on the neuropsychology of aggression. It elucidated the professionals' view on using neuropsychological tests to study different types of aggression and their diverse aspects, which are included in the RDoC framework. Balestrino et al. studied the usefulness of a handwriting test in assessing cognitive impairments. They found that a single score, the COGnitive Impairment Through handwriting (COGITAT) score, reliably assessed the writer's cognitive state. Because previous research has shown that psychological treatment, such as Cognitive Behavioral Therapy, is ineffective for all forensic clients, Hendriks et al. studied the feasibility and usability of neuromodulation training in a forensic outpatient clinic. They found that the training was rated sufficiently usable and feasible by patients and their therapists. Given the emergence of neurotechnological applications in forensic psychiatry and criminal law more broadly, such as the above-described neurofeedback treatment, there is a call for the implementation of so-called neurorights. Díaz Soto and Borbón evaluated the status of this matter and concluded that, although the interpretation of the current human rights should be made in such a way as to protect the dignity of the accused or client, new neurorights may offer reduced protection of human rights.

Wearable technology

Two notable studies investigated the role of wearable technology in forensic psychiatry. de Looff et al. conducted a randomized crossover trial to evaluate the usability and acceptance of four wearable devices among forensic psychiatric patients and staff. Their findings revealed that while fitness trackers like Fitbit and Garmin were more user-friendly, none of the devices met international usability standards, highlighting the need for improved gamification and motivational features. In a complementary study, ter Harmsel et al. assessed the Sense-IT bio cueing app's effectiveness when added to Aggression Regulation Therapy (ART) for forensic outpatients. Although the app increased interoceptive awareness in most participants, its impact on aggression and emotion regulation was inconsistent.

These studies contribute valuable insights into the forensic field, emphasizing the potential of wearable technology to enhance therapeutic outcomes. Key lessons include the critical importance of usability, personalized interventions, and seamless integration into therapy for successful adoption. Both studies demonstrate solid methodologies, though limitations such as small sample sizes and the need for better algorithm validation highlight areas for future improvement. Overall, these studies underscore the necessity of tailoring technology-based interventions to individual needs for effective forensic psychiatric treatment.

Qualitative studies on VR interventions

The implementation of Virtual Reality in forensic psychiatric treatment is relatively new. Several authors have applied qualitative methods to explore the experience of patients and clinicians with VR applications such as DEEP, Virtual Reality Aggression Prevention Training (VRAPT), or Triggers and Helpers. Two out of three studies had been conducted in the Netherlands (Klein Haneveld et al.; Kouijzer et al.) and one in Sweden (González Moraga et al.).

Whereas VRAPT and Triggers and Helpers are blended applications in which roleplaying assists patients in improving awareness and social skills, e.g., reduction of aggressive behavior, DEEP is an application in which a patient practices deep breathing in a gamified biofeedback underwater world.

These three studies used interviews to explore the participants' experiences and seek answers to the questions of for whom and when these applications are helpful in forensic psychiatric treatment. In the DEEP study, the authors sought answers to which application method would suit whom best. Apart from suggestions on improving immersiveness, ideas on implementing DEEP in clinical practice emerged. The study by Kouijzer et al. focused on implementation and used Triggers and Helpers as a showcase. Patient characteristics must be considered when deciding to whom this method should be offered, and continuously assisting clinicians when they use VR seems a vital necessity. The Swedish qualitative evaluation of patients' experience with VRAPT also highlighted the need to thoroughly implement innovative treatments such as VR and personalize treatment goals for which VR can be used.

Quantitative studies on VR interventions

There were six quantitative studies, most from Europe (the Netherlands, Sweden) and one from Canada. All studies except one tested different VR interventions in clinical forensic settings, while one investigated a chatbot developed for risk assessment training. Given the state of the field, the quantitative studies overall had a feasibility and effect approach and described interventions with the need for continued development and evaluation. Common findings were that all users' attitudes toward technology-driven interventions were generally positive and that outcomes depended on successful implementation in interventional settings.

Several studies focused on the treatment of aggression regulation for either forensic psychiatric patients or imprisoned offenders, evaluating either a method specifically designed for VR-assisted treatment only (Virtual Reality Aggression Prevention Training) or comparing outcomes from a treatment (Responsive Aggression Regulation Therapy) being delivered either in virtual environments or in real life settings. The results of the two studies presenting longitudinally followed outcomes over time were promising, with decreased levels of anger, aggression, and emotion regulation maintained over follow-up. Two studies focused on assessment, both with an experimental design, where one investigated the feasibility of paranoia assessment in virtual environments, and the other determined acceptance and trust of students on chatbot-assisted risk assessment training. For both studies, not only the technology but also the user characteristics seemed necessary for the usefulness of the assessment. So far, we cannot replace standard training and assessment with technology-driven versions. Finally, a study used a VR-based intervention to prepare forensic psychiatric patients for authorized leave, and the

potential of VR to increase patient motivation and reduce stress was evident.

In summary, many of the quantitative studies on this Research Topic benefited from the characteristics of VR, which facilitates exposure to specific environments in preparation for real-life occurrences. However, it is evident that much is left to investigate and that further developments, especially concerning individual tailoring, are needed. The current findings can guide clinicians and researchers in forensic settings in their coming ventures on these matters.

General conclusion

The collective insights from the 15 papers featured in our Research Topic highlight the transformative potential of integrating neuropsychology, biomarkers, and advanced technology into forensic psychology. Through diverse methodologies, these studies illustrate how neuropsychological assessments, wearable devices, and virtual reality (VR) interventions can enhance diagnosis, treatment, and overall therapeutic outcomes in forensic populations.

Key lessons from these studies include the importance of feasibility and usability in implementing new technologies. For instance, neuromodulation and neurofeedback have shown initial promise in addressing impulse control issues, but practical challenges must be addressed to ensure broader application. Similarly, using wearables, while beneficial in some cases, reveals that user engagement and device adaptability are crucial for sustained success. The adoption of VR in forensic settings stands out for its ability to simulate realistic scenarios, providing a safe environment for behavior assessment and skills training. Studies on VR-assisted aggression treatment, for example, indicate positive patient experiences and potential for improving therapeutic outcomes. However, these interventions require a careful introduction to avoid exacerbating psychological distress, especially in vulnerable populations.

Integrating these technologies necessitates a multifaceted approach, combining rigorous scientific validation with practical considerations. Future research should focus on refining these tools, tailoring interventions to individual needs, and ensuring ethical standards are met, particularly concerning mental privacy and dignity. The findings presented serve as a valuable guide for future research and clinical practices, emphasizing the need for tailored, technology-driven interventions to serve forensic populations better.

Author contributions

JvH: Writing – review & editing, Writing – original draft. MW: Writing – review & editing, Writing – original draft. YB: Writing – review & editing, Writing – original draft. PR: Writing – review & editing, Writing – original draft. JvD: Writing – review & editing, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 04 June 2024
DOI 10.3389/fpsyg.2024.1354997

OPEN ACCESS

EDITED BY
Shao-Wei Xue,
Affiliated Hospital of Hangzhou Normal University, China

REVIEWED BY
Anthony Beech,
University of Birmingham, United Kingdom
Alexander Nikolaevich Savostyanov,
State Scientific Research Institute of Physiology and Basic Medicine, Russia

*CORRESPONDENCE
J. Wilpert
jwilpert@dfzs.nl

RECEIVED 14 December 2023
ACCEPTED 13 May 2024
PUBLISHED 04 June 2024

CITATION
van der Schoot A, Wilpert J and van Horn JE (2024) Neurofeedback and meditation technology in outpatient offender treatment: a feasibility and usability pilot study.
Front. Psychol. 15:1354997.
doi: 10.3389/fpsyg.2024.1354997

Neurofeedback and meditation technology in outpatient offender treatment: a feasibility and usability pilot study

A. van der Schoot, J. Wilpert* and J. E. van Horn

Research Departement, De Forensische Zorgspecialisten (DFZS), Utrecht, Netherlands

Introduction: Although Cognitive Behavioral Therapy (CBT) is the most often used intervention in forensic treatment, its effectivity is not consistently supported. Interventions incorporating knowledge from neuroscience could provide for more successful intervention methods.

Methods: The current pilot study set out to assess the feasibility and usability of the study protocol of a 4-week neuromeditation training in adult forensic outpatients with impulse control problems. The neuromeditation training, which prompts awareness and control over brain states of restlessness with EEG neurofeedback, was offered in addition to treatment as usual (predominantly CBT).

Results: Eight patients completed the neuromeditation training under guidance of their therapists. Despite some emerging obstacles, overall, the training was rated sufficiently usable and feasible by patients and their therapists.

Discussion: The provided suggestions for improvement can be used to implement the intervention in treatment and set up future trials to study the effectiveness of neuromeditation in offender treatment.

KEYWORDS

feasibility, usability, neurofeedback, meditation, forensic outpatients, impulse control problems

Introduction

Cognitive Behavioral Therapy (CBT) is viewed as one of the most effective psychological interventions to reduce re-offending (recidivism) and is widely implemented as evidence based forensic treatment in various forensic settings (Pearson et al., 2002; Landenberger and Lipsey, 2005; Wilson et al., 2005; Lipsey et al., 2007; Henwood et al., 2015). However, studies show that not all offenders benefit from CBT to the same extent (Babcock et al., 2004; Feder and Wilson, 2005; Eckhardt et al., 2013; Beaudry et al., 2021). It is known that certain offender characteristics, such as (comorbid) psychiatric disorders, can interfere with the success of treatment (Babcock et al., 2004; Eckhardt et al., 2013). Brazil et al. (2018) provide an explanation: effects of existing methods are mostly measured by self-reported behavioral outcomes such as aggression without an operationalization of the specific underlying constructs that contribute to offending behavior, which they consider detrimental for the effectiveness of offender treatment. They propose that biological (e.g., genetics, brain, and physiology) and cognitive functioning measures, and clinical observations would provide more insight into effectiveness of treatment programs in reducing recidivism rates. Also, incorporating biopsychosocial components should improve treatment effectivity by tailoring

intervention techniques to specific perpetrator characteristics (Brazil et al., 2018; Beaudry et al., 2021).

CBT taps into cognitive and intellectual aspects and functioning, which are potentially less easily accessible to offenders with impulse control problems in particular. Impulse control problems impede the ability to foresee consequences, make achievable plans, choose from alternatives, control impulses, inhibit unwanted thoughts, and regulate social behavior (Heatherton and Wagner, 2011). Hence, impulse control problems are strongly related to the risk of general offending behavior (Moffitt et al., 2011; Loeber et al., 2012; Fergusson et al., 2013) and recidivism rates (Lloyd et al., 2014). As neurotechnology tunes into bodily mechanisms and experiential learning, as opposed to the cognitive methods which set out to finding explanations and alter thinking, it can be a valuable addition to offender treatment. Different neurotechnological methods have been developed to aid better self-regulation abilities, however, which of these methods are most suitable and effective to achieve effective treatment effects in offenders is yet unclear (Bijlsma et al., 2022).

As mentioned, in the search for treatment methods that are more effective, biopsychological factors should be taken into account (Brazil et al., 2018). Aggression can be the result of increased left frontal cortical activity (activity regarding approach) and decreased right frontal cortical activity (activity regarding inhibition) (Hortensius et al., 2012). It is therefore important to conduct research on interventions that work with left and right hemisphere asymmetry in aggression. Neurorehabilitation technology is an umbrella term for various technological applications and methods addressing specific brain functioning networks or pathways that are related to specific behaviors or symptoms. Some of these applications could yield promising prospects for offender treatment. For example, research suggests that both transcranial direct current stimulation (tDCS) and continuous theta burst stimulation (cTBS) are methods of neurorehabilitation that can play a role in the modulation of aggressive behavior by directly changing brain activity (Knehans et al., 2022). In a laboratory aggression task and questionnaire, Sergiou et al. (2022) demonstrated that HD-tDCS enhanced the frontal brain regions connectivity in a group of offenders, resulting in a decrease of aggressive responses. Subsequently, this could represent an innovative approach suitable for implementation in forensic outpatient treatment. However, HD-tDCS is a relatively expensive method which can only be administered by trained professionals. This renders it challenging for forensic outpatient clinics to offer this type of treatment. A cheaper and easier administrable method of neurorehabilitation, which has also been studied in offender populations, is EEG neurofeedback (Bijlsma et al., 2022).

Neurofeedback, also known as EEG biofeedback, is a technically supported form of real-time feedback of an individual's brain activity through a brainwave monitoring device. In neurofeedback training, users learn to manipulate their neural activity based on direct feedback from the device. Since neurofeedback has been successfully used in treatment of impulse control problems in non-offenders (Sokhadze et al., 2008; van Doren et al., 2019; Hong and Park, 2022; Lima et al., 2022; Moreno-García et al., 2022) and is thought to target neural and cognitive processes that underlie offending behaviors (Bijlsma et al., 2022), it could also be a meaningful neurorehabilitation method in offender treatment (Van Outsem, 2011).

Although neurofeedback research in offender populations is scarce, initial results exhibit promise. Larson (2019) studied a small group of domestic violence perpetrators ($N = 10$) which received an intensive neurofeedback training. The treatment group (neurofeedback training) showed significantly lower Beta wave frequencies (e.g., active, alert, and focused mental states) than the control (no neurofeedback training) group. However, no significant differences were found between pre- and post-tests of participants' self-reported feelings of anger, stress, and aggression. Furthermore, in a single case study on an adult with a history of sexual offending by Borghino et al. (2022), neurofeedback had a positive effect on control of sexual feelings, urges and behaviors. In yet another study on neurofeedback and recidivism, 20% of the treated incarcerated offenders (convicted of arson, sexual or violent offenses) had been rearrested, as opposed to 65% of the matched incarcerated offenders who did not receive neurofeedback (Von Hilsheimer and Quirk, 2006). More research is needed to fully understand how neurofeedback can contribute to offender treatment (Fielenbach et al., 2018).

Findings indicating a link between meditation practice and changes in brain regions and networks associated with impulsivity problems (Hölzel et al., 2011; Dambacher et al., 2015; Chaibi et al., 2023), suggest that neurofeedback combined with meditation, neuromeditation, could demonstrate an even larger positive effect on self-regulative behaviors, such as: attention regulation, body awareness, emotion regulation and change in perspective on the self (Hölzel et al., 2011; Sedlmeier et al., 2018). Since learning to enter calm states can be very challenging without feedback, insights into brain activity can facilitate to "get it right" in a more targeted manner. Through integration of real-time monitoring of brain waves and meditation practices, individuals can acquire the ability to swiftly enter a targeted state of brain relaxation and sustain this state over prolonged time (Tarrant, 2020). A benefit of neuromeditation in contrast to CBT, is that a patient can learn to (re)gain bodily and mind control in a top-down (internal self-direction) manner instead of bottom-up (self-direction through externally offered strategies).

An example of a neuromediation appliance is the Muse™ brain sensing wearable device. Via Bluetooth, the Muse EEG headband is connected to the Muse meditation app. It registers and recognizes Beta Waves (active, alert, and focused mental states) and Alpha Waves (relaxed and calm mental states) of the wearer and promotes Alpha states by providing auditory feedback (Muse, 2023).[1] The presence of high Beta wave brain frequencies could hinder the ability to self-regulate emotions, as asynchronicity in frontal frequencies is related to aggression in offenders (Hortensius et al., 2012; Sergiou et al., 2022). Therefore, promoting Alpha states through neuromeditation with Muse could be a promising neurorehabilitation method supporting self-regulation in offenders.

Research showed that neuromeditation with Muse increased a state of mindfulness (ability to focus attention on the here and now, to feel less stress/tension) in adult participants in a non-clinical setting, represented by less mind restlessness and accurate attention to the breath. Participants reported the neuromeditation method to be an effective addition to their meditation practice (Hunkin et al., 2021a). Mindfulness has been shown to reduce impulse control problems (Gallo et al., 2021). Also, applied in general health care, Muse has demonstrated improvement of focused attention, reduction of physical symptoms, and supporting accelerated mindfulness learning. As a result, stress levels reduced and cognitive performance

1 https://choosemuse.com

(such as faster reaction time and increased inhibition) improved (Bhayee et al., 2016; Taj-Eldin et al., 2018; Crivelli et al., 2019). Therefore, Muse EEG could be a beneficial technology in offender treatment. At the present, no documentation was found of prior studies involving the utilization of Muse neuromeditation technology within forensic settings.

Central aim of the study

In preparation of a Randomized Control Trial (RCT) study, a pilot study was conducted to investigate the feasibility and usability of the Muse neuromeditation technology, in adult forensic outpatient treatment. It was expected that neuromeditation can be valued as a feasible and usable addition to treatment as usual.

Method

The study was conducted in a Dutch forensic outpatient treatment facility between September 2022 and March 2023. An extensive test battery was applied using a multi-method design (self-report instruments, clinician-rated instruments, interrater-agreement, neuropsychological test, neuromeditation) at pre- and post-test (5 weeks after the pre-test) with weekly neuromeditation measurements. The research was approved by the Internal Review Board of the Van der Hoeven Clinic, indicating that it complies with the ethical guidelines of the institution and all laws and regulations in the Netherlands and Europe (2021-2-SC).

Setting

Two locations of a Dutch forensic outpatient treatment facility were involved in the study. At these facilities, outpatients from the age of 12 receive treatment aimed at transgressive inclinations or behavior on a voluntary or mandatory base. Voluntary indicates that a patient enters treatment on their own initiative, on referral of a general practitioner or another mental health care professional. Mandatory treatment is imposed by a judge. Excluded for treatment are patients who are in acute psychiatric crisis, for example psychosis, severe addiction problems or suicidal tendencies. They are referred to the appropriate specialized mental health care.

Treatment for outpatients consists of a combination of various CBT elements, such as psychoeducation, self-monitoring, cognitive restructuring, improvement of coping skills or other evidence-based intervention techniques, such as Eye Movement Desensitization and Reprocessing (EMDR) and Acceptance and Commitment Therapy (ACT).

Sample

Patients were eligible for participation if they were 18+ years of age and had at a score of 2 or higher on the dynamic risk factor "lack of impulse control" of the Forensic Outpatient Risk assessment and Evaluation (FORE V2; Van Horn et al., 2020; for more information, see Instruments section). A total of six female therapists, with a mean age of 28.83 years ($SD = 3.8$), registered to participate in the study and selected patients from their caseload who fitted the inclusion criteria.

Of the 11 patients enrolled in the Muse pilot study, eight completed the study: six males and two females with a mean age of 40.88 years ($SD = 12.28$, range 25–59 years). The three dropouts did not start with the neuromeditation sessions, because of other priorities such as treatment start-up, EMDR intervention and crises. The patients' mean impulse control score on the FORE dynamic risk factor was 2.75 ($SD = 0.71$), ranging from 2 to 4. At the outset of their involvement in the study, participants had an average treatment duration of 8.38 months ($SD = 5.98$), with a range from 2 to 19 months. Table 1 presents several additional characteristics of the patients' sample.

4-weeks neuromeditation training

At the start of each face-to-face session, patients received neuromeditation training using the Muse EEG headband (https://choosemuse.com; RRID:SCR_014418). In their study with Muse neuromeditation, Crivelli et al. (2019) concluded that a daily exercise of meditation (10–20 min) over the course of 4 weeks, resulted in positive effects. Following this, the study duration was set at 4 weeks. Therapists incorporated the neuromeditation training in the patient's regular weekly treatment session, assisted by a protocol with a description of all the necessities and sequence of steps per session. Also, instruction manuals for both therapists and patients were provided with guidelines on how to use the Muse headband and mobile application (including the meditation exercises). More information on the Muse headband and mobile application is provided in the Instruments section.

TABLE 1 Characteristics of the patients' sample ($N = 8$).

	n(%)
Education	
Secondary vocational education	4(50)
Pre-secondary vocational education	2(25)
Primary education	2(25)
Primary diagnosis	
Disruptive, impulse-control and conduct disorders NOS	4(50)
Personality disorder	2(25)
Autism Spectrum Disorder	2(25)
Substance abuse	
Cannabis	3(38)
Cocaine/speed	2(25)
Binge drinking	1(13)
Reason to enter treatment	
Verbal and physical aggression	7(88)
Compulsive stealing	1(13)
Treatment context	
Voluntary	7(88)
Mandatory	1(13)

Per session, 15 min were scheduled for the intervention, including pre- and post-inquiries and a brief evaluation. Each consecutive session, the headband wearing time was increased by 1 min, starting with 3 min in the first session. After the Muse intervention, the session was continued with treatment as usual. During the 4-weeks research period, patients were also asked to perform a daily mindfulness exercise at home.

Instruments

The dynamic risk factor *D8 Impaired impulse control* of the risk assessment instrument *Forensic Outpatient Risk assessment and Evaluation* (FORE V2; Van Horn et al., 2020) was used as an inclusion criterion. This item measures out-of-control behavior or poor control over emotions, leading to disruptions in the past 6 months in daily life functioning at home, at work, during education, socially or financially. Patients were included if they scored 2 or higher ("some problems") on a 5-point scale, with 0 "no impulse control problems" and 4 "severe impulse control problems in three or more life domains". The risk assessment is conducted by therapists. From 2017, all forensic outpatient services are obligated to assess the recidivism risk of every person in mandatory forensic outpatient treatment. In 2019, the FORE was advised to use for treatment outcome monitoring and the prediction of risk in outpatient forensic care (Zicht op forensische zorg [View on forensic care], 2023). A study of the validity and reliability of the FORE V2 has yielded good results (Eisenberg et al., 2020).

Muse headband and mobile application

The Muse headband (S) and mobile application (Muse, 2023) were used for neuromediation (see text footnote 1). The headband measures patterns of frequencies in brain waves, which represent brain status (i.e., brain activity) using electroencephalography (EEG).

Muse translates these brain waves into real-time audio feedback on the Muse app. Feedback consists of three different weather sounds as a proxy of the user's brain state. Users can choose between different soundscapes, such as rainforest (default) or beach. In the rainforest soundscape for example, sounds of storm and hard rain represent wandering thoughts, indicating that an individual is distracted, and attention is fluctuating (active mental state). Sounds of wind and soft rain, on the other hand, signal that the brain is in its natural state of rest. That is, the attention is not fluctuating, but there is no deep focusing either (neutral mental state). Lastly, dripping water and bird chirping convey a deep soothing focus (calm mental state). After a session, the Muse app displays the number of bird chirps "achieved" during the session and the percentage of calm mental state, as well as the number of "recoveries", which mark the ability to reinstate from an active to a calmer mind. A study by Hunkin et al. (2021b) provided initial evidence for the internal consistency and validity of two Muse metrics (mean calm states and recoveries from active states) as indicators of state mindfulness.

Using the auditory feedback, the receiver learns what state the brain is in and how a calm brain state "sounds" and feels like. Mindfulness and meditation are used to control brain turmoil, for instance, by doing relaxation exercises or focusing attention (for example on breathing). The Muse app offers various guided and non-guided meditation exercises that can be used with or without the headband, such as a non-guided breathing exercise or guided relaxation exercises.

Feasibility and usability

Feasibility was defined as the degree to which the study protocol can be performed in a practical way in terms of study procedure and use of the Muse headband and app. Usability was defined as the degree to which a product is experienced as efficient and satisfactory by designated users to accomplish predetermined objectives within a defined usage context.

The evaluation forms (completed at post-test) to assess the feasibility consisted of several questions for therapists and patients, covering the feasibility topics: study protocol (e.g., comprehensible information, pre- and post-test experience), Muse headband (e.g., wearing comfort, easy to use), and Muse app (e.g., suitable exercises).

The usability was assessed using qualitative and quantitative information. Qualitative information consisted of feedback from therapists and patients on, for instance, perceived changes in relaxation skills and reduction of impulsivity. Some of the questions in the evaluation form were open-ended (e.g., about the pros and cons of the training), but most were dichotomous (yes/no) with a text field to optionally elaborate on the scoring. Furthermore, at the end of each neuromeditation session, patients were asked about their experiences with the Muse headband and the app.

Quantitative information was gathered per session and pre- and post-research period. The *Modified Overt Aggression Scale* (MOAS; Buitelaar et al., 2014) and a rating scale for bodily tension were administered in each neuromeditation session. The MOAS was used to assess patients' level of occurred behavioral aggression in the past week. The MOAS is a check-off list to register incidents of aggression over the past week. The practitioner registers if any of the following aggression types occurred: verbal, physical, aggression against property and auto-aggression on a 5-point scale (0 "none" to 4 "frequent"). For example, for verbal aggression the scores represent 0 "no verbal aggression", 1 "shouting angrily", 2 "cursing viciously", 3 "impulsively threatens violence toward others or self" or 4 "threatening violence toward others or self repeatedly or deliberately". Subsequently, a higher score reflects a higher prevalence of aggression. The psychometric properties of the MOAS have been supported in earlier studies (Kay et al., 1988). Furthermore, therapists asked patients about the currently experienced *tension* in their body, mind, and breath directly before and after each headband usage, ranging from 1 "relaxed" to 10 "highly tense".

Pre- and post-test changes in impulsivity, body tension, awareness, executive functioning, and aggression were measured with self-report instruments. Impulse control was measured with the *Barratt Impulsiveness Scale version 11* (BIS-11; Patton et al., 1995). The BIS-11 is a 30-item self-report questionnaire that measures impulsivity. Responses are given on a 4-point Likert scale (1 "rarely/never" to 4 "almost always"). High scores indicate a higher degree of impulsiveness. Since the original factor structure could not be confirmed in an adult forensic population (Ireland and Archer, 2008), we used the total BIS-11 score in this study. According to a review by Vasconcelos et al. (2012), the internal consistency (Cronbach's $\alpha = 0.69$ to 0.83) and test–retest reliability (correlation coefficient $r = 0.66$ to 0.83) of the BIS-11 were satisfactory in most studies.

Body awareness was measured with the *Anger Bodily Sensations Questionnaire* (ABSQ; Dutch: Zwets et al., 2014) and the *Dutch Scale of Body Connection* (SBC; Van der Maas, 2015). The ABSQ is a self-report questionnaire with 18 items about specific bodily sensations

encountered when experiencing anger as a result of (perceived) provocation. Responses are given on a 5-point Likert scale (1 "not at all" to 5 "very much"). A higher score on the ABSQ indicates experiencing a higher amount of body signals when angry. In a study of Dutch offenders, the ABSQ showed good internal consistency (Cronbach's $\alpha = 0.93$) and test–retest reliability ($r = 0.82$; Zwets et al., 2014). The SBC is a 20-item self-report measure, designed to assess the subscales body awareness (12 items) and bodily dissociation (8 items). Items are based on common expressions of awareness of the body such as 'I notice that my breath becomes superficial when I'm nervous'. Responses are given on a 5-point Likert scale (1 "not at all" to 5 "always"). Higher scores on the subscales represent more awareness/dissociation. The internal consistency of the subscales body awareness and bodily dissociation has been shown to be adequate (Cronbach's $\alpha = 0.83$ and 0.78 respectively; Price and Thompson, 2007).

To assess executive functioning (EF), the Dutch version of the *Parametric Go/No-Go Task* (PGNG; Langenecker et al., 2007; Dutch version: Van Horn et al., 2023) was used. The PGNG is a computerized task designed to measure cognitive flexibility (set shifting), response inhibition, and working memory. Validity studies demonstrate that the PGNG measures the core EFs in a psychometrically sound, brief, and ecologically valid way (Langenecker et al., 2007; Votruba and Langenecker, 2013). The task comprises three levels with increasing difficulty assessing attention, set shifting, and inhibitory control in levels 2 and 3. In all three levels, a series of letters is presented (x, y and z) at a fairly rapid rate. The aim is to follow certain rules as instructed, while responding to specific letters as quickly as possible by pressing the space bar. Outcome measures in the three levels are the percentage of correct target trials (PCTT, indicative of attention), and the percentage of correct inhibition trials (PCIT, indicative of inhibitory control, not assessed in level 1).

The short *Dutch version of the Aggression Questionnaire* (AVL-AV; Buss and Perry, 1992) was used to measure aggression. The 12-item AVL-AV is a self-report questionnaire and measures physical aggression, verbal aggression, anger, and hostility (with three items each). Items were scored on a 5-point Likert scale (1 "completely disagree" to 5 "completely agree"). Higher scores indicate more aggression, anger, and hostility. The AVL-AV shows good psychometric properties in aggressive offenders (Hornsveld et al., 2009).

The *System Usability Scale* (SUS; Brooke, 1996) was administered at post-test to quantify the overall usability of the Muse headband and app. The SUS consists of 10 items with scoring options on a 5-point Likert scale (1 "completely disagree" to 5 "completely agree"). The total score of the SUS can range from 0 to 100, with higher scores indicating better usability. Based on study mean quartiles, Bangor et al. (2008) consider usability scores from 52.01 as acceptable and from 72.75 as good. They found a good internal consistency of the SUS-items of Cronbach's $\alpha = 0.85$ and higher.

Procedure

Therapists received documents from the research team containing information about the study content and procedures. The six participating therapists reached out to their patients (regardless of their treatment phase) when they met the inclusion criteria. Patients received an information leaflet and, after agreeing to participate, signed the informed consent form and filled out pre-test measures online, including several questions about their prior experience with relaxation exercises and motivation to participate in the study. This pre-test battery was accessed by an email link to Qualtrics (Version: March 2023),[2] a secure online survey platform. The PGNG was administered under the guidance of the therapist at the outpatient facility. After the research period, post-tests were conducted with the same instruments following the same procedures. In addition, at post-test, both patients and therapists filled out an evaluation questionnaire via Qualtrics to assess the feasibility and usability of the study protocol and neuromeditation training.

Strategy of analysis

Information from the pre- and post-tests (quantitative) and evaluation forms (partly qualitative) was analyzed using IBM SPSS version 27. Since this pilot study is not primarily aimed at quantifying effects, and the sample size is insufficient for generating statistically reliable insights into preliminary findings as well, any assertions or conclusions drawn from the data must be approached with caution. Instead, the presentation of frequencies and percentages, as well as averages and standard deviations, is undertaken solely for the purpose of providing a descriptive overview and to offer an insight into how these results contributed to the perception of the usability of the intervention. Regarding the qualitative data from the open-ended questions and possibility to elaborate on chosen answer categories, the following procedure was followed. Firstly, the first and second author independently rated the information in two main categories: 'negative evaluation' or 'positive evaluation'. Statements from patients and therapists were categorized as negative if there were (to some degree) indications of problems and cons (e.g., too time consuming, unclear, difficult, without added value, etc.). Statements from patients and therapists were categorized as positive if there was (some degree of) evidence to the contrary, implicating no problems and pros. Secondly, Cohen's kappa values were calculated for the feasibility (10 items) and usability (3 items) separately. Interpretation guidelines for Kappa values are as follows: $\kappa < 0$ = no agreement, 0.0–0.20 = slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement and 0.81–1.0 = perfect agreement (Landis and Koch, 1977). Agreement on the feasibility items was moderate ($\kappa = 0.55$) and on the usability items fair ($\kappa = 0.25$). Following that, an agreement score was achieved per item, and these consensus scores are presented as the ultimate findings. Feasibility and usability were considered acceptable when at least 60 percent of the patients and therapists rated the items as positive, and good when 80 percent rated them as positive.

Results

Prior experience and expectations

Table 2 describes, among other things, the patients' prior experience with relaxation exercise and their motivation and

2 https://www.qualtrics.com

TABLE 2 Pre-test survey questions patients (*N* = 8).

Question	*N*(%)
Prior experience with relaxation exercise	5(63)
Current practice of relaxation exercise	4(50)
Motivation participation study	
Learning relaxation exercises	3(38)
Improving capability to relax	7(88)
Body awareness could be an important addition to treatment	7(88)
Using the Muse headband	1(13)
Other	2(25)
Willing to perform daily relaxation exercise at home	8(100)
No difficulties with the use of English in Muse app	7(88)
Intention to plan an extra weekly session for Muse	5(63)

TABLE 3 Feasibility and usability rating based on therapists' experiences per patient (*N* = 8).

Topic	*n*(%)
Feasibility	
Study protocol	
Patients experience filling out questionnaires*	
No problems	5(63)
Some problems	3(38)
Patients experience performing PGNG*	
No problems	3(38)
Some problems	5(63)
Enough time (15 min) for Muse session	4(50)
Enough information in study protocol	7(88)
Muse app and headband	
Easy to use Muse app and headband	5(63)
Usability	
Changes and added value of neuromeditation	
Added value of Muse headband and app	7(88)
Observed changes in patient's relaxation skills and/or impulsivity	7(88)
Inclined to use neuromeditation with other patients	8(100)

*Coded qualitative information by first and second author.

expectations entering the study. The two other motives to participate, as mentioned in Table 2, were to reduce stress levels and improve calmness. Patients expected their participation to result in increasing capability to relax and experience calmness, more bodily awareness and control and reduction of impulsive behavior.

Feasibility and usability as assessed by therapists

Information on the feasibility and usability of the study protocol and the neuromeditation as experienced by the therapists, is presented in Table 3.

Table 3 shows that usability items were rated as good (>80%) by therapists and most of the feasibility items were acceptable (>60%), except for PGNG and session time which were rated below the acceptable threshold. The evaluation form responses provided by the therapists revealed several noteworthy findings. Among these, were the amount of questionnaires, and one patient's challenges in interpreting certain items too literally. According to the therapists, patients experienced some frustration engaging in the PGNG task due to heightened difficulty levels. Furthermore, approximately half of the patients required more time (ranging from 20 to 25 min) during sessions to elaborate on their experiences with at-home meditation and the usage of the headband. Also, some patients experienced technical failure of the app, when pausing during an exercise.

Therapists were asked if patients indicated pros and cons of the Muse intervention. Therapists indicated that patients' comments on feasibility were more often negative than positive (1 pro versus 10 cons). The drawbacks primarily centered around issues such as headband discomfort, lack of applicability of exercises, frustration due to not being able to change the weather, distraction by a foreign language (English to non-native speakers) during stress, much effort to perform exercises at home and uncertainty of the exercise's impact as a result of the absence of feedback during home sessions. As for the usability, comments were more often positive than negative (11 pros versus 0 cons). The essence of the pros was that Muse provided patients with a straightforward method of acquiring relaxation skills and attuning to one's bodily sensations.

Therapists reported 9 pros versus 10 cons regarding feasibility and 2 pros versus 0 cons concerning usability. As feasibility drawbacks, therapists mentioned several points of concern: equipment and procedure take time to get acquainted to, interference of the use of Muse when patients might prefer to discuss pressing matters first, complexity and inapplicability of specific Muse app exercises, additional time needed for session preparation (particularly during the pre-treatment phase which is frequently filled with diagnostic and working relationship-establishing activities).

In terms of feasibility advantages, therapists noted that Muse offers a structured approach with clear objectives, facilitates accessible feedback and learning, and appears to be more comprehensible for patients compared to merely receiving explanations. Specific relaxation moments within the treatment session also appeared to be beneficial.

As for usability, therapists observed that their patients exhibited increased calmness, reduced stress and impulsiveness, a shift towards "think first and act later," and improved relaxation abilities. All therapists were inclined to use the Muse intervention again in the future with other patients, especially with problems such as stress, ADHD, trauma and sleeping difficulties.

Feasibility and usability as assessed by patients

Table 4 summarizes the raters' consensus of the evaluation form responses of the patients.

As can be seen in Table 4, most feasibility topics were rated good (>80%). The only aspect patients considered unfeasible, was the daily meditation exercise at home. All usability items were rated acceptable (>60%).

Patients reported four pros and three cons regarding feasibility topics and one pro and zero cons on usability. Patients experiencing difficulties with the questionnaires, found them confronting and containing some hard questions. Also, two patients struggled with the PGNG. One of them retook the test (with better results) because of impeding fatigue during the first administration. The three patients

TABLE 4 Feasibility and usability rating based on patients' experiences ($N = 8$).

Topic	*n*(%)
Feasibility	
Study protocol	
Comprehensible information	7(88)
Experience filling out questionnaires*	
No problems	6(75)
Some problems	2(25)
Experience performing PGNG*	
No problems	6(75)
Some problems	2(25)
Prior expectations met regarding participation in study	5(63)
Muse app and headband	
Clear explanation from therapists on how to handle Muse app	8(100)
At ease during headband session	6(75)
Headband comfortable	7(88)
Managed to do the daily exercise	2(25)
Suitable exercises in Muse app	7(88)
Usability	
Changes and added value of neuromeditation with Muse	
Muse contributed to earlier experience with mindfulness ($n = 5$)	3(60)
Muse contributed to changes in impulsivity	5(63)
Intention to use Muse/relaxation exercises in the future*	5(63)
	M(SD)
SUS score	70.94 (17.88)

*Coded qualitative information by first and second author.

10.3389/fpsyg.2024.1354997

TABLE 5 Within session assessment of prior aggression, current level of tension and neuromeditation results ($N = 8$).

	T1	T2	T3	T4
	M(SD)	M(SD)	M(SD)	M(SD)
Aggression (MOAS)				
Auto-aggression	0.25(0.46)	0.25(0.71)	0.88(1.25)	0.13(0.35)
Aggression objects	0	0.13(0.35)	0.50(0.76)	0.13(0.35)
Verbal others	0.88(0.84)	1(1.20)	0.75(0.89)	1.13(1.46)
Physical others	0	0	0	0
Tension				
General tension				
Before	4(1.51)	4(2)	5.25(2.38)	4.25(1.83)
After	3.13(1.13)	2.88(1.46)	3.88(2.30)	3.13(1.81)
Breath				
Before	2.38(1.30)	3.13(1.73)	3.88(2.59)	3.25(2.05)
After	2(0.93)	2.38(1.06)	2.75(1.98)	2.38(1.06)
Body				
Before	4.63(1.60)	4.88(1.89)	5.50(2.27)	4(1.93)
After	3(1.07)	3.50(1.31)	3.75(1.98)	2.75(0.89)
Thoughts				
Before	4.63(3.07)	4.13(2.23)	4.75(3.01)	4.63(1.69)
After	2.88(2.48)	3.38(2.67)	4.13(2.59)	3.25(1.98)
Neuromeditation				
Birds	5.50(4.75)	8.63(8.75)	14.38(15.34)	21.13(22.63)
Recoveries	2.38(2.45)	4.25(3.73)	7.13(8.90)	5(6.91)
Calm state	31.63(12.19)	30.75(19.86)	31.38(25.48)	35.25(28.29)

indicating their expectations of participation were not met, emphasized that their experience exceeded their expectations in a positive way. Patients varied in their preferred soundscapes, some considered the rainforest sounds irritating and switched to another option. Performing a daily exercise at home proved to be a challenge for most patients because of a variety of reasons: difficulty finding a applicable exercise in the Muse app, uncertainty due to lacking Muse headband feedback, distraction from the environment and lack of daily structure to incorporate a regular time to exercise. In case patients found the Muse app exercises inapplicable, they resorted to YouTube for alternatives. Nevertheless, seven patients indicated that there were enough applicable exercises in the Muse app. The average rating (a report mark from 1 for very bad to 10 for very good) for the contribution of the home exercise in promoting relaxation was 4.25 (SD = 2.43, range 1–7). The report mark for the Muse headband session was 7 (SD = 1.69, range 5–10). Patients estimated that their optimal amount of time of headband usage would be 7 to 8 min. Three of five patients with prior experience with mindfulness, reported added value of Muse to this experience.

Regarding the usability, patients reported improved body awareness and control/ability to regulate peace of mind. Asked what they had learned from the Muse intervention, patients stated: discovering ways of breathing/thoughts to help calm down and relax, slow down thoughts, and build in moments of rest. Overall, patients' feedback indicated they regarded the neuromeditation training feasible and usable, except for the perpetuation of the daily exercises at home. Also, the mean SUS-score of the sample indicated an acceptable, bordering good, usability of the Muse device and app.

Usability: outcomes neuromeditation sessions

In Table 5, means and standard deviations of the within session measurements are presented. MOAS scores indicated no physical aggression towards others and some fluctuating auto-aggression, aggression towards object and verbal aggression towards others during the research period. All tension measures show a decrease after the headband usage. Participants encountered tension more frequently related to their bodies than their thoughts. The neuromeditation indicators, birds and recoveries, increased with every session, except for a drop in recoveries in the last session. Calm state remained similar during the first three sessions and increased slightly in the last.

Usability: outcomes pre- and post-test

Table 6 describes the means and standard deviations of the pre- and post-test outcome measures of the 4-weeks research period.

Due to the small sample size, pre- and post-test results can only be compared on face value, see Table 6. Impulsivity, body awareness and aggression decreased, except for the aggression subscale anger,

TABLE 6 Means and standard deviations of pre- and post-test outcome measures ($N = 8$).

	T1	T2
	M(SD)	M(SD)
Impulsivity		
BIS-11	72.38(7.29)	68.38(9.16)
Body awareness		
ABSQ	46.63(13.84)	45(11.07)
SBQ Body awareness	3.35(0.42)	3.33(0.47)
SBQ Dissociation	3.06(0.49)	2.81(0.48)
Executive functioning		
PGNG Task PCTT	77.49(17.50)	79.21(15.20)
PGNG Task PCIT	66.01(28.54)	68.55(29.86)
Aggression		
AVL-AV Physical	11(3.16)	9.38(3.74)
AVL-AV Verbal	8.75(1.28)	7.25(0.89)
AVL-AV Anger	9.63(2.20)	9.63(2.50)
AVL-AV Hostility	9.88(3.56)	9.38(3.20)

which remained the same. Executive function increased with more correct and inhibition trials.

Discussion

In this multi-method pilot study, the primary goal was to assess the feasibility and usability of the research protocol of a neuromeditation training offered to forensic outpatients. This neuromeditation training was introduced as an additional intervention to the usual treatment (primarily CBT), using the integrative neurofeedback and meditation technology of Muse. Six therapists (all females) and eight patients (6 males and 2 females) participated in the study.

The feasibility and usability of neuromeditation were evaluated through participants' feedback, providing insights into its practicality and perceived benefit. Patients expressed motivation to engage in neuromeditation to enhance relaxation and bodily connectedness. Notably, patient expectations were either met or exceeded following the neuromeditation training, corresponding with findings from Hunkin et al. (2021a) regarding mindfulness experiences.

Both therapists and patients acknowledged the study protocol's comprehensiveness. However, both also voiced a preference toward lengthier and more frequent neuromeditation sessions. Patients suggested an optimal session duration of 7–8 min, in agreement with the findings of Hunkin et al. (2021a) who reported favorable effects from 7-min neuromeditation sessions. Furthermore, Hunkin and colleagues underscored the potential benefits of introducing a training phase for app familiarity and auditory feedback to overcome performance hindrances caused by stress. The inclusion of a training phase holds particular promise, considering that participants in future studies might have less prior meditation experience than the current sample.

In terms of questionnaire relevance, it became apparent that not all measures were equally pertinent. The absence of improvements in the Modified Overt Aggression Scale (MOAS), might suggest that aggression is an unsuitable outcome measure of the neuromeditation intervention. This aligns with patients being in forensic treatment not only for aggression but also for other issues, such as compulsive stealing. The common denominator was impulse control problems. Brazil et al. (2018) also emphasize that in identifying appropriate treatment methods for offenders, the underlying biopsychosocial constructs are better indicators of treatment applicability than the operationalization of aggression through behavioral constructs. Hence, using the MOAS to operationalize aggression may also be inadequate for this purpose.

Furthermore, the lack of applicability of exercises in the Muse app could be addressed by consulting patients about exercise preferences and offering alternatives from established treatment protocols (e.g., Re-Art, Hoogsteder and Bogaerts, 2018) or external resources like YouTube. Additionally, collaborative planning for at-home practice, considering suitable settings and timing to minimize disruptions during exercises, emerged as an important strategy to address challenges. Selecting an appropriate phase for the introduction of the intervention is also vital; one therapist noted that the pre-treatment phase might not be the most suitable time due to other priorities. The dropout of three of the 11 enrolled patients further emphasizes the significance of appropriate timing to prevent disruption in treatment progress.

Therapists had the impression that five out of eight patients struggled with the cognitive executive function task, while only two patients self-reported difficulties. This discrepancy in perspectives might be attributed to therapists' professional focus on patients' behaviors observed in sessions. In contrast, patients likely evaluated based on a wider range of situations, including those not discussed or observed during the sessions. Both perspectives, one not necessarily more reliable than the other, contribute to a more complete picture. These divergences only underscore the necessity of acknowledging and discussing both therapist and patient perspectives in treatment. In this study, therapists' and patients' agreement in wishing to continue neuromeditation in regular treatment, reflects their shared acknowledgment of its supplementary value.

Usability

At first glance, improvements were evident across all self-report measures, except the anger scale of the aggression questionnaire, and the executive functioning task. Although no statistical analyses could be performed due to small sample size, patients reported experienced improvements in overall impulsivity problems after neuromeditation. This substantiated in modest yet impactful enhancements such as improved performance in achieving "calm" states during neuromeditation training, reduced overall tension post-neuromeditation, lowered scores in impulsivity and aggression-related metrics (excluding anger), heightened body connectedness, and augmented inhibitory control performance in neuropsychological tasks. Therapists also corroborated these positive changes, observing heightened calmness, diminished stress, reduced impulsiveness, and increased relaxation capacity in patients. Notably, therapists expressed more optimism regarding changes than patients, indicating improvement in seven of eight patients, whereas five of eight patients acknowledged experiencing positive changes.

Based on the motivational direction model of frontal asymmetry in which relative left frontal cortical activity is associated with approach motivation (Harmon-Jones, 2003), there is a risk that neuromeditation could also suppress positive behaviors such as social behavior. Various studies show that meditation promotes social behaviors (Engert et al., 2023). There is however no research that takes into account the effects of neuromeditation on social behavior. Future studies should look at a range of emotional and social behaviors to better understand the full impact of the treatment.

Although this was primarily a feasibility and usability study, and we cannot conclusively determine its effectiveness, the positive results observed in some patients are noteworthy. These results across various measures indicate the potential benefits of combining meditation with neurofeedback.

Strengths and limitations

One of the notable strengths of this pilot study lies in its pioneering approach, integrating neuromeditation technology into the treatment of forensic outpatients with impulse control problems. The study's emphasis on user experience and feedback from both therapists and patients provides valuable insights into the feasibility and usability of this innovative intervention for offenders.

However, several limitations should be acknowledged. First and foremost, the small sample size restricts the generalizability of the findings. The participants were characterized by their familiarity with substance use (75%) and meditation (63%). Individuals with substance use problems have shown to exhibit higher impulsivity than non–substance-users (Moeller and Dougherty, 2002), which is a concern in forensic treatment since it poses a risk for violence in offenders (Pickard and Fazel, 2013). These factors might have resulted in overrepresentation of these characteristics in our study sample. The presence of previous meditation experience might have positively impacted motivation, potentially lowering the reluctance to participate. Conversely, individuals lacking familiarity with meditation might be less inclined to engage with novel meditation-based interventions. This selection bias limits the extent to which the results can be extended to a broader range of forensic patients. Additionally, the relatively short duration of the intervention (4 weeks) and relatively little headband usage may have hindered the emergence of more pronounced changes in measured outcomes. A longer intervention period might be necessary to observe substantial alterations in behaviors and self-regulation. Another aspect requiring attention pertains to the length and format of the pre- and post-intervention assessment battery. While the inclusion of such measures is crucial for evaluating the outcomes, the potential burden of lengthy questionnaires and evaluation forms should not be underestimated. To enhance participant engagement and minimize response fatigue, these assessment tools should remain concise, focused, and pertinent to the objectives at hand.

A final disadvantage, is that as non-developers we have little insight into Muse's headband and app specifications and therefore lack certain information to get a better idea of its working mechanism processes. For the benefit of effect studies, this would need further investigation. To enhance the effectiveness of studies, further exploration is necessary. Future research should consider how technological factors, like the exact placement of EEG electrodes, relates to treatment effectiveness for impulse control problems in offenders.

Clinical implications

The outcomes of this pilot study hold clinical implications for the treatment of forensic outpatients with impulse control problems. Neuromeditation, as facilitated by the Muse technology, offers a unique avenue for enhancing self-regulatory skills and reducing impulsivity. By focusing on real-time monitoring of brain activity and coupling it with meditation practices, individuals can gain the ability to access and maintain desired states of relaxation. The encouraging feedback from both therapists and patients, indicating positive changes in relaxation skills and reductions in impulsivity, highlights the potential of this approach. The study's findings underscore the need for a nuanced understanding of intervention timing within the treatment process. The introduction of neuromeditation might be most effective when patients are not overwhelmed with other treatment priorities, allowing them to fully engage with the intervention. The varied experiences of the study's participants point to the importance of tailoring neuromeditation exercises to suit individual preferences, and ensuring that patients can incorporate them into their daily routines.

General conclusion

In conclusion, this pilot study ventures into the realm of neuromeditation as an innovative method for enhancing self-regulation and reducing impulsivity among forensic outpatients. The results suggest that neuromeditation, implemented through the Muse technology, holds promise in employing this intervention in this population. While the study's small sample size and short intervention duration warrant caution in drawing conclusions, the findings provide valuable insights into the feasibility and usability of this approach. The positive feedback from both therapists and patients, coupled with the observed improvements in relaxation skills and reductions in impulsivity, point towards a potential value of neuromeditation as a supplementary treatment modality.

This study lays the foundation for future research endeavors in this domain, emphasizing the importance of larger-scale studies to evaluate the impact of neuromeditation on forensic patient outcomes. By addressing the methodological limitations and incorporating the suggestions for protocol improvement, further investigations can shed light on the true potential of neuromeditation in augmenting offender treatment. While the findings are preliminary, they signal a promising direction for advancing the field of forensic psychology through the integration of neuroscientific knowledge and innovative technologies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Internal Review Board of the Van der Hoeven Clinic, indicating that it complies with the ethical guidelines of the institution and all laws and regulations in the Netherlands and Europe (2021-2-SC). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

AS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. JW: Data curation, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing. JH: Methodology, Project administration, Resources, Supervision, Writing – review & editing, Conceptualization, Formal analysis, Validation.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Babcock, J. C., Green, C. E., and Robie, C. (2004). Does batterers' treatment work? An analytic review of domestic violence treatment. *Clin. Psychol. Rev.* 23, 1023–1053. doi: 10.1016/j.cpr.2002.07.001

Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Int. J. Hum. Comput.* 24, 574–594. doi: 10.1080/10447310802205776

Beaudry, G., Yu, R., Perry, A. E., and Fazel, S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: a systematic review and meta-analysis of randomised controlled trials. *Lancet Psychiatry* 8, 759–773. doi: 10.1016/S2215-0366(21)00170-X

Bertsch, K., Florange, J., and Herpertz, S. C. (2020). Understanding brain mechanisms of reactive aggression. *Curr. Psychiatry Rep.* 22, 81–16. doi: 10.1007/s11920-020-01208-6

Bhayee, S., Tomaszewski, P., Lee, D. H., Moffat, G., Pino, L., Moreno, S., et al. (2016). Attentional and affective consequences of technology supported mindfulness training: a randomised, active control, efficacy trial. *BMC Psychol.* 4:60. doi: 10.1186/s40359-016-0168-6

Bijlsma, J., Geukes, S. H., Meynen, G., Raemaekers, M. A. H., Ramsey, N. F., Thomas, S., et al. (2022). *Kansen en risico's van de toepassing van neurotechnologie in het strafrecht [opportunities and risks of the application of neurotechnology in criminal law]*. Utrecht, The Netherlands: Utrecht University Repository.

Borghino, S., Gallacher, M., and Lawrence, D. (2022). Exploring the use of neurofeedback with an adult male with an autistic Spectrum condition and a history of sexual offending: a single case study. *J. Forensic Psychol. Res. Pract.* 24, 372–388. doi: 10.1080/24732850.2022.2130736

Brazil, I. A., van Dongen, J. D., Maes, J. H., Mars, R. B., and Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: from behavior to biocognition. *Neurosci. Biobehav. Rev.* 91, 259–277. doi: 10.1016/j.neubiorev.2016.10.010

Brooke, J. (1996). "SUS: a "quick and dirty" usability scale" in *Usability evaluation in industry*. eds. P. W. Jordan, B. A. Thomas, B. Weerdmeester and I. L. McClelland (London, UK: Taylor & Francis).

Buitelaar, N. J., Posthumus, J. A., Scholing, A., and Buitelaar, J. K. (2014). Impact of treatment of ADHD on intimate partner violence (ITAP), a study protocol. *BMC Psychiatry* 14, 336–339. doi: 10.1186/s12888-014-0336-2

Buss, A. H., and Perry, M. (1992). The aggression questionnaire. *J. Pers. Soc. Psychol.* 63, 452–459. doi: 10.1037/0022-3514.63.3.452

Chaibi, I., Bouchatta, O., Bennis, M., and Ba-M'hamed, S. (2023). The role of the anterior cingulate cortex in aggression and impulsivity. *Behav. Neurosci.* 137, 155–169. doi: 10.1037/bne0000552

Crivelli, D., Fronda, G., Venturella, I., and Balconi, M. (2019). Supporting mindfulness practices with brain-sensing devices. Cognitive and electrophysiological evidences. *Mindfulness (N Y)* 10, 301–311. doi: 10.1007/s12671-018-0975-3

Dambacher, F., Sack, A. T., Lobbestael, J., Arntz, A., Brugman, S., and Schuhmann, T. (2015). Out of control: evidence for anterior insula involvement in motor impulsivity and reactive aggression. *Soc. Cogn. Affect. Neurosci.* 10, 508–516. doi: 10.1093/scan/nsu077

Eckhardt, C. I., Murphy, C. M., Whitaker, D. J., Sprunger, J., Dykstra, R., and Woodard, K. (2013). The effectiveness of intervention programs for perpetrators and victims of intimate partner violence. *Partn. Abus.* 4, 196–231. doi: 10.1891/1946-6560.4.2.196

Eisenberg, M. J., Hutten, J. C., Van der Veeken, F. C. A., Bouman, Y. H. A., Bogaerts, S., and Van Horn, J. E. (2020). *Validatie FARE: Een multi-center studie naar de interbeoordelaarsbetrouwbaarheid, convergente validiteit, veranderbaarheid en de predictieve validiteit van de Forensische Ambulante Risico Evaluatie (FARE) [Validation FORE: A multi-center study of the inter-rater reliability, convergent validity, changeability and predictive validity of the forensic outpatient risk evaluation (FORE)]*. Utrecht, The Netherlands: Kwaliteit Forensische Zorg.

Engert, V., Klimecki, O. M., Kanske, P., and Universities Research Group. (2023). Spreading positive change: Societal benefits of meditation. *Frontiers in Psychiatry.* 14, 1038051. doi: 10.3389/fpsyt.2023.1038051

Feder, L., and Wilson, D. B. (2005). A meta-analytic review of court-mandated batterer intervention programs: can courts affect abusers' behavior? *J. Exp. Criminol.* 1, 239–262. doi: 10.1007/s11292-005-1179-0

Fergusson, D. M., Boden, J. M., and Horwood, L. J. (2013). Childhood self-control and adult outcomes: results from a 30-year longitudinal study. *J. Am. Acad. Child Adolesc. Psychiatry* 52, 709–717.e1. doi: 10.1016/j.jaac.2013.04.008

Fielenbach, S., Donkers, F. C., Spreen, M., Visser, H. A., and Bogaerts, S. (2018). Neurofeedback training for psychiatric disorders associated with criminal offending: a review. *Front. Psych.* 8:313. doi: 10.3389/fpsyt.2017.00313

Gallo, G. G., Curado, D. F., Opaleye, E. S., Donate, A. P. G., Scattone, V. V., and Noto, A. R. (2021). Impulsivity and mindfulness among inpatients with alcohol use disorder. *Subst. Use Misuse* 56, 25–32. doi: 10.1080/10826084.2020.1833924

Harmon-Jones, E. (2003). Clarifying the emotive functions of asymmetrical frontal cortical activity. *Psychophysiology* 40, 838–848. doi: 10.1111/1469-8986.00121

Heatherton, T. F., and Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends Cogn. Sci.* 15, 132–139. doi: 10.1016/j.tics.2010.12.005

Henwood, K. S., Chou, S., and Browne, K. D. (2015). A systematic review and meta-analysis on the effectiveness of CBT informed anger management. *Aggress. Violent Behav.* 25, 280–292. doi: 10.1016/j.avb.2015.09.011

Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., and Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspect. Psychol. Sci.* 6, 537–559. doi: 10.1177/1745691611419671

Hong, J., and Park, J. H. (2022). Efficacy of neuro-feedback training for PTSD symptoms: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* 19:13096. doi: 10.3390/ijerph192013096

Hoogsteder, L. M., and Bogaerts, S. (2018). *Justitieleinterventies.nl: beschrijving Agressie Regulatie op Maat Ambulant voor Volwassenen [Description custom aggression regulation ambulatory for adults]*. Nederlands Jeugdinstituut, Movisie en Trimbos instituut. Available at: www.justitieleinterventies.nl

Hornsveld, R. H., Murris, P., and Kraaimaat, F. W. (2009). *Drie zelfrapportage vragenlijsten voor de forensische psychiatrie [Three self-report questionnaires for forensic psychiatry]*. De Bondt grafimedia communicatie.

Hortensius, R., Schutter, D., and Harmon-Jones, E. (2012). When anger leads to aggression: induction of relative left frontal cortical activity with transcranial direct current stimulation increases the anger–aggression relationship. *Soc. Cogn. Affect. Neurosci.* 7, 342–347. doi: 10.1093/scan/nsr012

Hunkin, H., King, D. L., and Zajac, I. T. (2021a). EEG neurofeedback during focused attention meditation: effects on state mindfulness and meditation experiences. *Mindfulness* 12, 841–851. doi: 10.1007/s12671-020-01541-0

Hunkin, H., King, D. L., and Zajac, I. T. (2021b). Evaluating the feasibility of a consumer-grade wearable EEG headband to aid assessment of state and trait mindfulness. *J. Clin. Psychol.* 77, 2559–2575. doi: 10.1002/jclp.23189

Ireland, J. L., and Archer, J. (2008). Impulsivity among adult prisoners: a confirmatory factor analysis study of the Barratt impulsivity scale. *Personal. Individ. Differ.* 45, 286–292. doi: 10.1016/j.paid.2008.04.012

Kay, S. R., Wolkenfeld, F., and Murrill, L. M. (1988). Profiles of aggression among psychiatric patients: I. Nature and prevalence. *J. Nerv. Ment. Dis.* 176, 539–546. doi: 10.1097/00005053-198809000-00007

Knehans, R., Schuhmann, T., Roef, D., Nelen, H., à Campo, J., and Lobbestael, J. (2022). Modulating behavioural and self-reported aggression with non-invasive brain stimulation: a literature review. *Brain Sci.* 12:200. doi: 10.3390/brainsci12020200

Landenberger, N. A., and Lipsey, M. W. (2005). The positive effects of cognitive–behavioral programs for offenders: a meta-analysis of factors associated with effective treatment. *J. Exp. Criminol.* 1, 451–476. doi: 10.1007/s11292-005-3541-7

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

Langenecker, S. A., Zubieta, J. K., Young, E. A., Akil, H., and Nielson, A. K. (2007). A task to manipulate attentional load, set-shifting and inhibitory control: convergent validity and test-retest reliability of the parametric go/no-go test. *J. Clin. Exp. Neuropsychol.* 29, 842–853. doi: 10.1080/13803390601147611

Larson, L. I. (2019). An exploration of neurofeedback treatment in a sample of men arrested for domestic violence with correlates of anger, aggression, and stress. Unpublished master's thesis, Idaho State University, Department of Nursing, College of Health Professions.

Lima, K. H. M., Gomes, J. S., and Tucci, A. M. (2022). Electroencephalographic neurofeedback as a tool for reducing harm and risk associated with alcohol use disorder: a critical review. *Drug Alcohol Rev.* 41, 594–602. doi: 10.1111/dar.13387

Lipsey, M. W., Landenberger, N. A., and Wilson, S. J. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Syst. Rev.* 3, 1–27. doi: 10.4073/csr.2007.6

Lloyd, C. D., Chadwick, N., and Serin, R. C. (2014). Associations between gambling, substance misuse and recidivism among Canadian offenders: a multifaceted exploration of poor impulse control traits and behaviours. *Int. Gambl. Stud.* 14, 279–300. doi: 10.1080/14459795.2014.913301

Loeber, R., Menting, B., Lynam, D. R., Moffitt, T. E., Stouthamer-Loeber, M., Stallings, R., et al. (2012). Findings from the Pittsburgh youth study: cognitive impulsivity and intelligence as predictors of the age–crime curve. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 1136–1149. doi: 10.1016/j.jaac.2012.08.019

Moeller, F. G., and Dougherty, D. M. (2002). Impulsivity and substance abuse: what is the connection? *Addict. Disord. Treat.* 1, 3–10. doi: 10.1097/00132576-200205000-00002

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci.* 108, 2693–2698. doi: 10.1073/pnas.1010076108

Moreno-García, I., Cano-Crespo, A., and Rivera, F. (2022). Results of neurofeedback in treatment of children with ADHD: a systematic review of randomized controlled trials. *Appl. Psychophysiol. Biofeedback* 47, 145–181. doi: 10.1007/s10484-022-09547-1

Muse. (2023). Muse: The Brain Sensing Headband. Available at: https://choosemuse.com/ (Accessed April 7, 2023)

Patton, J. H., Stanford, M. S., and Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *J. Clin. Psychol.* 51, 768–774. doi: 10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Pearson, F. S., Lipton, D. S., Cleland, C. M., and Yee, D. S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime Delinq.* 48, 476–496. doi: 10.1177/001112870204800306

Pickard, H., and Fazel, S. (2013). Substance abuse as a risk factor for violence in mental illness: some implications for forensic psychiatric practice and clinical ethics. *Curr. Opin. Psychiatry* 26, 349–354. doi: 10.1097/YCO.0b013e328361e798

Price, C. J., and Thompson, E. A. (2007). Measuring dimensions of body connection: body awareness and bodily dissociation. *J. Altern. Complement. Med.* 13, 945–953. doi: 10.1089/acm.2007.0537

Sedlmeier, P., Loße, C., and Quasten, L. C. (2018). Psychological effects of meditation for healthy practitioners: an update. *Mindfulness* 9, 371–387. doi: 10.1007/s12671-017-0780-4

Sergiou, C. S., Santarnecchi, E., Romanella, S. M., Wieser, M. J., Franken, I. H. A., Rassin, E. G. C., et al. (2022). tDCS targeting the ventromedial prefrontal cortex reduces reactive aggression and modulates electrophysiological responses in a forensic population. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 7, 95–107. doi: 10.1016/j.bpsc.2021.05.007

Sokhadze, T. M., Cannon, R. L., and Trudeau, D. L. (2008). EEG biofeedback as a treatment for substance use disorders: review, rating of efficacy, and recommendations for further research. *Appl. Psychophysiol. Biofeedback* 33, 1–28. doi: 10.1007/s10484-007-9047-5

Taj-Eldin, M., Ryan, C., O'Flynn, B., and Galvin, P. (2018). A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers. *Sensors* 18:4271. doi: 10.3390/s18124271

Tarrant, J. (2020). "Neuromeditation: the science and practice of combining neurofeedback and meditation for improved mental health" in *Smart biofeedback perspectives and applications*. ed. E. D.-Y. Liao (London, UK: IntechOpen).

Van der Maas, C. C. (2015). *Psychomotor therapy in chronic pain rehabilitation enhancing body awareness in multidisciplinary treatment*. [Doctoral thesis] Vrije Universiteit Amsterdam.

van Doren, J., Arns, M., Heinrich, H., Vollebregt, M. A., Strehl, U., and Loo, S. K. (2019). Sustained effects of neurofeedback in ADHD: a systematic review and meta-analysis. *Eur. Child Adolesc. Psychiatry* 28, 293–305. doi: 10.1007/s00787-018-1121-4

Van Horn, J. E., Eisenberg, M. J., Bouman, Y. H. A., Van den Hanenberg, F. J. A. C., Van der Put, C. E., and Bogaerts, S. (2020). *Handleiding Forensisch Ambulante Risico Evaluatie versie 2 –FARE v2 Volwassen cliënten vanaf 18 jaar [Manual Forensic Outpatient Risk Evaluation version 2 –FORE v2 Adult clients]*. Utrecht: Kwaliteit Forensische Zorg.

van Horn, J., van der Schoot, A., Wilpert, J., Engelbregt, H. J., and Brand, N. (2023). A pilot study of the convergent and discriminant validity of the Dutch version of the parametric go/no-go task. *Cogn. Behav. Neurol.* doi: 10.1097/WNN.0000000000000363

Van Outsem, R. (2011). The applicability of neurofeedback in forensic psychotherapy: a literature review. *J. Forensic Psychiatry Psychol.* 22, 223–242. doi: 10.1080/14789949.2010.528012

Vasconcelos, A. G., Malloy-Diniz, L., and Correa, H. (2012). Systematic review of psychometric proprieties of Barratt impulsiveness scale version 11 (BIS-11). *Clin. Neuropsychiatry* 9, 61–74. doi: 10.1037/t05661-000

Von Hilsheimer, G., and Quirk, D. A. (2006). Origin of EEG biofeedback for remediating misbehavior. *Behav. Anal. Today* 7, 492–507. doi: 10.1037/h0100095

Votruba, K. L., and Langenecker, S. A. (2013). Factor structure, construct validity, and age-and education based normative data for the parametric go/no-go test. *J. Clin. Exp. Neuropsychol.* 35, 132–146. doi: 10.1080/13803395.2012.758239

Wilson, D. B., Bouffard, L. A., and MacKenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Crim. Justice Behav.* 32, 172–204. doi: 10.1177/0093854804272889

Zicht op forensische zorg [View on forensic care]. (2023). Available at: https://zichtopforensischezorg.nl/expertisecentrum-forensische-psychiatrie-efp/forensisch-netwerk/programma-kfz/calls/veldnormen (Accessed April 3, 2023)

Zwets, A. J., Hornsveld, R., Kraaimaat, F. W., Kanters, T., Muris, P., and Van Marle, H. (2014). The psychometric properties of the anger bodily sensations questionnaire (ABSQ). *J. Forensic Psychiatry Psychol.* 25, 432–450. doi: 10.1080/14789949.2014.925957

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 23 February 2024
DOI 10.3389/fpsyg.2024.1328839

OPEN ACCESS

EDITED BY
Heng Choon Chan,
University of Birmingham, United Kingdom

REVIEWED BY
Alexander Hochdorn,
University of Brasilia, Brazil
Ezequiel Norberto Mercurio,
University of Buenos Aires, Argentina

*CORRESPONDENCE
Juliette C. Hutten
jhutten@dewaagnederland.nl

RECEIVED 27 October 2023
ACCEPTED 05 February 2024
PUBLISHED 23 February 2024

CITATION
Hutten JC, van Horn JE, Hoppenbrouwers SS, Ziermans TB, Geurts HM and the Forensic Neuropsychology Consortium (2024) Neuropsychological assessment of aggressive offenders: a Delphi consensus study. *Front. Psychol.* 15:1328839. doi: 10.3389/fpsyg.2024.1328839

Neuropsychological assessment of aggressive offenders: a Delphi consensus study

Juliette C. Hutten[1,2]*, Joan E. van Horn[1], Sylco S. Hoppenbrouwers[1], Tim B. Ziermans[2], Hilde M. Geurts[2] and the Forensic Neuropsychology Consortium

[1]De Waag (Outpatient Forensic Mental Health Clinic), Forensic Care Specialists, Utrecht, Netherlands, [2]Brain and Cognition, Department of Psychology, Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, Netherlands

Objective: This study explores the intricate relationship between cognitive functioning and aggression, with a specific focus on individuals prone to reactive or proactive aggression. The purpose of the study was to identify important neuropsychological constructs and suitable tests for comprehending and addressing aggression.

Methods: An international panel of 32 forensic neuropsychology experts participated in this three-round Delphi study consisting of iterative online questionnaires. The experts rated the importance of constructs based on the Research Domain Criteria (RDoC) framework. Subsequently, they suggested tests that can be used to assess these constructs and rated their suitability.

Results: The panel identified the RDoC domains Negative Valence Systems, Social Processes, Cognitive Systems and Positive Valence Systems as most important in understanding aggression. Notably, the results underscore the significance of Positive Valence Systems in proactive aggression and Negative Valence Systems in reactive aggression. The panel suggested a diverse array of 223 different tests, although they noted that not every RDoC construct can be effectively measured through a neuropsychological test. The added value of a multimodal assessment strategy is discussed.

Conclusions: This research advances our understanding of the RDoC constructs related to aggression and provides valuable insights for assessment strategies. Rather than suggesting a fixed set of tests, our study takes a flexible approach by presenting a top-3 list for each construct. This approach allows for tailored assessment to meet specific clinical or research needs. An important limitation is the predominantly Dutch composition of the expert panel, despite extensive efforts to diversify.

KEYWORDS

neuropsychological tests, forensic psychology, aggression, violence, Delphi technique, Research Domain Criteria (RDoC)

1 Introduction

Aggressive offenses have far-reaching consequences for individuals and society, including financial strain on health and justice sectors, public safety issues, reduced quality of life for victims, their relatives, and the offenders (Patel and Taylor, 2012; Langton et al., 2014; Rivara et al., 2019). Neuropsychological profiling is an underused clinical tool to assess the complex web of risk factors for aggressive behavior.

This is surprising as the intricate relationship between cognitive functioning and reactive and proactive aggression has been widely studied. Although empirical studies and systematic reviews have uncovered neurocognitive mechanisms underlying reactive and proactive aggression (Alcázar-Córcoles et al., 2010; Kuin et al., 2015; Van De Kant et al., 2020), expert knowledge on individual neuropsychological assessment has not been integrated into the current body of research. An overarching framework that bridges the gap between fundamental research and clinical experience is therefore much needed. In the current study, a panel of experts is asked (i) which Research Domain Criteria (RDoC) domains (Insel et al., 2010; further explained below) are important in explaining reactive vs. proactive aggression and, (ii) which neuropsychological tasks are suitable for assessing those domains.

A common definition of aggression is "behavior that is intended to harm another person who is motivated to avoid that harm" (Allen and Anderson, 2017). Notably, aggressive offenders make up a substantial proportion (up to 70%) of prisons, forensic hospitals and outpatient mental health facilities (McMurran et al., 2000; Völlm et al., 2018). There is a great need for research into the risk factors of aggressive behavior to help reduce recidivism (Smeijers, 2017; Wigham et al., 2022). One of the potential risk factors for aggression that warrants exploration is cognitive functioning, particularly through neuropsychological assessments. Cognitive limitations are more prevalent among offenders than in the general population (Ogilvie et al., 2011), particularly among aggressive offenders (Cruz et al., 2020). For example, research found a prevalence of clinically significant executive deficits (a subset of cognitive functions) in an offender population ranging from 5.2% to 27.2% (correctional offenders) and 9.5–35.7% (forensic psychiatric patients), compared to 2.5% in the general population (Shumlich et al., 2019). Furthermore, multiple factors can be at the root of cognitive limitations, including traumatic brain injury, substance abuse, and attention deficit hyperactivity disorder, all of which are more prevalent among offenders (Harris, 2006; Ginsberg et al., 2010; Farrer and Hedges, 2011; Frost et al., 2013; Fayyad et al., 2017; Hellenbach et al., 2017; Muñoz García-Largo et al., 2020; Matheson et al., 2022). As such, it is necessary to further highlight the role of cognitive factors in the context of offending behavior, and this study aims to do so by improving knowledge about neuropsychological assessment in forensic populations.

1.1 Reactive and proactive aggression

The term "aggression" refers to a spectrum of acts that range from shouting or pushing to aggravated assault or homicide. By the definition stated above, rape, sexual assault, and robbery would also be classified as aggressive offenses. As the literature shows, most offenders are generalists, meaning they commit more than one type of crime in their lives (Simon, 1997; Soothill et al., 2000; Sullivan et al., 2006). Therefore, we chose to include aggressive sexual- or property crimes while non-aggressive crimes such as fraud were outside the scope of this study. Understanding the different determinants of aggression has been a subject of interest in various fields such as psychology, criminology, and neuroscience since the mid-20th century. Several taxonomies have been proposed in the literature (Parrott and Giancola, 2007; Krahé, 2013), but there is no consensus yet about which categorization is most appropriate. The most well-known distinction is the reactive-proactive dichotomy, sometimes referred to as hostile-instrumental (Buss, 1961). Reactive aggression occurs in reaction to a provocation or frustration and is impulsive in nature. Proactive aggression on the other hand is generally goal-directed and premeditated. Both types of aggression can occur within an individual, and thus, the strict classification into one of these two categories has been disputed in the literature (Bushman and Anderson, 2001). Currently, a dimensional view of aggression is favored, acknowledging that individuals often exhibit varying degrees of both reactive and proactive aggression rather than rigidly categorizing them into distinct types. Interestingly, research on factors associated with or related to reactive and proactive aggression provides empirical support for the usefulness of the distinction. For example, reactive aggression has been linked to heightened emotional reactivity, impulsivity, verbal impairments and impairments in executive functioning, and hostile attribution bias. Proactive aggression on the other hand is linked to a lack of moral emotions, callous and unemotional traits, and low physiological arousal (Cima and Raine, 2009). To summarize, individuals can exhibit both types of aggression, with a tendency toward one type, reflecting a predominant behavioral disposition. As both types of aggression appear to be related to different constructs, the current study considers both types of aggression separately.

1.2 Research Domain Criteria (RDoC)

The National Institute of Mental Health (NIMH) developed the RDoC (Insel et al., 2010), to—as opposed to traditional categorial diagnostic systems such as the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2022)— investigate core dimensions of functioning that underlie various mental health conditions. In addition, as aggression can occur within various mental health conditions such as personality disorders, intermittent explosive disorder and conduct disorder, the RDoC framework provides a transdiagnostic perspective to uncover shared mechanisms that contribute to aggression across these diverse disorders. The RDoC describes six domains: (1) Negative Valence Systems, responsible for responses to aversive situations or context, such as fear, anxiety, and loss; (2) Positive Valence Systems, responsible for responses to positive motivational situations or contexts, such as reward seeking, consummatory behavior, and reward/habit learning; (3) Cognitive Systems, responsible for various cognitive processes; (4) Social Processes, which mediate responses to interpersonal settings of various types, including perception and interpretation of others' actions; (5) Sensorimotor Systems, responsible for the control and execution of motor behaviors, and their refinement during learning and development; and (6) Arousal/Regulatory Systems responsible for generating activation of neural systems as appropriate for various contexts, and providing appropriate homeostatic regulation of such systems as energy balance and sleep (see Figure 1).

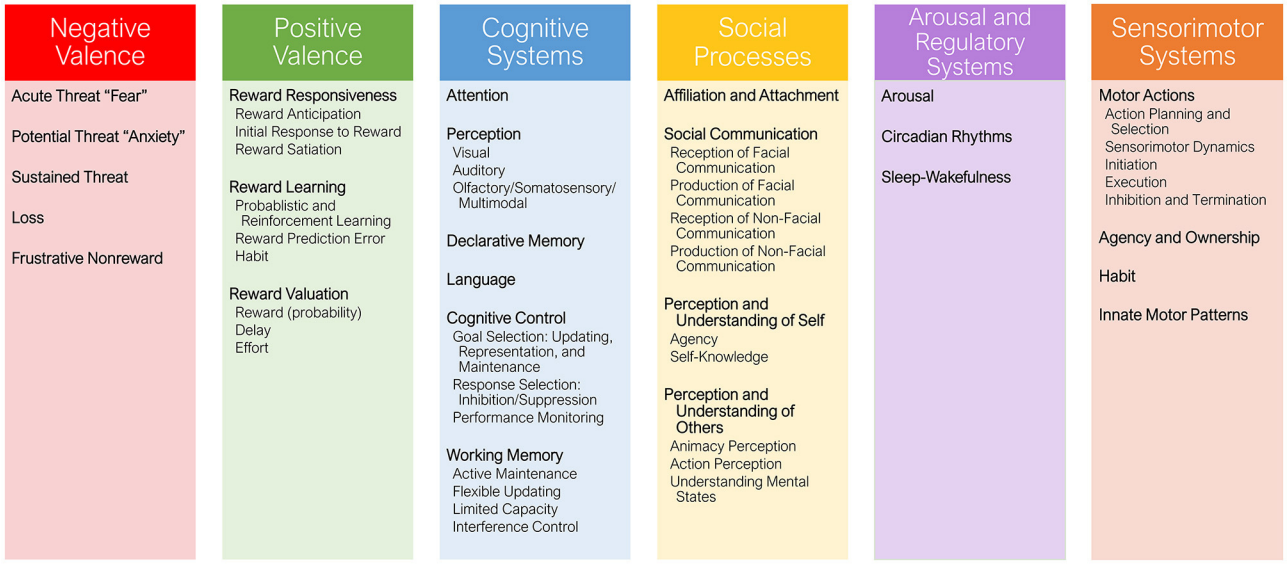


FIGURE 1
Overview of the Research Domain Criteria (RDoC) domains, including 25 constructs (bold) and 33 subconstructs (regular text).

1.3 Aggression and cognitive domains

In this section, the existing knowledge regarding the interplay between aggression and the domains outlined in the RDoC framework is briefly elucidated. If available, we refer to systematic reviews and/or meta-analyses. Neuropsychological studies have revealed differences between (aggressive) offenders and non-offending controls in different RDoC domains, such as Cognitive Systems (including executive functions, attention, and language) (Cohen et al., 2003; Ogilvie et al., 2011; Anderson et al., 2016; Burgess, 2020; Chow et al., 2022), Social Processes (Marsh and Blair, 2008; Karoglu et al., 2022), and Positive/Negative Valence systems (Estrada et al., 2019; Manning, 2020; mainly reward and threat processing). In our recent multi-level meta-analysis, we have studied all domains of cognitive functioning in relation to offending behavior (Hutten et al., preprint). Overall, offenders performed worse on neuropsychological tests than non-offending controls, and this was the case for all of the cognitive domains studied. A notable observation from this meta-analysis was the substantial variation in tests (146 different tests), and the lack of studies from non-Western countries. Through the Delphi method, we aim to gather insights from forensic neuropsychology experts across the world to obtain consensus on the most suitable tests to measure neuropsychological functioning in aggressive offenders. With this, we aim to expand on this empirical knowledge by connecting research findings and their translational application in forensic practice.

The primary goal of offender rehabilitation is reducing recidivism. A recent global systematic review found 2-year recidivism rates of 18–55% after incarceration and 10–47% after community sentences (Yukhnenko et al., 2023). Psychological treatment has a small but positive effect on recidivism in violent offenders, with a 10.2% difference in recidivism between treated vs. non-treated offenders (Papalia et al., 2019). Despite these findings, there remains a need for further enhancements in intervention strategies to reduce recidivism more effectively. More knowledge on the relation between the RDoC domains and aggression could enhance offender rehabilitation in several ways. Studies have found worse executive functioning in recidivists compared to first time offenders (Ross and Hoaken, 2011; Sánchez de Ribera et al., 2022). Conventional risk assessment tools appear to have reached their ceiling effect, achieving a moderate area under the curve of 0.70, (Monahan and Skeem, 2014; Ogonah et al., 2023). Risk-assessment tools often measure cognitive factors like impulsivity and self-control through less objective methods such as observer ratings and self-reports. Neuropsychological tasks are considered more objective, excluding the impact of compromised self-insight (Steward and Kretzmer, 2022). Accordingly, expanding risk assessment to include neuropsychological and neurobiological factors alongside the existing psychosocial risk factors may enhance the accuracy recidivism predictions (Aharoni et al., 2013, 2014; Haarsma et al., 2020; Zijlmans et al., 2021; Nauta-Jansen, 2022). In addition to predicting recidivism, cognitive functioning—in particular inhibitory and cognitive flexibility difficulties—also appears to predict treatment dropout and treatment success (Fishbein et al., 2009; Cornet et al., 2014). Identifying the specific cognitive domains that are impaired in offenders and related to aggression is crucial to providing targeted interventions and reducing the risk of criminal behavior. For example, an aggression regulation training could be suitable for individuals with aggression arising from inhibitory problems, while people with difficulties in emotion recognition might benefit more from an emotion recognition training (Li et al., 2023). Hence, misidentification of the determinants of the aggression may lead to suboptimal treatment.

Although a clear link has been demonstrated between cognitive limitations and aggression, the use of neuropsychology in forensic settings has not reached its full potential. For example, incorporation of neurobiological information in Dutch pretrial

forensic reports was low and did not rise significantly from 2005 to 2015 (Kempes et al., 2019). Additionally, even when neurobiological factors were acknowledged in relation to the offense, they were often overlooked in discussions about future risk assessment and -management. There are three explanations for this observation which are not mutually exclusive. First of all, clinicians are likely to struggle identifying the most suitable instruments as there is a plethora of neuropsychological tests available. A systematic review on neuropsychological assessment practices in forensic settings found a notable diversity in assessment tools, with 140 different types of tests. The authors conclude that a wide range of neuropsychological functions are being measured by a large number of instruments (Venturi Da Silva and Cavalheiro Hamdan, 2022). Related to this, many tests have multiple outcomes—often measuring different cognitive functions—or multiple ways to calculate the outcomes. This heterogeneity may compromise the reliability of test results and raises questions about how information is understood by clinicians and legal practitioners (Serafim et al., 2015). Second, for most neuropsychological tests normative data are collected from general population samples and have not been validated for the offender population. Possibly, the use of default norm scores leads to insufficient differentiation among individuals in the offender setting (Cornet et al., 2016). As such, it remains unclear which tests are most sensitive and suitable for the aggressive offender population. Third, offender populations present unique challenges in conducting neuropsychological assessments, such as high rates of noncompliance, low motivation (for treatment and/or assessment), and limited education and literacy levels (Hetland et al., 2007; Tuominen et al., 2014). Cultural and linguistic differences may also need to be considered when conducting neuropsychological assessments with offender populations. Considering these challenges, further research and tailored approaches are required to address the selection of suitable tests and norms for the aggressive offender population, to ensure accurate and reliable assessments.

1.4 Study objectives

This study aims to identify the most suitable neuropsychological tests for cognitive assessment within the aggressive offender population, distinguishing between predominantly reactive vs. proactive aggressive offenders. With this, our research contributes to the advancement of forensic neuropsychology. By pinpointing the specific cognitive domains associated with both reactive and proactive aggression, we aim to pave the way for more targeted assessments and interventions in aggressive offender populations. To achieve this goal, we need to bridge the gap between research and clinical practice and strife toward consensus among an international panel of experts from the field of forensic neuropsychology. In the current study, we will apply the Delphi methodology for this purpose. Our objectives encompass two categories of questions posed to the expert panel: firstly, we seek theoretical insights into the constructs commonly associated with aggression, emphasizing their significance in the evaluation of aggressive offenders; secondly, we aim to pinpoint the most suitable tests for this evaluation, thereby facilitating future test selection in forensic contexts.

2 Materials and methods

This study was preregistered at AsPredicted (#103758) and has been approved by the Ethics Review Board (ERB) of the University of Amsterdam (ERB number: 2022-BC-15289).

2.1 The Delphi methodology

We conducted a Delphi consensus study to obtain consensus among an international panel of experts in forensic neuropsychology. While meta-analyses and reviews allow us to have and overview of the current scientific knowledge, the Delphi method allows us to obtain insight into the existing clinical expertise. The Delphi method is a technique used to achieve consensus among a group of experts by soliciting their opinions through a series of questionnaires and providing them with controlled feedback (Dalkey and Helmer, 1963). The Delphi method is based on the concept of collective wisdom, which assumes that the combined opinion of multiple people is closer to the truth than a single individual's perspective (Habibi et al., 2014). Recently, researchers have been striving to achieve consensus on various neuropsychological topics, such as the definition of the term 'impairment' or inconsistent use of test score labels (e.g., Guilmette et al., 2020). Our study aligns with these developments. The Delphi methodology, with its collaborative and iterative nature, serves as an effective tool within this context, facilitating the establishment of a shared foundation for understanding and addressing diverse neuropsychological considerations in the field. This is carried out by aggregating the results of online, anonymous questionnaires in a systematic way. The current study consisted of three rounds, which are described in 2.4 Procedure and data analysis.

2.2 Expert panel selection

Both researchers and clinicians employed in the field of forensic neuropsychology were invited to participate in the panel. Potential researchers were identified based on the articles that emerged from our literature review which is in review (Hutten et al., preprint). The researchers who had a minimum of two publications on the topic of forensic neuropsychology, of which one in the last five years (to confirm that they were still actively engaged in the field) were approached to participate in the Delphi study. For clinicians, the inclusion criterium is at least 4 years' experience as a (clinical) neuropsychologist in the forensic setting. Recruitment took place through the author's networks, (international) societies or networks for neuropsychology/forensic psychology, social media, and through the "snowballing technique" (Iqbal and Pipon-Young, 2009). Panels with 10 to 50 members are recommended for Delphi studies (Turoff, 1975). In total, 127 potential experts were invited personally by email. Sixty-three potential experts started

the questionnaire and provided digital informed consent. Thirty potential experts responded they could not participate (no time: 13, questioned their own expertise: 15, no reason: 2). Finally, 32 experts completed the first-round questionnaire.

2.3 Research Domain Criteria (RDoC)

This Delphi study was based on the RDoC framework (Insel et al., 2010). The RDoC model is a research framework that approaches mental health and psychopathology by examining major domains of basic human neurobehavioral functioning, rather than relying on traditional diagnostic categories. The model consists of six major functional domains (see Figure 1), and each domain is studied by exploring different aspects using constructs that are examined across a range of functioning from normal to abnormal.

2.4 Procedure and data analysis

In three consecutive rounds of online questionnaires (compiled through Qualtrics, 2023), experts rated the importance of a predetermined list of the RDoC constructs on a 5-point scale from 1 "not important" to 5 "essential", with a non-neutral midpoint of 3 (moderately important). Using a non-neutral midpoint forces panelists to deliberate and to decide about the importance of the constructs. If they felt incompetent to answer a question, a "don't know" option was available (Linstone and Turoff, 1975). Subsequently, the panel members provided suggestions for tests that can be used to measure the constructs they rated at least moderately important (rating 3 or higher). In addition, they rated each other's test suggestions as suitable or not suitable for aggressive offenders. Throughout the questionnaires, panel members can provide explanations or reasoning. Before distributing the questionnaire for the first round, two clinical neuropsychologists filled in the questionnaire to provide feedback and ensure clarity of the questions.

After each round, the constructs that did not achieve consensus (about their importance) moved into the subsequent round for re-rating. Our operationalization of consensus is interquartile range (IQR) ≤ 1. For a four to five-point Likert scale, an IQR of 1 or less is considered a high level of consensus (Raskin, 1994; Rayens and Hahn, 2000).

For the importance ratings of the RDoC constructs, means and standard deviations are reported. We conducted Mann-Whitney U tests to analyze the difference in importance scores between reactive and proactive aggression, primarily due to the ordinal nature of the data. For the suitability of tests, we reported the percentage of the panel that rated the test as suitable.

2.4.1 Round 1

The objectives of the first round were (1) to identify the most important RDoC constructs that should be included in the assessment of aggressive offenders, and (2) to collect suggestions for tests that are recommended to assess these constructs. Before the experts started with the main questions, they were asked to fill in some information about their age, gender, profession, current workplace, and academic degree.

Then, the panel members were asked to rate the importance of the RDoC constructs. For the constructs they rated as at least moderately important, they were also asked to rate the importance of the underlying subconstructs. The experts were able to suggest additional constructs not delineated in the RDoC. The research team (JH, JvH, SH, TZ, and HG) evaluated these suggestions to confirm they were not already covered in the RDoC, they were clearly described, and they were within the scope of the RDoC [as suggested by Jorm (2015)]. These additional constructs were then added to subsequent rounds.

Next, for the constructs that they rated as at least moderately important, the experts gave suggestions for tests that they recommend for the assessment of this construct. They could give several suggestions per construct.

2.4.2 Round 2

The 32 panel members who completed round 1 were invited to participate in round 2 of the study, which 26 of them did. (Sub)constructs that did not reach consensus in round 1 were rated again. These constructs were presented along with feedback outlining the average panel rating, each expert's own previous response, and a synopsis of comments that were offered by experts in support of their opinion. In addition, the additional constructs added by the panelists in round 1 were rated for importance. Then, the experts scored the suitability of the recommended tests suggested in round 1 (suitable/not suitable/don't know). If the round 1 tests suggestions were not specific enough—e.g., a test category such as "gambling tests" or a measurement goal such as "verbal comprehension tests"—the panel was asked to specify in this round.

2.4.3 Round 3

The 26 panel members who completed round 2 were invited to participate in round 3 of the study. Round 3 was completed by 24 panel members. This round was mostly similar to round 2. In addition, the top-3 tests that were rated most frequently as suitable and were known by at least half of the panel were presented to the panel members. They were asked to rank these tests from most to least suitable.

3 Results

Thirty-two experts completed round 1 of the study (mean age = 43.44, SD = 11.20, 15 males, 17 females). Characteristics of the expert panel are displayed in Table 1. Despite repeated attempts (see paragraph 2.2) to gather an international expert panel, most experts were currently working/living in the Netherlands (n = 17). Seven of the experts were researchers, nine were clinicians, fifteen professionals integrated their therapeutic work with scientific research, and one was currently employed as manager. Of the original panel, twenty-four completed all three rounds of questionnaires and were included in the consortium.

TABLE 1 Characteristics of the initial expert panel ($N = 32$).

Demographics	
Gender, male/female *N*	15/17
Age M (SD)	43.4 (10.7)
Profession[a] *N* (%)	
Researcher	22 (68.8)
Clinician	24 (75.0)
Other (e.g., teaching, management)	8 (25.0)
Country, *N* (%)	
Netherlands	17 (53.1)
USA	5 (15.6)
Italy	5 (15.6)
India	2 (6.3)
UK	1 (3.1)
Australia	1 (3.1)
Sweden	1 (3.1)
Setting[a] *N* (%)	
University	13 (40.1)
Outpatient/ambulatory	12 (37.5)
(Forensic) hospital/inpatient	12 (37.5)
Diagnostics/Assessment	5 (15.6)
Research center	4 (12.5)
Prison/correctional facility	2 (6.25)
Assisted living	1 (3.1)
Academic rank[b] *N* (%)	
Professor	5 (15.6)
Associate professor	3 (9.4)
Assistant professor	4 (12.5)
Post-doctoral researcher	5 (15.6)
PhD candidate	2 (6.3)
Master of Science	3 (9.4)
Type of clinician[c] *N* (%)	
Clinical neuropsychologist	10 (31.3)
Neuropsychologist	5 (15.6)
Clinical psychologist	3 (9.4)
Psychiatrist	2 (6.3)
Neurologist	1 (3.1)
Other	3 (9.4)
Years of experience in the forensic setting[c] *N* (%)	
<4	5 (15.6)
4–7	7 (21.9)
8–11	5 (15.6)
12–15	2 (6.3)
16+	5 (16.1)

[a]Panel members could give multiple answers to this question. [b]This question was only answered for the panel members who were employed as researcher. [c]This question was only answered for the panel members who were employed as clinician.

In round 1, for each construct a panel member rated as at least moderately important (rating 3 or higher), the panel member was asked to suggest one or more tests to measure this construct. In total, 223 different tests were suggested by the panel.

In round 2, the panel rated these tests as "suitable", "not suitable" or "don't know". In round 3, we presented the panel with the three most-suitable tests (that were known by at least half of the panel) per construct, and we asked them whether they agreed with this top three. However, many did not fill in these questions. One explanation is that they did not know one or more of the tests, making it impossible to rank them. Another possibility is a decrease in motivation as the questionnaires were quite extensive and time consuming. Because of this, we based the top-3 tests in Table 3 on the suitability ratings from round 2. For certain constructs, fewer than three tests were familiar to at least half of the panel, resulting in less than three test suggestions (or even zero) being included in the overview.

To aid clinicians in their test selection, we included some practical information about the administration time, age range, manual, and psychometric properties of the tests. We derived this information from test manuals, systematic reviews, and books. If these were not available, we reported on single studies with a sample that was most similar to the aggressive offender population. Our goal was not to create an exhaustive and comprehensive overview, as it falls beyond the scope of this study. Therefore, we refer readers to the British Psychological Society test reviews using the EFPA review model (British Psychological Society, n.d.), the Buros Center for Testing (Buros Center for Testing, n.d.), or for Dutch readers the COTAN (NIP, n.d.) for more information about the psychometric properties of tests.

Below, we discuss the results per domain, sorted by importance-rating (see Table 2 and Figure 2). First, the importance ratings are discussed, including the reasoning provided by the panel members. Then, the test suggestions are discussed.

3.1 Negative valence systems

Negative Valence Systems were rated as the most important domain (M = 3.81, SD = 0.14), with a significant difference between reactive (M = 4.08, SD = 0.32) and proactive (M = 3.54, SD = 0.19) aggression ($U = 2$, $p = 0.009$). The ability to learn from one's own errors was suggested as an addition to this domain in round 1. Reaction to threat was rated as more important for reactive than for proactive aggression (acute: M = 4.50 vs. 3.24, potential: M = 4.31 vs. 3.44, sustained: M = 4.12 vs 3.68). The panel reasoned that as reactive aggression is driven by an immediate emotional reaction to a perceived threat or provocation, these constructs are more relevant in reactive aggression. Anxiety might make individuals more sensitive to perceived provocations, increasing the likelihood of aggression. Prolonged exposure to threat (sustained threat) might result in chronic stress and might cause individuals to use aggression to end the threat. Loss and being unable to achieve goals or experience rewards (frustrative non-reward) can lead to feelings of anger, sadness and disappointment. Aggression might be a way to cope with these feelings.

TABLE 2 Final ratings of the RDoC constructs.

	Reactive					Proactive				
RDoC constructs	*n*	Range	M	SD	IQR	*n*	Range	M	SD	IQR
Cognitive Systems			3.72	0.48	1			3.58	0.44	1
Attention	26	1–5	3.65	0.80	1	24	1–5	3.58	0.88	1
Perception	23	1–4	3.30	0.76	1	22	1–4	3.23	0.87	1
Visual	23	2–5	4.00	0.74	0	23	2–5	3.35	0.83	1
Auditory	23	1–5	3.74	0.96	1	22	1–5	3.05	0.90	0
Olfactory/somatosensory/multimodal	24	1–4	3.00	0.72	0	22	1–4	2.86	0.89	1
Declarative memory	25	1–4	2.80	0.71	0	24	1–5	3.13	0.85	1
Language	26	1–4	3.19	0.69	0	24	1–5	3.08	0.83	1
Cognitive control	31	2–5	4.45	0.72	1	31	2–5	4.16	0.93	1
Goal selection, updating	30	2–5	4.03	0.89	1	31	3–5	4.26	0.68	1
Representation, and maintenance response selection	31	3–5	4.68	0.60	1	25	3–5	4.28	0.54	1
Inhibition/suppression performance monitoring	30	3–5	4.23	0.68	1	30	2–5	4.17	0.79	1
Working memory	31	1–5	3.45	1.15	1	30	1–5	3.37	1.19	1
Active maintenance	27	1–5	3.59	1.08	1	24	2–5	3.67	0.70	1
Flexible updating	29	2–5	3.97	0.91	1	24	1–5	3.75	0.90	0
Capacity	25	1–5	3.64	0.86	0	28	1–5	3.50	1.11	1
Interference control	25	2–5	3.92	0.64	1	24	2–5	3.88	0.74	1
Counterfactual reasoning*	20	2–5	3.75	0.79	1	19	2–5	3.68	0.75	1
Information processing speed*	25	2–5	3.64	0.64	1	24	2–5	3.50	0.72	1
Arousal/regulatory			3.66	0.63	0			3.21	0.38	0
Arousal	32	1–5	4.38	1.10	1	24	2–5	3.63	0.65	1
Circadian rhythms	23	2–4	3.26	0.54	1	24	2–4	2.88	0.54	0
Sleep-wakefulness	29	1–5	3.34	1.20	1	23	2–4	3.13	0.76	1
Negative valence systems			4.08	0.32	1			3.54	0.19	0
Acute threat "fear"	32	1–5	4.50	0.84	1	25	1–5	3.24	0.83	1
Potential threat "anxiety"	32	3–5	4.31	0.69	1	32	1–5	3.44	1.05	1
Sustained threat	25	3–5	4.12	0.60	1	25	1–5	3.68	0.80	1
Loss	25	1–5	3.72	0.84	0	23	1–5	3.48	0.85	1
Frustrative nonreward	32	1–5	4.13	0.94	1	25	1–5	3.76	0.72	0
The ability to learn from one's own errors*	24	3–5	3.71	0.55	1	24	2–4	3.63	0.58	1
Positive valence systems			3.32	0.14	0			3.76	0.11	0
Reward responsiveness	23	2–4	3.26	0.62	1	22	2–4	3.64	0.58	1
Reward anticipation	28	1–5	3.46	1.10	1	21	3–4	3.76	0.44	1
Initial response to reward	28	1–5	3.39	1.13	1	23	2–5	3.70	0.63	1
Reward satiation	23	2–4	3.04	0.64	0	22	2–5	3.68	0.65	1
Reward learning	22	3–4	3.50	0.51	1	21	3–5	3.76	0.54	1
Probabilistic and reinforcement learning	21	2–4	3.48	0.60	1	22	3–5	3.91	0.43	0

(Continued)

TABLE 2 (Continued)

	Reactive					Proactive				
RDoC constructs	*n*	Range	M	SD	IQR	*n*	Range	M	SD	IQR
Reward prediction error	21	2–4	3.43	0.60	1	22	3–5	4.00	0.53	0
Habit - pvs	24	2–4	3.21	0.59	1	24	2–5	3.63	1.01	1
Reward valuation	22	2–4	3.18	0.73	1	21	3–4	3.86	0.36	0
Reward (probability)	23	2–4	3.35	0.65	1	23	2–5	3.78	0.60	1
Delay	22	2–4	3.27	0.63	1	21	2–5	3.71	0.72	1
Effort	20	2–4	3.25	0.64	1	20	2–5	3.70	0.73	1
Sensorimotor			3.25	0.48	1			3.10	0.51	1
Motor actions	28	1–5	3.21	1.03	1	22	2–4	2.72	0.55	1
Action planning and selection	22	2–4	3.18	0.59	1	22	2–5	3.45	0.80	1
Sensorimotor dynamics	22	1–4	2.86	0.56	0	24	1–5	2.94	0.85	0
Initiation	22	1–4	3.36	0.73	1	24	1–5	3.38	1.09	1
Execution	22	1–5	3.41	0.85	1	24	1–5	3.56	1.09	1
Inhibition and termination	20	2–5	4.40	0.82	1	24	2–5	4.00	0.82	1
Agency and ownership	27	1–5	3.48	1.01	1	21	1–5	3.24	0.94	1
Habit – sensorimotor	26	1–5	2.81	1.06	1	21	1–3	2.33	0.58	1
Innate motor patterns	26	1–5	2.69	1.16	1	20	1–4	2.60	0.75	1
Sensorimotor integration*	16	1–4	3.06	0.85	1	17	1–4	2.82	0.81	1
Social processes			3.90	0.35	1			3.70	0.37	1
Affiliation and attachment	23	1–5	3.74	0.81	1	31	2–5	4.19	0.83	1
Social communication	23	3–5	4.00	0.67	1	21	3–5	3.86	0.48	1
Reception of facial communication	29	3–5	4.38	0.68	1	22	3–4	3.68	0.48	1
Production of facial communication	23	3–4	3.57	0.51	1	21	2–4	3.33	0.66	1
Reception of non-facial communication	29	4–5	4.41	0.50	1	21	3–4	3.57	0.51	1
Production of non-facial communication	23	2–5	3.52	0.67	1	28	1–5	3.39	1.20	1
Perception and understanding of self	32	2–5	4.16	0.77	1	31	2–5	4.19	0.75	1
Agency	30	1–5	3.97	1.03	1	21	3–5	3.86	0.57	1
Self-knowledge	30	1–5	3.73	1.11	1	29	1–5	3.59	1.05	1
Perception and understanding of others	32	1–5	4.25	0.95	1	31	3–5	4.35	0.66	1
Animacy perception	23	2–5	3.87	0.55	0	27	1–5	3.41	1.12	1
Action perception	29	1–5	4.00	1.04	1	29	1–5	3.55	1.06	1
Understanding mental states	31	3–5	4.35	0.71	1	31	2–5	4.06	0.77	1
Ability to correctly understand the authenticity of others emotions*	22	3–5	4.05	0.49	0	19	2–5	3.79	0.71	1
Ability to understand absurdities*	23	2–5	3.13	0.69	0	22	1–4	2.91	0.61	0
Emotional contagion*	19	2–5	3.53	0.84	1	18	2–5	3.33	0.84	1
Moral reasoning*	23	2–5	3.87	0.97	1	21	1–5	3.90	1.00	1
Sympathy*	21	2–5	3.62	0.81	1	21	1–5	3.57	0.98	1
Other										
Intelligence/IQ*	24	2–5	3.54	0.83	1	24	2–5	3.75	0.79	1
Symptom/performance validity*	21	1–5	2.90	1.22	2	21	1–5	3.14	0.12	2
Cognitive distortions*	23	2–5	3.78	0.74	1	23	2–5	3.87	0.87	1
Emotion regulation*	23	2–5	4.39	0.84	1	23	2–5	3.74	1.01	2

*These constructs are not part of the RDoC, but were suggested as additions in round 1.

TABLE 3 Characteristics of the top-3 most suitable tests per construct, sorted by highest importance rating ($N = 25$).

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Cognitive control (R: 4.45, P: 4.16)						
Go/No Go task*	84%	84%	Computerized task that measures the ability to stop automatic reactions (impulse control).	Varies	18-65	Convergence with other types of self-control measures (Duckworth and Kern, 2011) • Executive functions: $r = 0.16$, N = 4855 • Delay tasks: $r = 0.12$, N = 523 • Self-report: $r = 0.11$, N = 1969 • Informant-report: $r = 0.15$, N = 1883
Wisconsin Card Sorting Task (WCST) (Heaton et al., 1993)*	84%	88%	Sorting cards based on changing rules and adapting. Measures ability to shift strategies, problem-solve, and assesses frontal lobe function.	20-30	5-89	Strauss et al. (2006) • Test-retest: generally low • Evidence of sensitivity to frontal damage: yes – but poor sensitivity and specificity • Evidence of ecological ability: yes
D-KEFS: Color Word Interference test (CWIT) (Delis et al., 2001)	80%	84%	Assesses cognitive functions like inhibiting automatic responses and shifting. Naming the ink color of words while inhibiting reading.	10	8-89	Internal consistency • Adequate (0.70–0.79) (Strauss et al., 2006), 0.75 (combined naming + reading) (Delis et al., 2007a) Test-retest • Marginal (0.60–0.69) to adequate (0.70–0.79) (Strauss et al., 2006) • Test-retest: 0.74 (cond. 1), 0.61 (cond. 2), 0.72 (cond. 3), 0.64 (cond. 4) (Delis et al., 2007a)
Perception and understanding of others (R: 4.25, P: 4.35)						
Faux Pas test (Stone et al., 1998; Gregory et al., 2002)	83%	83%	Measures detection social blunders in conversations through a number of stories.	15-20	18-65	(Söderstrand and Almkvist (2012) • Internal consistency: 0.905 • Split-half: 0.954 • Interrater: 0.916 • correlated significantly with the Eyes Test ($r = 0.302$, $p \leq 0.05$) and the Dewey Story Test ($r = -0.276$, $p \leq 0.05$) Osterhaus and Bosacki (2022) • Internal consistency: reported by 3 studies (0.78, 0.81, 0.91; M = 0.83, SD = 0.07)
Emotion Recognition Task (ERT) (Montagne et al., 2007)*	78%	78%	Computer-generated paradigm designed to assess the recognition of 6 basic facial emotional expressions: anger, disgust, fear, happiness, sadness, and surprise.	6-10	8-75	• Studies demonstrated the validity in a wide range of patient groups, often showing impairments that are selective for specific emotions (see Kessels et al., 2014 for an overview)
Strange stories (Happé, 1994)	70%	74%	Gauges social understanding. Participants interpret social scenarios, assessing comprehension of subtle social cues.	30-45 (original version) 15-20 (SS-R)	Developed for children, but can also be used for adults	Osterhaus and Bosacki (2022) • Internal consistency: reported by 5 studies (0.67, 0.69, 0.69, 0.73, and 0.79; M = 0.71, SD = 0.04)
Perception and understanding of self (R: 4.16, P: 4.19)						

(Continued)

10.3389/fpsyg.2024.1328839

TABLE 3 (Continued)

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Thematic Apperception Test (TAT) (Morgan and Murray, 1935)	52%	65%	Creating stories based on ambiguous pictures, revealing inner perceptions and imagination.	2 sessions of 50 min.	5+	Hilsenroth and Segal (2004) • Interrater reliability: 0.80−0.86
Emotion Recognition Task (ERT)	48%	70%	See above			
Faux pas test	43%	74%	See above			
Strange stories test	43%	70%	See above			
Arousal (R: 4.38, P: 3.63)						
Go/No Go task	54%	83%	See above			
Affiliation and attachment (R: 3.74, P: 4.19)						
Faux Pas test	74%	78%	See above			
Strange Stories Test	65%	74%	See above			
Reading the Mind in the Eyes Test (RMET or Eyes Test) (Baron-Cohen et al., 2001)*	61%	74%	Participants infer emotions and thoughts from images of eyes, gauging social cognition.	4	16+	Osterhaus and Bosacki (2022) • Internal consistency: reported by 6 studies (0.41, 0.53, 0.61, 0.62, 0.75. 0.82; M = 0.62, SD = 0.15)
Frustrative non-reward (R: 4.13, P: 3.67)						
Iowa Gambling Task (IGT) (Bechara, 2016)*	78%	83%	Assesses decision making through a card game where participants choose cards, learning to avoid risky options for long-term gains.	15-20 minutes to administer and score	8-79	• Split-half and test-retest: not testable (Lezak, 2012) • Low correlations with self-reported risk taking and personality traits related to risk-taking (Schmitz et al., 2020)
Social communication (R: 4.00, P: 3.86)						
Emotion Recognition Task (ERT)	78%	83%	See above			
Facial Expressions of Emotion – Stimuli and Tests (FEEST) (Young et al., 2002)*	78%	78%	Inferring emotions and thoughts from images of eyes, testing emotional perception and recognition. The FEEST is a combination of the Ekman 60 Faces Test and the Emotion Hexagon Test.	25-30	18+	Short version (Kuhlmann and Margraf, 2023) • Cronbach's α was on average 0.70 for prototype and 0.67 for morphed stimuli • Test-retest reliability: 0.60 for prototype and 0.62 for morphed stimuli Young et al. (2002) • Ekman 60 faces – split half: 0.62 (total score), 0.62 (anger), 0.66 (disgust), 0.53 (fear), 0.21 (happiness), 0.60 (sadness), 0.61 (surprise) • Emotion Hexagon – split half: 0.92 (total score), 0.68 (anger), 0.92 (disgust), 0.88 (fear), 0.18 (happiness), 0.65 (sadness), 0.33 (surprise) • Correlation between Ekman 60 faces and Emotion Hexagon: 0.68 (total score), 0.51 (anger), 0.27 (fear), 0.52 (disgust), −0.05 (happiness), 0.54 (sadness), 0.42 (surprise)

(Continued)

10.3389/fpsyg.2024.1328839

TABLE 3 (Continued)

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Faux pas test	74%	83%	See above			
Reading the Mind in the Eyes Test (RMET or Eyes Test)	74%	74%	See above			
Sustained threat (R: 4.12, P: 3.68)						
Reading the Mind in the Eyes Test (RMET or Eyes Test)	43%	52%	See above			
Potential threat ("Anxiety"; R: 4.31, P: 3.44)						
Emotion Recognition Task (ERT)	87%	91%	See above			
Affective/Emotional Go/No-Go Task*	78%	83%	Participants respond to emotional and neutral stimuli, measuring impulse regulation and emotional control.	Varies	Varies	Correlations between commission errors across the emotional and non-emotional tasks: 0.51-0.56, supporting the construct validity of behavioral inhibition (Schulz et al., 2007)
Facial Expressions of Emotion – Stimuli and Tests (FEEST)	65%	65%	See above			
Acute Threat ("Fear"; R: 4.50, P: 3.24)						
Emotion Recognition Task (ERT)	78%	83%	See above			
Affective/Emotional Go/No-Go Task	74%	78%	See above			
Reward learning (R: 3.50, P: 3.76)						
Iowa Gambling Task (IGT)	100%	100%	See above			
Wisconsin Card Sorting Task (WCST)	70%	87%	See above			
Tower of London (TOL)*	39%	78%	Participants move disks on pegs, aiming to recreate a specific tower arrangement, assessing strategic thinking, planning and problem solving.	10-15	7-80	Köstering et al. (2015) • Across samples, mean split-half and lower bound indices of reliability of accuracy scores were adequate ($r \geq 0.7$) or higher, with the lower-bound estimate uniformly indicating high reliability ($glb \geq 0.8$) • TOL-F planning accuracy possesses adequate criterion-related concurrent validity Humes et al. (1997) • Correlation with TOH: 0.37
Tower of Hanoi (TOH)*	39%	78%	Similar to TOL, except disks now vary in size (making the task more difficult)	15-20	7-80	Humes et al. (1997) • Correlation with TOL: 0.37 • Batista et al. (2007): Internal consistency 0.37 (original version) 0.40-0.77 (revised version)

(Continued)

10.3389/fpsyg.2024.1328839

TABLE 3 (Continued)

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Attention (R: 3.65, P: 3.58)						
D-KEFS: Trail Making Test (TMT) (Delis et al., 2001)	80%	88%	Connecting numbered circles while alternating between numbers and letters, evaluating cognitive flexibility, visual attention and attention shifting.	15-20	8-89	• Internal consistency 0.57 to 0.81 (Shunk et al., 2006); Low (≤ 0.59) (conditions 1-4) to Adequate (0.70–0.79) (condition 5) (Strauss et al., 2006) 0.72 (Combined Number + Letter Sequencing) (Delis et al., 2007c) Test-retest • Marginal (0.60–0.69) (Combined Number + Letter Sequencing) and Adequate (0.70–0.79) (motor speed and condition 5) (Strauss et al., 2006) 0.56 (cond. 1), 0.57 (cond. 2), 0.59 (cond. 3), 0.37 (cond. 4), 0.77 (cond. 5), 0.66 (combination) (Delis et al., 2007c)
WAIS: Digit Symbol Coding (Wechsler, 2012)	80%	92%	Matching symbols to numbers as quickly as possible, testing processing speed and sustained attention.	5	16-90	Test-retest • 0.86 (Pearson NL, 2012) • High (0.80–0.89) (Strauss et al., 2006) Internal reliability • High (0.80–0.89) (Strauss et al., 2006)
WAIS: Symbol Search (Wechsler, 2012)	80%	92%	Scanning sets of symbols, identifying target symbol presence or absence, assessing visual attention and processing speed.	5	16-90	Test-retest: • 0.75 (Pearson NL, 2012) • High (0.80–0.89) (Strauss et al., 2006) Internal reliability • Adequate (0.70–0.79) (Strauss et al., 2006)
Loss (R: 3.72, P: 3.48)						
Iowa Gambling Task (IGT)	42%	75%	See above			
Cambridge Gambling Task (CANTAB; (Cambridge Cognition, 2012)*	25%	50%	Participants choose between options to win or lose money, evaluating risk-taking, decision-making and reward-seeking behavior.	12-18	18+	Has been shown to be sensitive to impairment in gambling addition (Lawrence et al., 2009) and substance use disorder (Rogers, 1999).
Reward Valuation (R: 3.18, P: 3.86)						
Iowa Gambling Task (IGT)	96%	100%	See above			
Wisconsin Card Sorting Task (WCST)	57%	87%	See above			
Reward Responsiveness (R: 3.26, P: 3.64)						
Iowa Gambling Task (IGT)	96%	96%	See above			
Go/No-Go task	57%	65%	See above			

(Continued)

10.3389/fpsyg.2024.1328839

TABLE 3 (Continued)

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Stop Signal Task (SST) (Logan, 1994)*	43%	52%	Participants quickly respond to a visual or auditory signal but stop when a "stop" signal appears, assessing the ability to inhibit automatic responses.	Varies	Varies	Convergence with other types of self-control measures (Duckworth and Kern, 2011) • Executive functions: $r = 0.11$, N = 1982 • Delay tasks: $r = 0.17$, N = 189 • Self-report: $r = 0.17$, N = 402 • Informant-report: $r = 0.13$, N = 506
Working Memory (R: 3.45, P: 3.37)						
WAIS: Digit Span (Wechsler, 2012)	88%	92%	Participants repeat a series of numbers in the same order (forward) or reverse order (backward), assessing short-term/working memory capacity.	5	16-90	Test-retest: • 0.82 (Pearson NL, 2012) • Test-retest: High (0.80–0.89) (total), Adequate (0.70–0.79) (forward and backward) (Strauss et al., 2006) Split-half • 0.91 (Pearson NL, 2012) Internal reliability • High (0.80–0.89) (total, forward and backward) (Strauss et al., 2006)
WAIS: Letter Number Sequencing (Wechsler, 2012)	80%	88%	Participants listen to a sequence of numbers and letters, then repeat the numbers in ascending order followed by the letters in alphabetical order, evaluating working memory and attention.	5	16-90	Split-half • 0.81 (Pearson NL, 2012) Test-retest • 0.78 (Pearson NL, 2012) • High (0.80–0.89) (Strauss et al., 2006) Internal reliability • Very high (0.90+) (Strauss et al., 2006)
WMS-III: Spatial Span (Wechsler, 1997)	72%	88%	Participants recreate a sequence of blocks tapped by the examiner in the same order, testing visuospatial working memory.	5	16-90	Strauss et al. (2006) • Internal consistency: Adequate (0.70 to 0.79) • Generalizability coefficients: Adequate (0.70 to 0.79) • Test-retest: Adequate (0.70 to 0.79)
Agency and Ownership (R: 3.48, P: 3.24)						
Tower of London (TOL)	54%	79%	See above			
Tower of Hanoi (TOH)	50%	75%	See above			
D-KEFS: Tower Test (Delis et al., 2001)	42%	67%	measures executive functioning and planning abilities by assessing their capacity to rearrange a set of colored disks on pegs to match a target configuration while adhering to specific rules	15-20	8-89	Internal consistency • Marginal (0.60–0.69) (total achievement) (Strauss et al., 2006) 0.64 (Delis et al., 2007b) Test-retest • Low (≤0.59) (total achievement) (Strauss et al., 2006) • Test-retest: 0.44 (Delis et al., 2007b)

(Continued)

10.3389/fpsyg.2024.1328839

TABLE 3 (Continued)

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Perception (R: 3.30, P: 3.23)						
Rey Complex Figure Test (RCFT) (Meyers and Meyers, 1995)	68%	88%	Reproducing a complex figure from memory. Measures Visuospatial memory and organizational skills.	10-15 (excl. delays)	6-93	Strauss et al. (2006) • Test-retest: adequate to high for intervals of 6 months or less • Practice effects: yes
WAIS: Block Design (Wechsler, 2012)	56%	84%	Participants arrange blocks to match a given design as quickly as possible, assessing spatial reasoning, visual-motor skills. perceptual organization.	10-15	16-90	• Split-half: 0.84 (Pearson NL, 2012) • Internal reliability: High (0.80–0.89) (Strauss et al., 2006) • Test-retest: Adequate (0.70–0.79) to High (0.80–0.89) (Strauss et al., 2006)
Judgment of Line Orientation (JLO) (Benton et al., 1978)	52%	68%	Participants match lines in a diagram to angles in another diagram, evaluating spatial orientation and visual perception.	15-20	7-96	• Test-retest: high (Strauss et al., 2006)
Sleep/Wakefulness (R: 3.34, P: 3.13): No tests were suggested						
Language (R: 3.19, P: 3.08)						
Phonological fluency tests	83%	92%	Generating words starting with a specific letter	5	2-95 (depends on version)	Strauss et al. (2006) • Test-retest: adequate • Evidence of sensitivity to frontal damage: yes, but poor sensitivity and specificity • Evidence of ecological ability: yes
Semantic fluency tests	83%	92%	Generating words within a specific semantic category	5	2-95 (depends on version)	Strauss et al. (2006) • Test-retest: adequate • Evidence of sensitivity to frontal damage: yes, but poor sensitivity and specificity • Evidence of ecological ability: yes
WAIS: Comprehension (Wechsler, 2012)	83%	92%	Answering questions about social situations. Measures verbal comprehension and social knowledge	5-10	16-90	Pearson NL (2012) • Split-half: 0.84 • Test-retest: 0.78
Circadian Rhythms (R: 3.26, P: 2.88): No tests were suggested						
Declarative Memory (R: 2.80, P: 3.13)						
15 Words Test (15WT)/RAVLT	76%	80%	Recalling a list of words immediately after hearing. Measures verbal memory.	10-15	6-97	Strauss et al. (2006) • Test-retest: marginal to adequate for total recall, trail 5 and delayed recall trails • Practice effects: yes

(Continued)

10.3389/fpsyg.2024.1328839

TABLE 3 (Continued)

Construct (importance rating: reactive, proactive) tests (authors)	Suitable (%)	Known by (%)	Short description	Administration time (min)	Age	Psychometric properties
Californian Verbal Learning Test (CVLT)	76%	76%	Memorizing and recalling a list of words over time. Measures verbal memory and learning over trials.	30 min, plus 30-min delay	16-90	Strauss et al. (2006) • Test-retest: high for scores of level of performance, low for scores of process/strategy • Practice effects: yes
Rey Complex Figure Test (RCFT)	68%	88%	See above			
Motor Actions (R: 3.21, P: 2.72)						
Go/No Go task	78%	83%	See above			
Trail Making Task (TMT)	65%	83%	Connecting numbers and letters in sequential order. Measures cognitive flexibility and visual attention.	5-10	9-89	Strauss et al. (2006) • Internal reliability: N/A • Test-retest: for the most part adequate • Evidence of sensitivity to attentional impairments: good • Evidence of ecological validity: good
Finger Tapping Test	61%	70%	Measures motor speed and coordination by tapping a finger as quickly as possible.	<10	5-85	Strauss et al. (2006) • Test-retest: variable (low to high)
Innate Motor Patterns (R: 2.69, P: 2.72)						
Finger Tapping Test	43%	57%	See above			
Tower of Hanoi (TOH)	39%	61%	See above			
Tower of London (TOL)	39%	61%	See above			
Habit - Sensorimotor (R: 2.81, P: 2.33)						
WAIS: Digit Symbol Coding	50%	75%	See above			
Conner's Continuous Performance test (CCPT-II)*	46%	63%	Responding to specific target stimuli while ignoring distractions. Measures attention and impulse control.	14	6-55+	Strauss et al. (2006) • Internal reliability: acceptable to high • Test-retest: limited • Evidence of sensitivity to attentional impairments: moderate • Evidence of ecological validity: limited information
Stroop Test (Golden and Freshwater, 2002)	43%	70%	Naming the ink color of words while ignoring their meaning. Measures cognitive control and inhibition.	5	5-90 (depends on version)	Strauss et al. (2006) • Test-retest: adequate for the interference trial • Evidence of sensitivity to frontal damage: yes, but poor sensitivity and specificity • Evidence of ecological ability: yes

Note. * = digital version available; % suitable = percentage of the experts that rated the test as suitable to measure the given construct; % known = percentage of the experts that knew the test well enough to rate it's suitability. Only tests that were known by 50% of the panel were included in the table.

10.3389/fpsyg.2024.1328839

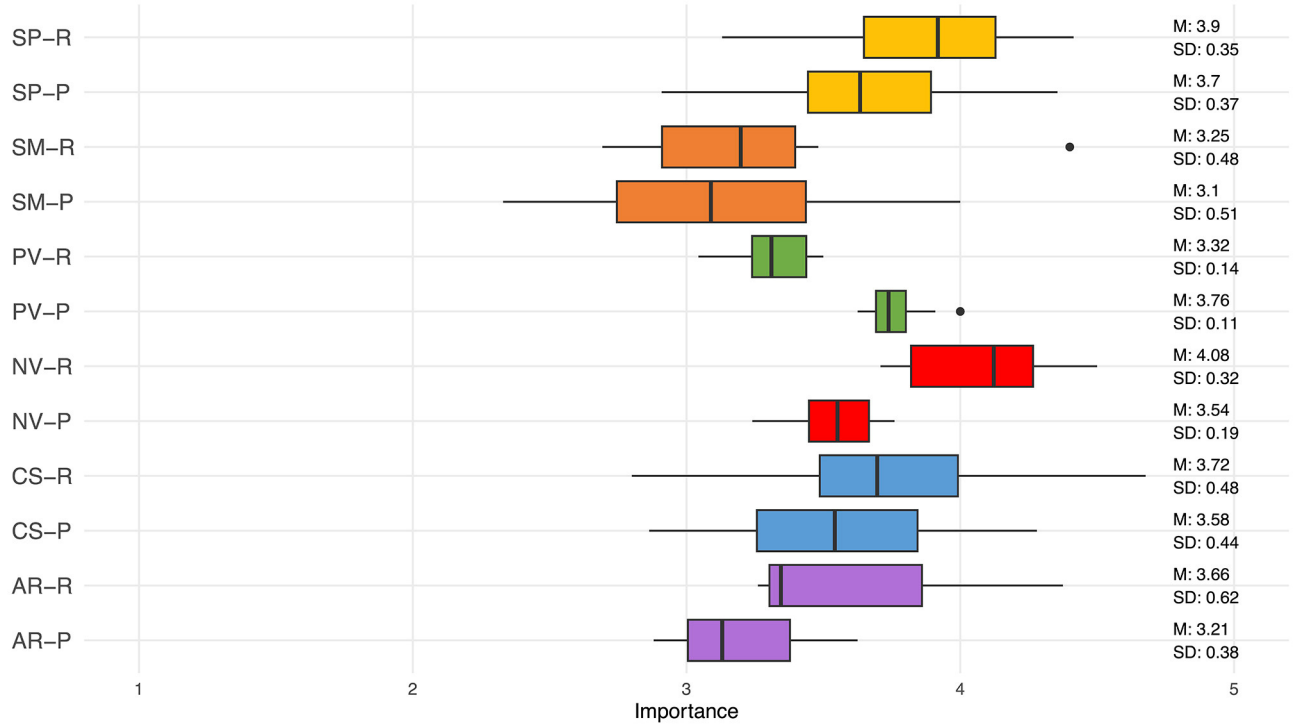


FIGURE 2
Boxplots of importance scores of the RDoC domains for reactive (R) and proactive (P) aggression. SP, Social Processes; SM, Sensorimotor Systems; PV, Positive Valence Systems; NV, Negative Valence Systems; CS, Cognitive Systems; AR, Arousal/Regulatory Systems.

In total, 22 neuropsychological tests were suggested by the panel to assess Negative Valence Systems. The panel commented that it might be better to assess this domain by including biological measures (heart rate, eye tracking/pupil size, skin conductance) or self-report. For frustration, observation from potentially frustrating tests was also suggested, however, it was noted that test observations should not be confused with objective test results. Frustration from not performing the test correctly is not the intended measurement of the test and therefore, not an objective test result. For some Negative Valence constructs, it was impossible to validly assemble a top-3 as many tests that were suggested were unknown by more than half of the panel. Therefore, Acute Threat and Loss have only two tests in the overview and Sustained threat only one.

3.2 Social processes

The domain Social Processes was rated as the second most important overall (M = 3.80, SD = 0.33), with a nonsignificant difference between reactive (M = 3.90, SD = 0.35) and proactive (M = 3.70, SD = 0.37) and aggression (U = 190.5, p = 0.097). Four additional constructs were added to this domain: sympathy, moral reasoning, ability to correctly understand the authenticity of others' emotions, and emotional contagion. The panel commented that the ability to interpret social cues (including other people's emotions and social ambiguity), is crucial in understanding and preventing aggression. Misinterpretations can increase the risk of both reactive and proactive aggression. However, proactive aggression may be less influenced by this as people with high callous-unemotional traits tend to be less concerned with other people's emotions. Another important aspect within this domain is empathy. High callous unemotional traits in individuals displaying proactive aggression often involves cognitive empathy without affective empathy, enabling manipulative behavior.

It was noted by the panel that it might be more feasible to assess Social Processes with interviews, questionnaires and observations instead of neuropsychological tests. In total, 34 tests were suggested for this domain, resulting in a top-3 tests for each construct.

3.3 Cognitive systems

Next, Cognitive Systems were rated as M = 3.65 (SD = 0.44), with a non-significant difference between reactive (M = 3.72, SD = 0.48) and proactive aggression (M = 3.58, SD = 0.44; U = 134.5, p = 0.389). The panel reflected on why the Cognitive Systems are (not) important to consider in aggressive patients. Working memory is needed to process and react to triggers (reactive aggression), but on the other hand, working memory is required to plan proactive aggressive behaviors. Attention was deemed important in reactive aggression, as it can be biased toward potential threats, while ignoring neutral or friendly information. Cognitive control is linked to inhibition, which can help prevent future (especially reactive) aggression. In addition, cognitive control is important to be able to find non-aggressive solutions to problems, and to apply lessons from therapy into daily life (also related to declarative memory). Language was considered important in assessing aggression because poor verbal skills can

hinder the ability to find non-aggressive solutions in conflicts, potentially leading to misunderstandings and frustration. The role of perception was somewhat unclear among the panel members. Counterfactual reasoning and information processing speed were added as additions to this domain.

For Cognitive Systems, a large number (154) of different tests was suggested. The top-3 tests per construct are displayed in Table 3.

3.4 Positive valence systems

Positive Valence Systems were rated with M = 3.54 (SD = 0.11). Interestingly, this was the only domain that was deemed significantly more important for proactive (M = 3.76, SD = 0.11) than for reactive aggression (M = 3.32, SD = 0.14; $U = 144$, $p < 0.001$). The panel members reasoned that individuals with high reward responsiveness may be more motivated to engage in proactive aggression, driven by the pursuit of rewards and experiencing greater pleasure and motivation when such rewards are at stake. Reward learning is important for proactive aggression, as individuals who have learned that aggressive behavior leads to desired outcomes are more likely to repeat such behavior to achieve their goals. Lastly, the value of potential rewards shape proactive aggression, with those highly valuing rewards associated with aggression, like financial gain or social status, being more inclined to engage in this form of aggression. Reactive aggression is more indirectly related to reward as the alleviation of distress or protection can be considered the reward in this context.

The experts suggested 14 different tests to assess Positive Valence Systems. The top-3 tests are displayed in Table 3. For Reward Valuation, there were only two tests known by half of the panel. It was noted by the panel that many of these tests do not directly measure reactions to rewards, but this can be inferred through observation.

3.5 Arousal/regulatory systems

Arousal/Regulatory Systems were rated with M = 3.44 (SD = 0.50) importance overall, with non-significant difference between reactive (M = 3.66, SD = 0.63) and proactive (M = 3.21, SD = 0.38) aggression ($U = 2$, $p = 0.400$). The construct Arousal was considered very important in reactive aggression (M = 4.38), where impulsive and emotionally charged responses are common. Conversely, in proactive aggression (M = 3.63), the issue often revolves around the absence of arousal or under-arousal, suggesting a potential opposite relationship. It was highlighted that arousal is a state rather than a trait and is subject to rapid fluctuations influenced by environmental factors that can be challenging to measure. Disturbances in sleep and circadian rhythms could have consequences on daily mood patterns, possibly affecting emotional regulation and impulsivity.

The panel noted the absence of tests to measure arousal. Instead, they proposed physiological measures (such as EEG, heart rate variability, skin conductance, pupil dilation) behavioral/observational methods (such as wearables, questionnaires, or diaries), and neuroimaging. The seven tests that were suggested by the panel are often developed to measure different constructs such as motor skills, attention, and inhibition, and were all—except for the go/no-go task—unknown by half of the panel. Therefore, no top-3 could be validly constructed.

3.6 Sensorimotor systems

The domain with the lowest importance rating was Sensorimotor Systems (M = 3.18, SD = 0.48), with non-significant difference between reactive (M = 3.25, SD = 0.48) and proactive aggression (M = 3.10, SD = 0.51; U = 45, $p = 739$). The construct "sensorimotor integration" was suggested as an addition to this domain. The panel reasoned that sensorimotor systems might be relevant in understanding reactive aggression, which can be impulsive and driven by limbic responses, particularly in individuals with trauma or dissociation. These motor reactions can lead to a loss of agency and ownership over actions, potentially becoming self-fulfilling. Automatic aggressive behaviors learned from early experiences may be tied to sensorimotor patterns, particularly in reactive aggression. However, there's debate over whether these constructs can be clinically measured and if they directly correlate with quantifiable aggression.

Nevertheless, the panel suggested 39 different tests to measure Sensorimotor Systems. These tests encompass a wide range from executive functioning/planning tests (e.g., tower tests) to tests that more directly measure motor skills and coordination. A panel member proposed the idea of using advanced technology like movement sensors and virtual reality to understand how people physically react to challenging situations.

3.7 Additional suggestions

Lastly, the panel suggested four constructs that do not fit within the RDoC domains but might be worth considering when assessing aggressive offenders. For intelligence, the panel agreed (IQr = 1) that this is moderately to very important to include (intelligence: reactive 3.54, proactive: 3.75). It was noted that general intelligence might not provide additional information beyond the specific cognitive functions already encompassed within the model or if these specific functions might completely explain the association between intelligence and aggression. Secondly, cognitive distortions—which are biased or irrational patterns of thoughts and perception that can influence a person's beliefs, attitudes, and behaviors—were deemed moderately to very important (IQr = 1, reactive: 3.78, proactive: 3.87). A panel member noted that cognitive distortions are influenced by inner psychological patterns or past traumas and can cause a person to misinterpret what's going on, making them more likely to engage in violent behavior. Third, emotion regulation was rated as essential for reactive aggression (4.39, IQr = 1), but there was no consensus for proactive aggression (3.74, IQr = 2). Lastly, symptom/performance validity was added as a suggestion, but the panel did not reach consensus on this construct (reactive: 2.90, proactive: 3.14, IQr = 2). The panel members commented that

the addition of symptom/performance validity tests is valuable for detecting feigned or exaggerated symptoms and can help to ensure that decision about risk assessment/management and legal decisions are based accurate information. However, these type of tests are less directly related to understanding the origins of aggressive/offending behavior *per se*.

4 Discussion

In this Delphi study, we investigated two questions by surveying an international expert panel. Firstly, we sought theoretical insights into the constructs commonly associated with aggression, emphasizing their importance in the evaluation of predominantly reactive vs. predominantly proactive aggressive offenders. Secondly, we aimed to pinpoint the most suitable tests for this assessment, thereby facilitating future test selection in forensic contexts.

4.1 RDoC constructs

Overall, all RDoC domains were considered at least moderately important (>3) by the expert panel for the neuropsychological assessment of aggressive offenders. Taken together, Social Processes and Negative Valence Systems were rated as the most important in understanding aggression, while Sensorimotor Systems were considered least important. These findings are in line with studies that found a relation between aggression and executive functions and attention (Bergvall et al., 2001; Ogilvie et al., 2011; Burgess, 2020; Cruz et al., 2020), language (Cohen et al., 2003; Anderson et al., 2016; Chow et al., 2022), social cognition (Karoglu et al., 2022), and reward and threat processing (Estrada et al., 2019; Manning, 2020). Below, we will further explore the importance of the RDoC constructs considering the distinction between reactive and proactive aggression.

4.2 Reactive vs. proactive aggression

The extent to which experts differed in their opinion about the theoretical importance of the RDoC constructs for understanding reactive aggression compared to proactive aggression was rather small for most domains. The most pronounced difference was that Positive Valence Systems were deemed more important to understand proactive aggression, whereas Negative Valence Systems were considered most relevant for understanding reactive aggression. Both come as no surprise based on previous research. Differences in reward processing are found in children and adults with conduct disorder, callous unemotional treats, antisocial personality disorder and psychopathy (Estrada et al., 2019). As these diagnoses are generally related to proactive aggression (Merk et al., 2005; Cima and Raine, 2009; White et al., 2015; Zhang et al., 2017), this outcome fits well into what we know. In addition, studies have shown that in people with impulsive-antisocial traits linked to psychopathy, their brains released more dopamine in the nucleus accumbens when exposed to rewards, suggesting an hyperreactivity to rewards (Buckholtz et al., 2010). This highlights the relevance of trying to unravel the antecedents of aggression for assessment and treatment. Reactive aggression on the other hand is a primary reaction to perceived treat.

For the other domains, the difference in perceived importance between reactive and proactive aggression were rather small. This may be explained by the fact that some RDoC constructs, such as arousal, are quite broad. It has been reported in empirical studies that reactive aggression involves high affective-physiological arousal while proactive aggression is characterized by minimal autonomic arousal (Chase et al., 2001; Blair, 2003). In other words, arousal might be important in both types of aggression, albeit in different ways. Another example: compromised working memory might be associated with increased reactive aggression, as it is needed to process and react to triggers, while in proactive aggression, working memory is required for planning acts of violence, making it equally important but in a different manner. In other words, while RDoC constructs are important to evaluate to gain insights into the determinants of both forms of aggression, they may play different roles in the two types of aggression.

4.3 Expert recommendations for neuropsychological test usage

In total, 223 different tests were suggested by the panel. This indicates that an large number of neuropsychological tests have been developed in the past decades and attests to the field's rapid development. It also presents a challenge for clinicians in choosing the most suitable tests. In addition, aggression is a multifaceted construct that cannot be measured through a single test. The distinction between reactive and proactive aggression adds another layer of complexity. In response to these challenges, we constructed a guide for clinicians and researchers, a curated selection of the three most favored tests as assessed by our panel of experts.

It must be noted that our aim was to provide an overview that offers a selection of the most suitable tests to measure the RDoC constructs, rather than constructing a fixed battery of tests. By presenting an overview of the most important neuropsychological constructs along with the most suitable tests to measure them, clinicians and researchers can select specific constructs that are most relevant to their case. However, for certain subgroups, particularly when assessing patients with intellectual disabilities or patients who are illiterate, the tests suggested in our study might not be suitable. In those cases, clinicians are encouraged to seek for alternative tests. In the case of intellectual disabilities, it is proposed to use adapted versions of the original tests (such as the children's version) (Willner et al., 2010). In the case of illiteracy, the suggestion is to modify tests to resemble real-life situations instead of school-based procedures (Kosmidis, 2018). It is noteworthy that both intellectual disabilities and illiteracy more prevalent in forensic populations than in the general population (Harris, 2006; Tuominen et al., 2014; Hellenbach et al., 2017; Muñoz García-Largo et al., 2020), underscoring the importance of considering these factors in the selection of appropriate assessment tools. In addition, it is important to note that some of the tests that emerged from our study are subject to criticism, often in absence of better alternatives (e.g., the Thematic Apperception Test, see Lilienfeld et al., 2000). It is beyond the scope of this study to address tests individually.

Assessing these constructs may help to explain the determinants of the aggressive behavior which can provide valuable input for tailored treatment planning. Another important outcome is that the panel indicated that not every RDoC construct is appropriate to be measured by neuropsychological testing. For example, it was noted that the construct affiliation/attachment can be more effectively assessed through a structured interview, and arousal through observation or physiological measures. The RDoC matrix provides numerous examples of self-report and physiological measures for assessing its constructs (National Institute of Mental Health, 2023). Hence, a combination of neuropsychological tests, interviews, self-report, observation, and physiological measures might be needed to optimally measure the RDoC constructs.

4.4 Limitations

The findings of this Delphi study need to be considered in the context of a few limitations. Firstly, despite repeated and extensive attempts to include a representative global panel, half of the panel consisted of people from the Netherlands. The continents of South America and Africa were not represented at all and other continents were underrepresented (especially taking the number of inhabitants into account). Since neuropsychological practices are affected by, for example, the country's health care system, legal framework, and cultural norms, this is likely to have influenced the results of the study (Kasten et al., 2021). This may have also limited generalizability as certain recommendations might be more tailored to the Netherlands.

Furthermore, while every effort was made to ensure conciseness of the questionnaires, it is essential to recognize that participant motivation can influence the quality and consistency of expert input in iterative research endeavors like the Delphi method. Eight panel members (25% of the original panel, of which three from the Netherlands, 2 USA, 1 Italy, 1 Sweden, 1 Australia) did not complete all three rounds. This may have had implications for representativeness of the panel as they might have had different perspectives than the remaining 24 experts. Fortunately, most information was gathered in round 1 where experts rated all RDoC constructs and provided their test suggestions.

The panel generated a large number (223) of neuropsychological tests that can be used to measure the RDoC constructs. The panel was unfamiliar with many of the tests (56% of the tests were unknown to more than half of the panel), which prevented them from forming an opinion about their suitability. As a consequence, we could not validly construct a top-3 for each RDoC construct. For these constructs, we refer readers to the RDoC matrix for other assessment suggestions (National Institute of Mental Health, 2023).

Other limitations stem from the Delphi methodology. The approach toward consensus may exclude different but possibly important perspectives of individual panel members. The results of a Delphi study represent the ratings with the most overlap between the panel members, but this is not necessarily the "objective truth". Our study wasn't designed to uncover objective truths; instead, we aimed to identify best practices. Moreover, the Delphi procedure precludes direct contact between panel members to avoid group pressure toward conformity and possible effects of authority. However, a discussion can often lead to valuable insights. To address this, the panel members could read each other's comments and reasonings anonymously in round 2 and 3. This could help them in understanding the source of potential discrepancies in ratings and possibly change their opinion. We highlighted that they were not obliged to change their ratings if their opinion had not been changed.

4.5 Implications and future directions

While this study represents a significant step forward in the endeavor to achieve adequate neuropsychological assessment of aggressive offenders, it is essential to acknowledge that our understanding of the relationship between the RDoC domains and aggression remains complex. Studying the interrelations between the constructs might provide more insights into aggressive behavior. For example, a lack of attention might lead to misinterpretation of social cues and a compromised working memory can lead to difficulties in emotion regulation.

Another prominent challenge that emerges from our study is the validation of the neuropsychological tests proposed by the expert panel. Moreover, to ensure that neuropsychological assessments are meaningful and sensitive to the unique characteristics of aggressive offenders, the field should focus on collecting more appropriate normative data.

Furthermore, the possible incorporation of neuropsychological test findings into risk assessment and management should be studied more thoroughly. This approach aligns with the Risk-Need-Responsivity (RNR) model (Bonta and Andrews, 2023), a leading framework in the forensic field. The findings have two connections to the RNR model. First, previous studies have indicated the added value of including biopsychosocial factors for the prediction of recidivism (Aharoni et al., 2013, 2014; Haarsma et al., 2020; Zijlmans et al., 2021). This aligns with the 'Need' principle of the RNR model, which emphasizes the importance of targeting criminogenic needs that are associated with an individual's likelihood of reoffending. Second, beyond understanding the cognitive limitations associated with aggression, future research should explore how this knowledge can be translated into effective intervention strategies. Specifically, cognitive limitations may play a crucial role in an individual's responsiveness to treatment, adhering to the 'Responsivity' principle of the RNR model. One might expect that if offenders have attentional difficulties or memory problems, that will influence treatment effectiveness. Longitudinal studies can help to understand how changes in neuropsychological function are related to changes in aggression and recidivism, further strengthening the connection between the RNR principles and the incorporation of neuropsychological assessments in risk assessment and -management strategies.

4.6 Conclusion

This Delphi consensus study shed light on the role of the RDoC framework in understanding and assessing aggression in offenders. The experts' ratings underline the multidimensional

nature of aggression, calling for a holistic approach when assessing and addressing aggression. Furthermore, distinguishing between reactive and proactive aggression provides useful insights into the mechanisms involved in aggressive behavior.

The extensive list of proposed neuropsychological tests, as well as the construction of a top-3 list for each construct, provide clinicians and researchers with a useful resource when it comes to selecting suitable tests. This overview allows for a flexible approach by tailoring assessments to specific clinical or research requirements. Furthermore, the acknowledgment that certain constructs may be better examined through interviews, observations, or physiological measures emphasizes the added value of a multimodal assessment strategy.

Future research should focus on test validation, normative data collecting, and the integration of neuropsychological findings into risk assessment and intervention as our understanding of the complex relationship between RDoC domains and aggression advances. Our Delphi consensus study not only enhances our comprehension of aggression in offenders through the application of the RDoC framework but also provides a comprehensive guide for clinicians and researchers in the selection of neuropsychological tests. The findings of this Delphi study offer a steppingstone for advancing the field of neuropsychological assessment in understanding and addressing aggressive behavior.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethics Review Board (ERB) of the University of Amsterdam. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JCH: Conceptualization, Formal analysis, Investigation, Project administration, Writing – original draft. JEH: Supervision, Writing – review and editing, Conceptualization. SH: Writing – review and editing, Conceptualization. TZ: Supervision, Writing – review and editing, Conceptualization. HG: Writing – review and editing, Supervision, Conceptualization.

Forensic Neuropsychology Consortium Members

Hasanen Al-Taiar, Nadia Bolognini, Lisanne Breuer, Josanne van Dongen, William Garmoe, Beverly Griffor, Jochem Jansen, Frank Jonker, Mike van Kessel, Niki Kuin, Femke Kuipers, Gary de Man, Bernice Marcopulos, Jesse Meijers, Pieter Molleman, Milo Oudenhoven, Rajakumari Reddy, J. Rajeshwaran Jamuna, Cristina Scarpazza, Anna van der Schoot, Carmen Sergiou, Andrea Stracciari, Marijke Willems, Stefano Zago.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors thank Eva Warreman for sharing her insights and experiences in conducting a Delphi study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1328839/full#supplementary-material

References

Aharoni, E., Mallett, J., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W., et al. (2014). Predictive accuracy in the neuroprediction of rearrest. *Soc. Neurosci.* 9, 332–336. doi: 10.1080/17470919.2014.907201

Aharoni, E., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W., Gazzaniga, M. S., et al. (2013). Neuroprediction of future rearrest. *Proc. Nat. Acad. Sci.* 110, 6223–6228. doi: 10.1073/pnas.1219302110

Alcázar-Córcoles, M. Á., Verdejo-García, A., Bouso-Saiz, J. C., and Bezos-Saldaña, L. (2010). Neuropsychology of impulsive aggression. *Revista de Neurologica* 10, 2009316. doi: 10.33588/rn.5005.2009316

Allen, J. J., and Anderson, C. A. (2017). "Aggression and violence: definitions and distinctions," in *The Wiley Handbook of Violence and Aggression,* ed. P. Sturmey (New York: John Wiley and Sons, Ltd), 1–14.

American Psychiatric Association (2022). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)*. Washington, D.C.: American Psychiatric Association Publishing.

Anderson, S. A. S., Hawes, D. J., and Snow, P. C. (2016). Language impairments among youth offenders: a systematic review. *Child. Youth Serv. Rev.* 65, 195–203. doi: 10.1016/j.childyouth.2016.04.004

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The reading the mind in the eyes test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *J. Child Psychiat. Psychiat.* 42, 715. doi: 10.1111/1469-7610.00715

Batista, A. X., Adda, C. C., Miotto, E. C., Lúcia, M. C. S. D., and Scaff, M. (2007). Torre de Londres e Torre de Hanói: Contribuições distintas para avaliação do funcionamento executivo. *J. Bras. Psiquiatr.* 56, 134–139. doi: 10.1590/S0047-20852007000200010

Bechara, A. (2016). *Iowa Gambling Task, Version 2. PAR.* Lutz: Psychological Assessment Resources, Inc.

Benton, A. L., Varney, N. R., and De Hamsher, K. (1978). Visuospatial judgment: a clinical test. *Arch. Neurol.* 35, 364. doi: 10.1001/archneur.1978.00500300038006

Bergvall, A. H., Wessely, H., Forsman, A., and Hansen, S. (2001). A deficit in attentional set-shifting of violent offenders. *Psychol. Med.* 31, 1095–1105. doi: 10.1017/S0033291701004317

Blair, R. J. R. (2003). Neurobiological basis of psychopathy. *Br. J. Psychiat.* 182, 5–7. doi: 10.1192/bjp.182.1.5

Bonta, J., and Andrews, D. A. (2023). *The psychology of criminal conduct (Seventh Edition)*. Routledge: Taylor and Francis Group.

British Psychological Society (n.d.). *Test Reviews*. Available online at: https://explore.bps.org.uk/test-reviews (accessed September 21, 2023).

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., et al. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nat. Neurosci.* 13, 419–421. doi: 10.1038/nn.2510

Burgess, J. (2020). A brief review of the relationship of executive function assessment and violence. *Aggress. Violent Behav.* 4, 101414. doi: 10.1016/j.avb.2020.101414

Buros Center for Testing (n.d.). *Buros Online Shop: Neuropsychological*. Available online at: https://marketplace.unl.edu/buros/test-reviews/categories/neuropsychological?dir=desc&limit=80&order=name&p=1 (accessed September 21, 2023).

Bushman, B. J., and Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychol. Rev.* 108, 273–279. doi: 10.1037//0033-295X.108.1.273

Buss, A. H. (1961). *The Psychology of Aggression*. New York: John Wiley and Sons Inc., 307.

Cambridge Cognition (2012). *CANTABeclipse Test Administration Guide*. Cambdige, UK: Cambridge Cognition Ltd.

Chase, K. A., O'Leary, K. D., and Heyman, R. E. (2001). Categorizing partner-violent men within the reactive–proactive typology model. *J. Consult. Clin. Psychol.* 69, 567–572. doi: 10.1037/0022-006X.69.3.567

Chow, J. C., Wallace, E. S., Senter, R., Kumm, S., and Mason, C. Q. (2022). A systematic review and meta-analysis of the language skills of youth offenders. *J. Speech, Lang. Hear. Res.* 65, 1166–1182. doi: 10.1044/2021_JSLHR-20-00308

Cima, M., and Raine, A. (2009). Distinct characteristics of psychopathy relate to different subtypes of aggression. *Pers. Individ. Dif.* 47, 835–840. doi: 10.1016/j.paid.2009.06.031

Cohen, R. A., Brumm, V., Zawacki, T. M., Paul, R., Sweet, L., and Rosenbaum, A. (2003). Impulsivity and verbal deficits associated with domestic violence. *J. Int. Neuropsychol. Soc.* 9, 760–770. doi: 10.1017/S1355617703950090

Cornet, L. J. M., Bootsman, F., Alberda, D. L., and de Kogel, C. H. (2016). "Neurowetenschappelijke toepassingen in de jeugdstrafrechtketen: Inventarisatie instrumenten, preventie en interventie," in *Wetenschappelijk Onderzoek- en Documentatiecentrum* (Den Haag).

Cornet, L. J. M., de Kogel, C. H., Nijman, H. L. I., Raine, A., and van der Laan, P. H. (2014). Neurobiological factors as predictors of cognitive–behavioral therapy outcome in individuals with antisocial behavior: a review of the literature. *Int. J. Offender Ther. Comp. Criminol.* 58, 1279–1296. doi: 10.1177/0306624X13494694

Cruz, A. R., de Castro-Rodrigues, A., and Barbosa, F. (2020). Executive dysfunction, violence and aggression. *Aggress. Violent Behav.* 51, 101380. doi: 10.1016/j.avb.2020.101380

Dalkey, N., and Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Manage. Sci.* 9, 458–467. doi: 10.1287/mnsc.9.3.458

Delis, D. C., Kaplan, E., and Kramer, J. H. (2007a). *Delis-Kaplan Executive Function System (D-KEFS): Color Word Interference Test*. Handleiding: Pearson.

Delis, D. C., Kaplan, E., and Kramer, J. H. (2007b). *Delis-Kaplan Executive Function System (D-KEFS): Tower Test*. Handleiding: Pearson.

Delis, D. C., Kaplan, E., and Kramer, J. H. (2007c). *Delis-Kaplan Executive Function System (D-KEFS): Trail Making Test*. Handleiding: Pearson.

Delis, D. C., Kaplan, E., and Kramer, J. H. (2001). Delis-Kaplan Executive Function System (D-KEFS) [Database record]. *APA PsycTests*. doi: 10.1037/t15082-000

Duckworth, A. L., and Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *J. Res. Pers.* 45, 259–268. doi: 10.1016/j.jrp.2011.02.004

Estrada, S., Tillem, S., Stuppy-Sullivan, A., and Baskin-Sommers, A. (2019). "Specifying the connection between reward processing and antisocial psychopathology across development: review, integration, and future directions," in *The Oxford Handbook of Positive Emotion and Psychopathology*, ed. J. Gruber (Oxford: Oxford University Press), 311–332.

Farrer, T. J., and Hedges, D. W. (2011). Prevalence of traumatic brain injury in incarcerated groups compared to the general population: a meta-analysis. *Prog. Neuro-Psychopharmacol. Biol. Psychiat.* 35, 390–394. doi: 10.1016/j.pnpbp.2011.01.007

Fayyad, J., Sampson, N. A., Hwang, I., Adamowski, T., Aguilar-Gaxiola, S., Al-Hamzawi, A., et al. (2017). The descriptive epidemiology of DSM-IV Adult ADHD in the World Health Organization World Mental Health Surveys. *ADHD Attent. Deficit Hyperact. Disord.* 9, 47–65. doi: 10.1007/s12402-016-0208-3

Fishbein, D., Sheppard, M., Hyde, C., Hubal, R., Newlin, D., Serin, R., et al. (2009). Deficits in behavioral inhibition predict treatment engagement in prison inmates. *Law Hum. Behav.* 33, 419–435. doi: 10.1007/s10979-008-9163-7

Frost, R. B., Farrer, T. J., Primosch, M., and Hedges, D. W. (2013). Prevalence of traumatic brain injury in the general adult population: a meta-analysis. *Neuroepidemiology* 40, 154–159. doi: 10.1159/000343275

Ginsberg, Y., Hirvikoski, T., and Lindefors, N. (2010). Attention Deficit Hyperactivity Disorder (ADHD) among longer-term prison inmates is a prevalent, persistent and disabling disorder. *BMC Psychiatry* 10, 112. doi: 10.1186/1471-244X-10-112

Golden, C. J., and Freshwater, S. M. (2002). *Stroop Color and Word Test: Revised Examiner's Manual*. Wood Dale, IL: Stoelting Co.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., et al. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain* 125, 752–764. doi: 10.1093/brain/awf079

Guilmette, T. J., Sweet, J. J., Hebben, N., Koltai, D., Mahone, E. M., Spiegler, B. J., et al. (2020). American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores. *Clin. Neuropsychol.* 34, 437–453. doi: 10.1080/13854046.2020.1722244

Haarsma, G., Davenport, S., White, D. C., Ormachea, P. A., Sheena, E., and Eagleman, D. M. (2020). Assessing risk among correctional community probation populations: predicting reoffense with mobile neurocognitive assessment software. *Front. Psychol.* 10, 2926. doi: 10.3389/fpsyg.2019.02926

Habibi, A., Sarafrazi, A., and Izadyar, S. (2014). Delphi technique theoretical framework in qualitative research. *Int. J. Eng. Sci.* 3, 8–13.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093

Harris, J. C. (2006). *Intellectual Disability: Understanding its Development, Causes, Classification, Evaluation, and Treatment*. Oxford: Oxford University Press.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., and Curtis, G. (1993). *Wisconsin Card Sorting Test (WCST) Manual, Revised and Expanded*. Lutz, FL: Psychological Assessment Resources.

Hellenbach, M., Karatzias, T., and Brown, M. (2017). Intellectual disabilities among prisoners: prevalence and mental and physical health comorbidities. *J. Appl. Res. Intellect. Disabil.* 30: 230–241. doi: 10.1111/jar.12234

Hetland, H., Eikeland, O.-J., Manger, T., Diseth, Å., and Asbjørnsen, A. (2007). Educational background in a prison population. *J. Correct. Educ.* 58, 145–156.

Hilsenroth, M. J., and Segal, D. L. (2004). *Comprehensive Handbook of Psychological Assessment: Volume 2 Personality Assessment*. New York: John Wiley and Sons.

Humes, G. E., Welsh, M. C., Retzlaff, P., and Cookson, N. (1997). Towers of Hanoi and London: reliability and validity of two executive function tasks. *Assessment* 4, 249–257. doi: 10.1177/107319119700400305

Hutten, J. C., Ziermans, T. B., Geurts, H. M. G., and Van Horn, J. E. (preprint). Cognitive functioning in offenders compared to non-offending controls: a multi-level meta-analysis. *[Preprint]*. doi: 10.31234/osf.io/s48rn

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiat.* 167 748–751. doi: 10.1176/appi.ajp.2010.09091379

Iqbal, S., and Pipon-Young, l. (2009). The Delphi Method. *Psychol. Methods* 22.

Jorm, A. F. (2015). Using the Delphi expert consensus method in mental health research. *Aust. New Zeal. J. Psychiat.* 49, 887–897. doi: 10.1177/0004867415600891

Karoglu, N., Ferguson, H. J., and ÓCiardha, C. (2022). Theory of mind in offending: a systematic review. *Trauma Viol. Abuse*, 23, 1610–1628. doi: 10.1177/15248380211013143

Kasten, E., Barbosa, F., Kosmidis, M. H., Persson, B. A., Constantinou, M., Baker, G. A., et al. (2021). European clinical neuropsychology: role in healthcare and access to neuropsychological services. *Healthcare* 9, 734. doi: 10.3390/healthcare9060734

Kempes, M., Berends, I., Duits, N., and Brink, W. V. D. (2019). Neurobiological information and consideration in dutch pre-trial forensic reports of juvenile criminal offenders. *Int. J. Forensic Ment. Health* 18, 212–219. doi: 10.1080/14999013.2018.1525776

Kessels, R. P. C., Montagne, B., Hendriks, A. W., Perrett, D. I., and De Haan, E. H. F. (2014). Assessment of perception of morphed facial expressions using the Emotion Recognition Task: Normative data from healthy participants aged 8-75. *J. Neuropsychol.* 8, 75–93. doi: 10.1111/jnp.12009

Kosmidis, M. H. (2018). Challenges in the neuropsychological assessment of illiterate older adults. *Lang. Cognit. Neurosci.* 33, 373–386. doi: 10.1080/23273798.2017.1379605

Köstering, L., Schmidt, C. S. M., Egger, K., Amtage, F., Peter, J., Klöppel, S., et al. (2015). Assessment of planning performance in clinical samples: Reliability and validity of the Tower of London task (TOL-F). *Neuropsychologia* 75, 646–655. doi: 10.1016/j.neuropsychologia.2015.07.017

Krahé, B. (2013). *The Social Psychology of Aggression: 2nd Edition (0 ed.)*. New York: Psychology Press.

Kuhlmann, B., and Margraf, J. (2023). A new short version of the Facial Expressions of Emotions: Stimuli and Test (FEEST) including prototype and morphed emotional stimuli. *Front. Psychol.* 14, 1198386. doi: 10.3389/fpsyg.2023.1198386

Kuin, N., Masthoff, E., Kramer, M., and Scherder, E. (2015). The role of risky decision-making in aggression: a systematic review. *Aggress. Violent Behav.* 25, 159–172. doi: 10.1016/j.avb.2015.07.018

Langton, L., Truman, J., and BJS Statisticians. (2014). *Socio-Emotional Impact of Violent Crime (NCJ 247076)*. Washington, D. C.: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Lawrence, A. J., Luty, J., Bogdan, N. A., Sahakian, B. J., and Clark, L. (2009). Problem gamblers share deficits in impulsive decision-making with alcohol-dependent individuals. *Addiction* 104, 1006–1015. doi: 10.1111/j.1360-0443.2009.02533.x

Lezak, M. D. (Ed.). (2012). Neuropsychological assessment (5th ed). Oxford University Press.

Li, F., Li, X., and Kou, H. (2023). Emotional recognition training enhances attention to emotional stimuli among male juvenile delinquents. *Psychol. Res. Behav. Manage.* 16, 575–586. doi: 10.2147/PRBM.S403512

Lilienfeld, S. O., Wood, J. M., and Garb, H. N. (2000). The Scientific Status of Projective Techniques. *Psychol. Sci. Public Interest* 1, 27–66. doi: 10.1111/1529-1006.002

Linstone, H. A., and Turoff, M. (1975). *The Delphi Method: Techniques and Applications*. Boston: Addison-Wesley Publishing.

Logan, G. D. (1994). "On the ability to inhibit thought and action: A user's guide to the stop signal paradigm," in *Inhibitory Processes in Attention, Memory, and Language,* ed. D. Dagenbach and T. H. Carr (Cambridge, MA: Academic Press), 189–239.

Manning, K. E. (2020). Seeing red? A systematic review of the evidence for attentional biases to threat-relevant stimuli in propensity to reactive aggression. *Aggress. Violent Behav.* 50, 101359. doi: 10.1016/j.avb.2019.101359

Marsh, A. A., and Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: A meta-analysis. *Neurosci. Biobehav. Rev.* 32, 454–465. doi: 10.1016/j.neubiorev.2007.08.003

Matheson, F. I., Dastoori, P., Whittingham, L., Calzavara, A., Keown, L. A., Durbin, A., et al. (2022). Intellectual/developmental disabilities among people incarcerated in federal correctional facilities in Ontario, Canada: examining prevalence, health and correctional characteristics. *J. Appl. Res. Intellect. Disabil.* 35, 900–909. doi: 10.1111/jar.12995

McMurran, M., Egan, V., Richardson, C., Street, H., Ahmadi, S., and Cooper, G. (2000). Referrals for anger and aggression in forensic psychology outpatient services. *J. Forens. Psychiat.* 11, 206–213. doi: 10.1080/095851800362481

Merk, W., Orobio De Castro, B., Koops, W., and Matthys, W. (2005). The distinction between reactive and proactive aggression: utility for theory, diagnosis and treatment? *Eur. J. Dev. Psychol.* 2, 197–220. doi: 10.1080/17405620444000300

Meyers, J. E., and Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial: Professional Manual.* Lutz, FL: Psychological Assessment Resources.

Monahan, J., and Skeem, J. L. (2014). Risk Redux: the resurgence of risk assessment in criminal sanctioning. *Federal Sentenc. Report.* 26, 158–166. doi: 10.1525/fsr.2014.26.3.158

Montagne, B., Kessels, R. P. C., De Haan, E. H. F., and Perrett, D. I. (2007). The emotion recognition task: a paradigm to measure the perception of facial emotional expressions at different intensities. *Percept. Mot. Skills* 104, 589–598. doi: 10.2466/pms.104.2.589-598

Morgan, C. D., and Murray, H. A. (1935). A method for investigating fantasies: the thematic apperception test. *Arch. Neurol. Psychiat.* 34, 289. doi: 10.1001/archneurpsyc.1935.02250200049005

Muñoz García-Largo, L., Martí-Agustí, G., Martin-Fumadó, C., and Gómez-Durán, E. L. (2020). Intellectual disability rates among male prison inmates. *Int. J. Law Psychiat.* 70, 101566. doi: 10.1016/j.ijlp.2020.101566

National Institute of Mental Health (2023). *RDoC Matrix. National Institute of Mental Health*. Available online at: https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/rdoc-matrix (accessed February 9, 2024).

Nauta-Jansen, L. M. C. (2022). The neurobiology of antisocial behavior in adolescence; current knowledge and relevance for youth forensic clinical practice. *Curr. Opini. Psychol.* 47, 101356. doi: 10.1016/j.copsyc.2022.101356

NIP (n.d.). *COTAN Documentatie*. Available online at: https://www.cotandocumentatie.nl/beoordelingen/ (accessed September 21, 2023).

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., and Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavior: a meta-analysis. *Criminology* 49, 1063–1107. doi: 10.1111/j.1745-9125.2011.00252.x

Ogonah, M. G. T., Seyedsalehi, A., Whiting, D., and Fazel, S. (2023). Violence risk assessment instruments in forensic psychiatric populations: a systematic review and meta-analysis. *Lancet Psychiat.* 10, 780–789. doi: 10.1016/S2215-0366(23)00256-0

Osterhaus, C., and Bosacki, S. L. (2022). Looking for the lighthouse: a systematic review of advanced theory-of-mind tests beyond preschool. *Dev. Rev.* 64, 101021. doi: 10.1016/j.dr.2022.101021

Papalia, N., Spivak, B., Daffern, M., and Ogloff, J. R. P. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clin. Psychol.: Sci. Pract.* 26, e12282. doi: 10.1111/cpsp.12282

Parrott, D. J., and Giancola, P. R. (2007). Addressing "The criterion problem" in the assessment of aggressive behavior: Development of a new taxonomic system. *Aggress. Violent Behav.* 12, 280–299. doi: 10.1016/j.avb.2006.08.002

Patel, D. M., and Taylor, R. M. (2012). *Social and Economic Costs of Violence: Workshop Summary*. Washington, D.C.: National Academies Press, 13254.

Pearson NL (2012). *WAIS-IV-NL Psychometrische Eigenschappen*. Boston: Pearson.

Qualtrics (2023). *Qualtrics Software*. Available online at: https://www.qualtrics.com (accessed February 9, 2024).

Raskin, M. (1994). The Delphi study in-field instruction revisited—expert consensus on issues and research priorities. *J. Soc. Work Educ.*, 30: 75–89. doi: 10.1080/10437797.1994.10672215

Rayens, M. K., and Hahn, E. J. (2000). Building consensus using the policy Delphi method. *Policy, Polit. Nursing Pract.* 1, 308–315. doi: 10.1177/152715440000100409

Rivara, F., Adhia, A., Lyons, V., Massey, A., Mills, B., Morgan, E., et al. (2019). The effects of violence on health. *Health Aff.* 38, 1622–1629. doi: 10.1377/hlthaff.2019.00480

Rogers, R. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers evidence for monoaminergic mechanisms. *Neuropsychopharmacology* 20, 322–339. doi: 10.1016/S0893-133X(98)00091-8

Ross, E. H., and Hoaken, P. N. S. (2011). Executive cognitive functioning abilities of male first time and return canadian federal inmates. *Can. J. Criminol. Crimi. Just.* 53, 377–403. doi: 10.3138/cjccj.53.4.377

Sánchez de Ribera, O., Trajtenberg, N., and Cook, S. (2022). Executive functioning among first time and recidivist inmates in Uruguay. *Appl. Neuropsychol.: Adult* 29, 1242–1249. doi: 10.1080/23279095.2020.1864634

Schmitz, F., Kunina-Habenicht, O., Hildebrandt, A., Oberauer, K., and Wilhelm, O. (2020). Psychometrics of the Iowa and Berlin gambling tasks: unresolved issues with reliability and validity for risk taking. *Assessment* 27, 232–245. doi: 10.1177/1073191117750470

Schulz, K., Fan, J., Magidina, O., Marks, D., Hahn, B., and Halperin, J. (2007). Does the emotional go/no-go task really measure behavioral inhibition? Convergence with measures on a non-emotional analog. *Arch. Clini. Neuropsychol.* 22, 151–160. doi: 10.1016/j.acn.2006.12.001

Serafim, A. D. P., Saffi, F., Silva, T. G. B. D., Almeida, C. V. D., Hokama, E., Barros, D. M. D., et al. (2015). Forensic neuropsychological assessment: a review of its scope. *Arch. Clini. Psychiat.* 42, 63–67. doi: 10.1590/0101-60830000000049

Shumlich, E. J., Reid, G. J., Hancock, M., and Hoaken, P. (2019). Executive dysfunction in criminal populations: comparing forensic psychiatric patients and correctional offenders. *Int. J. Forensic Ment. Health* 18, 243–259. doi: 10.1080/14999013.2018.1495279

Shunk, A. W., Davis, A. S., and Dean, R. S. (2006). Test review: Dean C. Delis, Edith Kaplan and Joel H. Kramer, Delis Kaplan Executive Function System (D-KEFS), The Psychological Corporation, San Antonio, TX, 2001. $415.00 (complete kit). *Appl. Neuropsychol.* 13: 275–27. doi: 10.1207/s15324826an1304_9

Simon, L. M. J. (1997). Do criminal offenders specialize in crime types? *Appl. Prevent. Psychol.* 6, 35–53. doi: 10.1016/S0962-1849(05)80064-2

Smeijers, D. (2017). *Severe Aggressive Behavior. Towards Understanding some of the Underlying Characteristics and Treatment Responsivity*. Available online at: http://rgdoi.net/10.13140/RG.2.2.12181.42728

Söderstrand, P., and Almkvist, O. (2012). Psychometric data on the eyes test, the faux pas test, and the dewey social stories test in a population-based swedish adult sample. *Nordic Psychology* 64, 30–43. doi: 10.1080/19012276.2012.693729

Soothill, K., Francis, B., Sanderson, B., and Ackerley, E. (2000). Sex offenders: specialists, generalists—or both? *Br. J. Criminol.* 40, 56–67. doi: 10.1093/bjc/40.1.56

Steward, K. A., and Kretzmer, T. (2022). Anosognosia in moderate-to-severe traumatic brain injury: a review of prevalence, clinical correlates, and diversity considerations. *Clin. Neuropsychol.* 36, 2021–2040. doi: 10.1080/13854046.2021.1967452

Stone, V. E., Baron-Cohen, S., and Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* 10, 640–656. doi: 10.1162/089892998562942

Strauss, E., Sherman, E. M., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, Third Edition*. Oxford: Oxford University Press.

Sullivan, C. J., Mcgloin, J. M., Pratt, T. C., and Piquero, A. R. (2006). Rethinking the "norm" of offender generality: investigating specialization in the short-term. *Criminology* 44, 199–233. doi: 10.1111/j.1745-9125.2006.00047.x

Tuominen, T., Korhonen, T., Hamalainen, H., Temonen, S., Salo, H., Katajisto, J., et al. (2014). Functional illiteracy and neurocognitive deficits among male prisoners: implications for rehabilitation. *J. Foren. Pract.* 16, 268–280. doi: 10.1108/JFP-10-2013-0044

Turoff, M. (1975). "The policy Delphi," in *The Delphi Method. Techniques and Applications*, eds H. A. Linstone, and M. Turoff (London: Addison-Wesley), 84–101.

Van De Kant, T., Boers, S. F., Kempes, M., and Egger, J. I. M. (2020). Neuropsychological subtypes of violent behaviour: differences in inhibition between affective and instrumental violence. *J. Foren. Psychol.* 5, 7. doi: 10.35248/2475-319X.19.5.158

Venturi Da Silva, L., and Cavalheiro Hamdan, A. (2022). Neuropsychological assessment in the forensic context: a scoping review. *Brazil. J. Foren. Sci. Med. Law Bioethics* 12, 53–74. doi: 10.17063/bjfs12(1)y202253-74

Völlm, B. A., Edworthy, R., Huband, N., Talbot, E., Majid, S., Holley, J., et al. (2018). Characteristics and pathways of long-stay patients in high and medium secure settings in england; a secondary publication from a large mixed-methods study. *Front. Psychiatry* 9, 140. doi: 10.3389/fpsyt.2018.00140

Wechsler, D. (1997). *WMS-III Administration and Scoring Manual*. New York, NY: The Psychological Corporation.

Wechsler, D. (2012). *Wechsler Adult Intelligence Scale—Fourth Edition*. London, UK: Pearson.

White, B. A., Gordon, H., and Guerra, R. C. (2015). Callous–unemotional traits and empathy in proactive and reactive relational aggression in young women. *Pers. Individ. Dif.* 75, 185–189. doi: 10.1016/j.paid.2014.11.031

Wigham, S., McGovern, R., Kaner, E., and Hackett, S. S. (2022). A review of recent innovation in psychosocial interventions for reducing violence and aggression in adults using a horizon scanning approach. *Aggress. Violent Behav.* 62, 101685. doi: 10.1016/j.avb.2021.101685

Willner, P., Bailey, R., Parry, R., and Dymond, S. (2010). Evaluation of executive functioning in people with intellectual disabilities. *J. Intell. Disab. Res.* 54, 366–379. doi: 10.1111/j.1365-2788.2010.01249.x

Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., and Ekman, P. (2002). *Facial Expressions of Emotion: Stimuli and Tests* (FEEST). Suffolk: Thames Valley Test Company.

Yukhnenko, D., Farouki, L., and Fazel, S. (2023). Criminal recidivism rates globally: A 6-year systematic review update. *J. Crim. Justice* 88, 102115. doi: 10.1016/j.jcrimjus.2023.102115

Zhang, Z., Wang, Q., Liu, X., Song, P., and Yang, B. (2017). Differences in inhibitory control between impulsive and premeditated aggression in juvenile inmates. *Front. Hum. Neurosci.* 11, 373. doi: 10.3389/fnhum.2017.00373

Zijlmans, J., Marhe, R., Bevaart, F., Van Duin, L., Luijks, M.-J. A., Franken, I., et al. (2021). The predictive value of neurobiological measures for recidivism in delinquent male young adults. *J. Psychiat. Neurosci.* 46, 10. doi: 10.1503/jpn.200103

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 March 2024
DOI 10.3389/fpsyg.2024.1275315

OPEN ACCESS

EDITED BY
Märta Wallinius,
Lund University, Sweden

REVIEWED BY
Vaitsa Giannouli,
Aristotle University of Thessaloniki, Greece
Alessandra Scotto di Freca,
University of Cassino, Italy

*CORRESPONDENCE
Maurizio Balestrino
✉ mbalestrino@neurologia.unige.it

RECEIVED 21 November 2023
ACCEPTED 26 February 2024
PUBLISHED 28 March 2024

CITATION
Balestrino M, Brugnolo A, Girtler N, Pardini M, Rizzetto C, Alì PA, Cocito L and Schiavetti I (2024) Cognitive impairment assessment through handwriting (COGITAT) score: a novel tool that predicts cognitive state from handwriting for forensic and clinical applications.
Front. Psychol. 15:1275315.
doi: 10.3389/fpsyg.2024.1275315

Cognitive impairment assessment through handwriting (COGITAT) score: a novel tool that predicts cognitive state from handwriting for forensic and clinical applications

Maurizio Balestrino[1,2]*, Andrea Brugnolo[1,2], Nicola Girtler[1,2], Matteo Pardini[1,2], Cristiano Rizzetto[1], Paolo Alessandro Alì[1], Leonardo Cocito[1] and Irene Schiavetti[3]

[1]Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal and Child Sciences (DINOGMI), University of Genoa, Genoa, Italy, [2]IRCCS Ospedale Policlinico San Martino, Genoa, Italy, [3]Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy

Introduction: Handwriting deteriorates proportionally to the writer's cognitive state. Such knowledge is of special importance in the case of a contested will, where dementia of the testator is claimed, but medical records are often insufficient to decide what the testator's cognitive state really was. By contrast, if the will is handwritten, handwriting analysis allows us to gauge the testator's cognitive state at the precise moment when he/she was writing the will. However, quantitative methods are needed to precisely evaluate whether the writer's cognitive state was normal or not. We aim to provide a test that quantifies handwriting deterioration to gauge a writer's cognitive state.

Methods: We consecutively enrolled patients who came for the evaluation of cognitive impairment at the Outpatient Clinic for Cognitive Impairment of the Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal and Child Sciences (DINOGMI) of the University of Genoa, Italy. Additionally, we enrolled their caregivers. We asked them to write a short text by hand, and we administered the Mini Mental State Examination (MMSE). Then, we investigated which handwriting parameters correlated with cognitive state as gauged by the MMSE.

Results: Our study found that a single score, which we called the **COG**nitive **I**mpairment **T**hrough h**A**ndwri**T**ing (**COGITAT**) score, reliably allows us to predict the writer's cognitive state.

Conclusion: The COGITAT score may be a valuable tool to gage the cognitive state of the author of a manuscript. This score may be especially useful in contested handwritten wills, where clinical examination of the writer is precluded.

KEYWORDS

forensic science, will challenge, handwriting analysis, cognitive impairment, dementia, posthumous

1 Introduction

Available scientific data indicate that cognitive impairment significantly compromises handwriting. Already in 1911 Tamburini claimed, citing the works of his contemporaries Borri and Grilli, that in patients diagnosed with dementia, affected by what was then known as progressive general paralysis, "the handwriting had that special character which alone betrays the profound neuro-psychic alteration of the writer" (Tamburini, 1911). Coming to current times, healthy subjects make spelling mistakes in 2% of words versus 25% of patients with mild Alzheimer's disease and 83% of patients with severe Alzheimer's disease (Silveri et al., 2007). The handwriting of Alzheimer's disease patients worsens as the disease progresses (Luzzatti et al., 2003) and indeed the simple inspection of a handwritten sentence gives some hints on the cognitive status of the person who wrote it (Shenkin et al., 2008). Alzheimer's disease patients produce shorter and less informative writings, produce more paraphasias, and make more mistakes in letter formation (Forbes et al., 2004). There are strong and marked correlations between cognitive tests and parameters of handwriting such as the length of the text, the number of comprehensible words and the amount of errors (Renier et al., 2016). Lexical, semantic and syntactic parameters of the written text, as well as frequency of spelling errors, are not impaired by normal aging but they are by Alzheimer's disease (Croisile, 2005).

Analysis of handwriting to ascertain a possible cognitive impairment is of special importance in the case of a contested will, where dementia of the testator is often claimed. In such trials, medical records are often insufficient, and witnesses often offer contradictory or unreliable reports. By contrast, analysis of handwriting offers the possibility of gauging the testator's performance in the precise moment when he/she was making the will. A poor handwriting may then indicate a cognitive impairment.

To carry out such an analysis one needs a score that quantifies handwriting deterioration, and a cutoff for cognitive impairment. To this aim, we created and investigated the "writing score," which quantifies how much handwriting is compromised (Fontana et al., 2008; Balestrino et al., 2012). The "writing score" evaluates, in a semi-quantitative manner, the legibility of the text as well as its spatial orientation. We demonstrated (Fontana et al., 2008; Balestrino et al., 2012) that the writing score correlates significantly with both the Mini Mental State Examination (Folstein et al., 1975) and the Milan Overall Dementia Assessment (Brazzelli et al., 1994). Its predictive value is rather reliable for scores at either end of its scale; very low scores predict cognitive impairment while very high scores predict cognitive normality. However, intermediate scores are not very specific, occurring both in cognitively compromised persons and in normal controls (Balestrino et al., 2017).

With the present research, we further investigate the relationship between handwriting and cognitive status, and we attempt to identify a cutoff score that may reliably identify subjects with cognitive impairment.

2 Methods

2.1 Patients' enrollment

Patients were consecutively enrolled from those seeking clinical attention for cognitive impairment evaluation at the Outpatient Clinic for Cognitive Impairment of the Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal and Child Sciences (DINOGMI) of the University of Genoa, Italy. Caregivers accompanying them were also included in the investigation. To enhance sample representativeness for the general population, the only inclusion criteria were the absence of severe visual sensory deficits and Italian mother-tongue. Initially, all patients and caregivers underwent testing, but subsequently, all subjects (patients and caregivers) younger than 50-year-old were excluded. We selected this cutoff because caregivers were mostly in that age group, and 50 is the age when the earliest cases of cognitive impairment occur (Albert and Heaton, 1988). All subjects signed an institutional consent form before enrollment in the study. All participants provided written informed consent, agreeing to the use and processing of their data for scientific purposes. They received information about the study's purpose, data usage, and their right to withdraw without affecting their clinical care. Ethical review and approval were not necessary for the study in compliance with national legislation and institutional requirements. The study was conducted in accordance with the national legislation and institutional requirements. The participants provided their written informed consent for research participation.

2.2 Neuropsychological and handwriting test

In addition to the usual assessment, which routinely includes an MMSE, all patients were asked to write a spontaneous text on blank paper. Caregivers were separately administered an MMSE and asked to write a spontaneous text as well. Both patients and caregivers were instructed as follows: "Write whatever you like on this paper, using no more than 6 or 7 lines. Do not worry about errors or corrections, this is not a school examination, and no one will give you a grade."

For all handwriting samples, we assessed:

- The Writing Score (Fontana et al., 2008; Balestrino et al., 2012), a numerical measure of handwriting quality representing the sum of two scores. The first evaluates overall correctness and legibility from a verbal standpoint (the "Verbal and lexical skills" scale), while the second evaluates spatial orientation, specifically the horizontal alignment of lines and how closely margins correspond to those of the sheet (the "Spatial orientation" scale). Each scale ranges from 1 to 5, with higher score indicating better quality. Please refer to (Balestrino et al., 2012) for additional details and handwriting samples. In this manuscript, Figures 1–7 show handwriting samples illustrating various "Spatial orientation" scores. Moreover, Supplementary Table II illustrates the Writing Score, as originally published. Briefly, the "writing score" is a categorical, semiquantitative score, whose values are assigned based on the specific definitions that some of us ideated and published (Fontana et al., 2008; Balestrino et al., 2012). As such, it is not the result of measurements on the text, but it is the result of the evaluator's judgment. We emphasize that the score obtained in this way correlates significantly with formal neuropsychological tests of cognitive state (*ibidem*), and that the test has a significant inter-observer agreement (Fontana et al., 2008). To the best of our knowledge, the writing score is the only

10.3389/fpsyg.2024.1275315

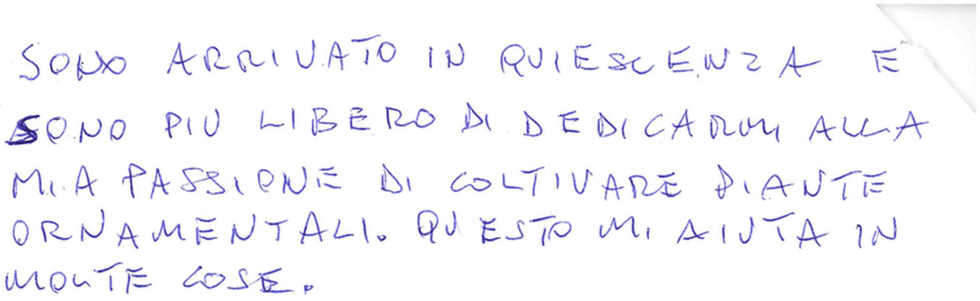


FIGURE 1
Example of a manuscript that was scored 5 ("normally oriented rows. In each row, beginning and end correspond to the page margins") in the "Spatial orientation" item of the "Writing Score."

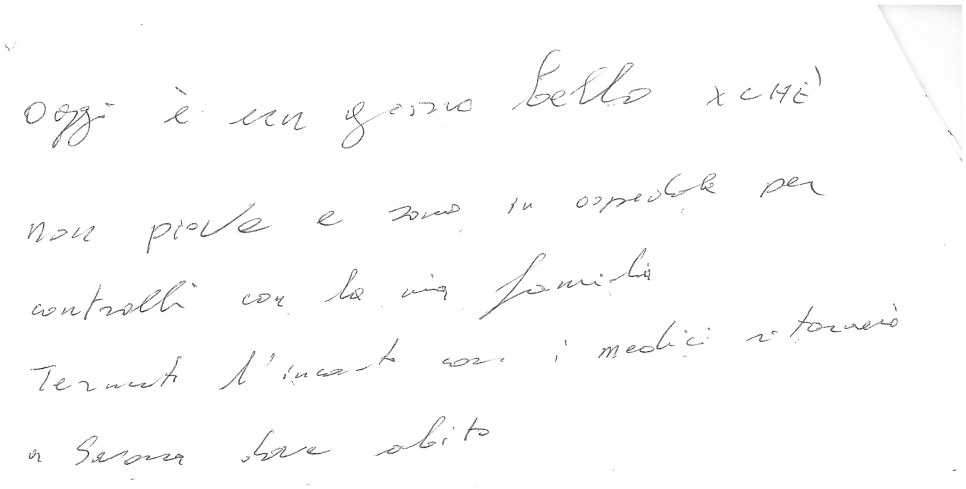


FIGURE 2
Example of a manuscript that was scored 4 ("rows slightly distorted or with beginning and end bearing little correspondence to the page margins") in the "Spatial orientation" item of the "Writing Score."

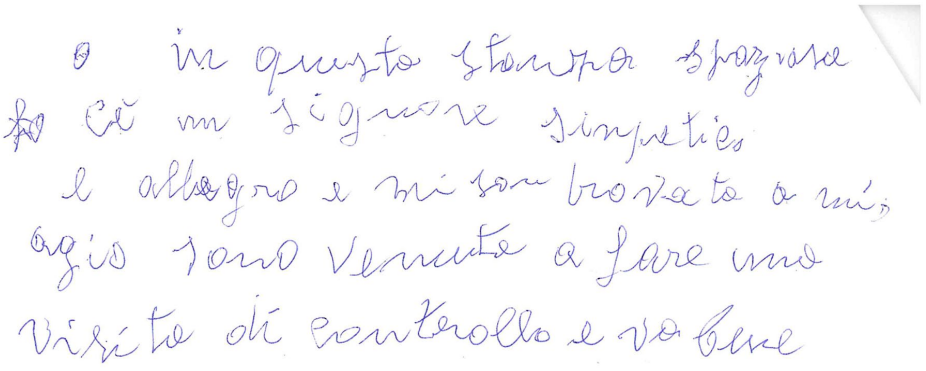


FIGURE 3
Example of a manuscript that was scored 4 ("rows slightly distorted or with beginning and end bearing little correspondence to the page margins") in the "Spatial orientation" item of the "Writing Score." Please note that in this case the score 4 was attributed because even if the rows are fairly horizontal the margins bear little correspondence to the page margins.

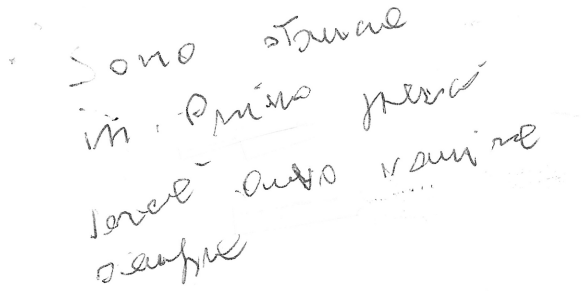


FIGURE 4
Example of a manuscript that was scored 3 ("rows clearly distorted or with beginning and end not corresponding to the page margins") in the "Spatial orientation" item of the "Writing Score." Please note that underwriting of some words was done by the examiners during handwriting analysis.

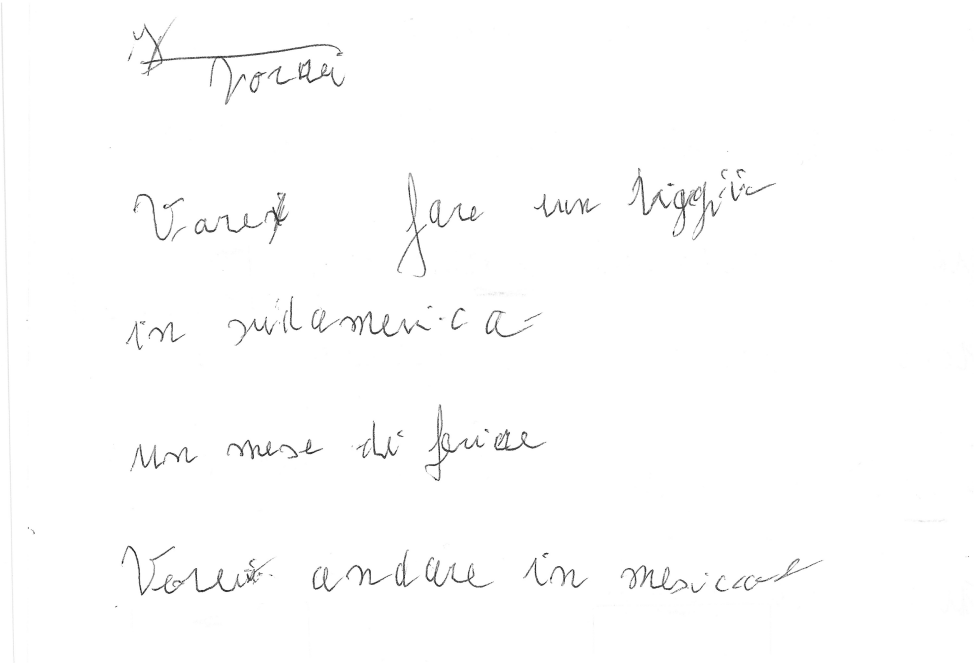


FIGURE 5
Example of a manuscript that was scored 3 ("rows clearly distorted or with beginning and end not corresponding to the page margins") in the "Spatial orientation" item of the "Writing Score." Please note that in this case the score 3 was attributed because even if the rows are fairly horizontal the margins are not corresponding to the page margins.

quantitative method that allows evaluation of handwriting in a forensically relevant way.

- The percentage of spelling and grammar errors, defined as the percentage of words in the text containing such errors. We considered as "spelling and grammar errors" those resulting in a mistake in how the word is written, for example letters missing or replaced. We did not considered as errors letters traced in an incorrect way (e.g., a letter "t" missing the horizontal tract)
- The total number of words written
- The percentage of words written, even partially, in capital letters.

2.3 Statistical analysis

Based on their MMSE score, all subjects were categorized as normal (MMSE≥24) or with cognitive impairment (MMSE <24). This

FIGURE 6
Example of a manuscript that was scored 2 ("Words or letters inserted where they do not belong in the text.") in the "Spatial orientation" item of the "Writing Score." Please note that in this example an indecipherable grapheme is placed out of context in the upper left corner.

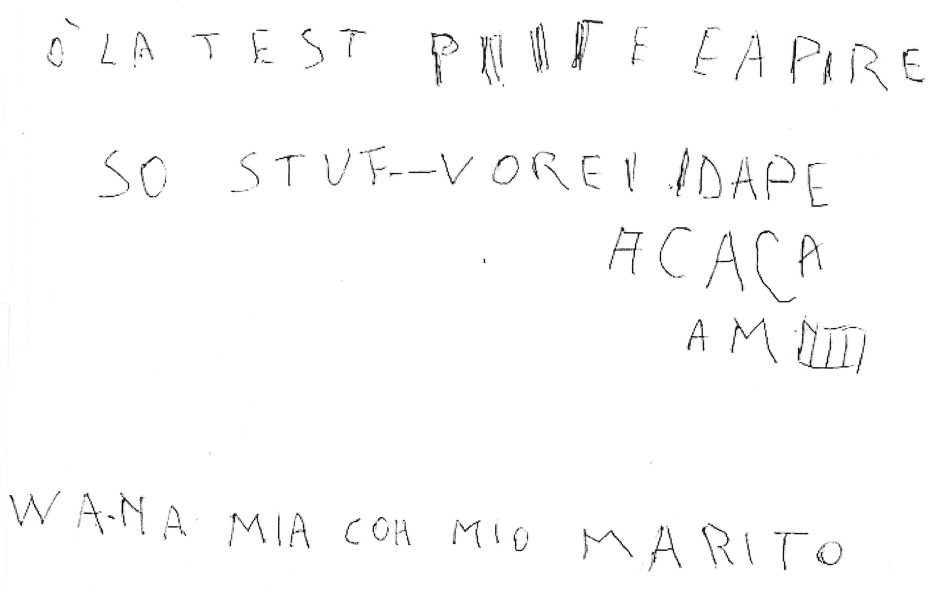


FIGURE 7
Example of a manuscript that was scored 1 ("chaotic orientation of the rows") in the "Spatial orientation" item of the "Writing Score."

score was the cutoff that separated normal from demented people in a large validation study in Italy (Measso et al., 1993) and it is still largely used by Italian neurologists. This cutoff score is also one of the most widely used worldwide (Tsoi et al., 2015).

Baseline characteristics were reported as mean ± standard deviation (s.d.) or count with frequency, as appropriate.

A novel predictive score useful for identifying the probability of cognitive impairment was derived and validated through a univariate and subsequent multivariate stepwise logistic regression analyses.

The model considered demographic variables as well as all items collected for spontaneous text. We chose to use spontaneous testing in building the score because it is more easily recoverable, even in normal real-life conditions, compared to dictated text that would need to be requested specifically. Variables with value of $p < 0.20$ in the univariate model (age, years of education, writing score—verbal and lexical skills, writing score - spatial orientation, total number of words, and percent of error) were candidates for multivariate analysis, where a backward stepwise variable selection with a value of $p < 0.10$ for inclusion and exclusion was applied.

Coefficients (β) with their standard error (S.E.), together with Odds Ratio (OR) and 95% Confidence Interval (CI) were estimated for each of the significant variables. ROC curve was graphed for identifying the Area Under the Curve (AUC) and assessing the discrimination of the fitted logistic model.

The new score was validated with a split-sample internal validation method. The whole sample was randomly divided into two groups, a training cohort (70%) and an internal validation cohort (30%) based on random computer generation. Characteristics of patients in the two data sets were compared using the chi-square test or Fisher's exact test for categorical variables and the Mann–Whitney U test for continuous variables.

The regression model applied to the whole group was firstly replicated on the training cohort to verify whether it produced the same subset of predictors. Coefficients (β) obtained from the regression analysis on the training cohort were used for deriving a

TABLE 1 Baseline characteristics of enrolled subjects ($N = 167$).

	Normal (MMSE ≥ 24)	Cognitive impairment (MMSE < 24)	Total
	$n = 113$	$n = 54$	$n = 167$
Sex, males—*n (%)*	41 (36.3)	17 (31.5)	58 (34.7)
Age (years)—*mean ± s.d.*	68.5 ± 11.62	79.2 ± 8.38	72.0 ± 11.78
Years of education—*mean ± s.d.*	12.1 ± 4.81	8.2 ± 3.86	10.8 ± 4.86
MMSE—*mean ± s.d.*	28.4 ± 1.89	17.1 ± 4.45	24.8 ± 6.07
Spontaneous Text			
Total writing score—*mean ± s.d.*	9.2 ± 1.34	6.8 ± 2.14	8.4 ± 2.00
Writing score—verbal and lexical skills—*mean ± s.d.*	4.6 ± 0.82	3.3 ± 1.31	4.2 ± 1.18
Writing score—spatial orientation—*mean ± s.d.*	4.6 ± 0.71	3.5 ± 1.06	4.3 ± 0.99
Total number of words—*mean ± s.d.*	35.1 ± 12.15	24.4 ± 10.97	31.6 ± 12.78
Percent of words in capital letters—*mean ± s.d.*	8.0 ± 25.90	14.0 ± 32.57	9.9 ± 28.28
Percent of errors—*mean ± s.d.*	1.4 ± 3.31	16.6 ± 25.77	6.4 ± 16.48
Dictated Text			
Total writing score—*mean ± s.d.*	9.4 ± 1.09	7.6 ± 1.87	8.9 ± 1.62
Writing score—verbal and lexical skills—*mean ± s.d.*	4.7 ± 0.75	3.5 ± 1.19	4.3 ± 1.07
Writing score—spatial orientation—*mean ± s.d.*	4.8 ± 0.50	4.2 ± 0.95	4.6 ± 0.73
Total number of errors—*mean ± s.d.*	1.1 ± 2.92	8.7 ± 9.27	3.6 ± 6.75

score defined as the linear combination of the coefficients multiplied by the corresponding value of the n variables (score = $\beta 1 \times var1 + \beta 2 \times var2 + \ldots + \beta n \times varn$), where higher scores represented a greater risk for cognitive impairment.

The discriminating performance of the score was evaluated in two steps. Firstly, in the training dataset two optimal cut-off scores were identified by maximizing, respectively, their specificity and their sensitivity, so as to detect with approximately 95% probability those subjects that were or, respectively, were not cognitively impaired. Subsequently, the performance of the score was assessed in the validation sample by applying a univariable logistic regression model with the binary score and by deriving sensitivity, specificity and AUC with relating 95% CI.

The probability of showing cognitive impairment based on the estimated coefficients as follows:

$$Probability\ of\ cognitive\ impairment\ (P) = \frac{e^{(\beta 0+\beta 1 var1+\ldots+\beta n varn)}}{1+e^{(\beta 0+\beta 1 var1+\ldots+\beta n varn)}}$$

The recommended sample-to-variable ratio suggests a minimum observation-to-variable ratio of 5:1, with preferred ratios of 15:1 or 20:1 (Hair et al., 2018). Consequently, for our internal validity study, which involves 11 independent variables in a logistic model, we have considered a minimum sample size of 165 patients (15:1), with a final enrollment of 167 patients.

Statistical analysis was performed using SPSS (RRID:SCR_002865) version 24.0.

3 Results

One hundred and sixty-seven adult individuals (patients and caregivers) aged 50 years old or older (range 50–93 years) were enrolled. Fifty-four of them (32.3%) had cognitive impairment (defined as MMSE<24), whereas the remaining 113 subjects (67.7%) reported a normal value of MMSE (≥ 24). Table 1 summarizes baseline data and identified scores and evaluations-related writing characteristics of the whole sample and for each MMSE group. The full database is included in the Supplementary material.

Results for the evaluation of predictors for cognitive impairment are shown in Table 2. The multivariate analysis confirmed age (OR: 1.07; 95%CI: 1.02–1.13; $p = 0.008$) together with three other (even if not fully significant) characteristics of spontaneous text (writing score - spatial orientation; total number of words; and percent of errors) as independent factors associated with cognitive impairment.

A ROC curve was derived, showing high area-under-the-curve (AUC) values: 0.901 (95% CI: 0.853–0.948, $p < 0.001$), indicating good diagnostic performance in predicting the outcome.

The whole sample was then randomly split into a training cohort ($N = 118$) and a validation cohort ($N = 49$) for performing an internal validation of the model (Table 3).

From the training cohort we calculated the coefficients of the multivariable logistic regression model (Table 4).

Replication of the original regression model on the training cohort confirmed the significance of the same subset of predictors (including those with borderline significance on the whole sample) and the coefficients in Table 4 were used for setting the final equation of the score (**COGITAT—COG**nitive **I**mpairment assessment

TABLE 2 Logistic regression models evaluating predictors for cognitive impairment (MMSE <24; *N* = 167).

	Univariable analysis		Multivariable analysis		
Variable	OR (95% C.I.)	*p*	β + S.E.	OR (95% C.I.)	*p*
Age (years)	1.11 (1.06–1.15)	<0.001	0.069 ± 0.026	1.07 (1.02–1.13)	0.008
Sex (Male vs. Female)	0.81 (0.40–1.61)	0.54			
Years of education	0.81 (0.74–0.89)	<0.001			
Spontaneous Text					
Writing score—verbal and lexical skills	0.34 (0.24–0.49)	<0.001			
Writing score—spatial orientation	0.26 (0.16–0.40)	<0.001	−0.552 ± 0.293	0.58 (0.32–1.02)	0.059
Total number of words	0.91 (0.88–0.95)	<0.001	−0.044 ± 0.025	0.96 (0.91–1.00)	0.076
Percent of words in capital letters	1.01 (1.00–1.02)	0.21			
Percent of error	1.22 (1.13–1.32)	<0.001	0.133 ± 0.039	1.14 (1.06–1.23)	0.001

TABLE 3 Internal validation, random selection of cohorts.

	Normal (MMSE ≥ 24)	Cognitive impairment (MMSE < 24)	Total
Training cohort, n (%)	80 (67.8)	38 (32.2)	118 (100.0)
Validation cohort, n (%)	33 (67.3)	16 (32.7)	49 (100.0)

TABLE 4 Coefficient of multivariable logistic regression model obtained from the training cohort (*N* = 118).

	β
Age (years)	0.085
Spontaneous Text	
Writing score—spatial orientation	−0.654
Total number of words	−0.055
Percent of errors	0.127
Constant	−3.498

Through h**A**ndwri**T**ing) and subsequently the probability of having cognitive impairment (P):

$$COGITAT\ Score = 0.085 * AGE - 0.654 * WSSO - 0.055 * WORD + 0.127 * PERR$$

$$P = \frac{e^{(-3.498+0.085\mathrm{AGE}-0.654\mathrm{WSSO}-0.055\mathrm{WORD}+0.127\mathrm{PERR})}}{1+e^{(-3.498+0.085\mathrm{AGE}-0.654\mathrm{WSSO}-0.055\mathrm{WORD}+0.127\mathrm{PERR})}}$$

Where:
P: Probability of having cognitive impairment.
AGE: Age.
WSSO: Writing Score Spatial Orientation.
WORD: Total number of words.
PERR: Percent of errors.

The AUC for the training cohort (Figure 8) was 0.907 [95% CI: 0.851–0.963], $p < 0.001$, suggesting a very good predictive performance of the model.

Two optimal cut-off points for the COGITAT score were identified in the training dataset, one that maximizes specificity and the other that favors sensitivity. The first one was found to be 4.258 (specificity 95.0% and sensitivity 47.4%) and patients scoring this high or higher were classified with a 95% probability as being cognitively impaired. The second cut-off score was set to 1.959 (sensitivity 94.7% and specificity 71.2%) and patients scoring this low or lower were classified with a 95% probability as being cognitively normal.

ROC curve graphed in the validation cohort (Figure 8) with the same score showed an AUC of 0.883 [95% CI: 0.788–0.977], $p < 0.001$. The application of the first cut-off produced a sensitivity of 50.0% and a specificity of 96.9%, while the application of the second score produced a sensitivity of 81.2%, a specificity of 28.1%.

To facilitate the clinical application of these findings, an instrument was developed that automatically calculates the score and estimated probability of cognitive impairment after entering the predictor variable data. Higher scores indicate a higher probability of cognitive deterioration. The higher the score compared to the cut-off value, the greater the patient's risk, and the lower the score compared to the cut-off value, the lower the patient's risk. The Excel spreadsheet can be downloaded at the following link.[1]

4 Discussion

It has been frequently shown that handwriting conveys useful information about the cognitive state of the person who wrote it (Forbes et al., 2004; Croisile, 2005; Shenkin et al., 2008; Renier et al., 2016). However, for clinical or forensic purposes, it is necessary to

1 https://osf.io/xvt9j/?view_only=68be825d55ae467282d7268f92f06ca4

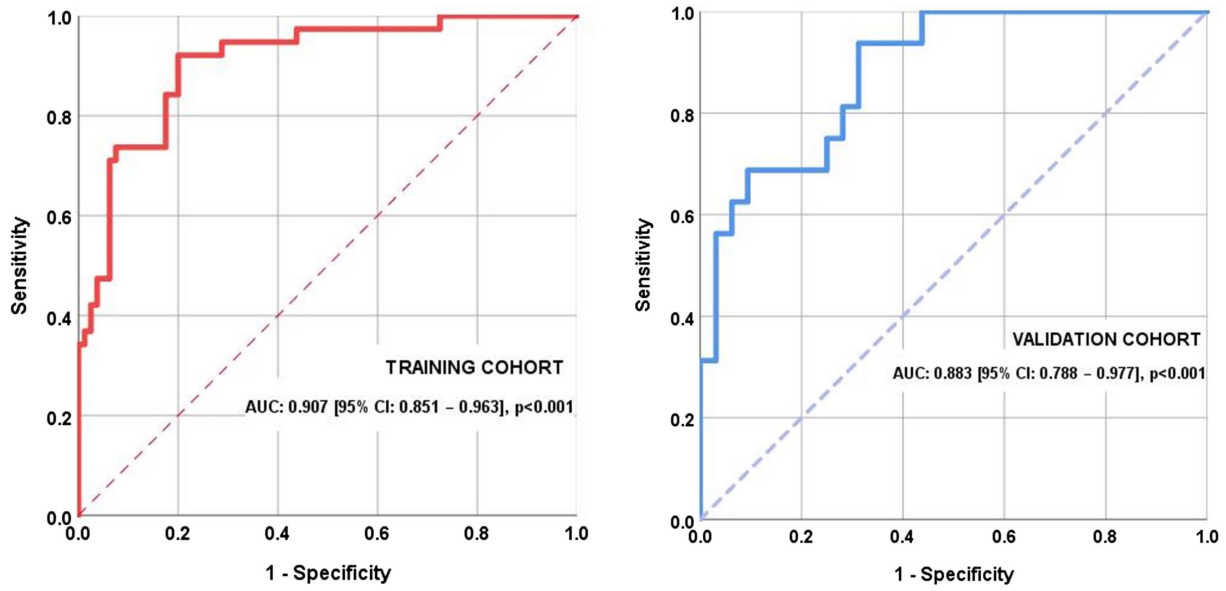


FIGURE 8
Receiver operating characteristic (ROC) curves for the training and the internal validation cohorts.

have an instrument that quantifies the alterations of pathological handwriting, rather than merely describing them. To address this need, some of us earlier created and tested the "writing score" (Fontana et al., 2008; Balestrino et al., 2012), which we briefly summarized above. Unlike other investigations, the writing score provides a numerical value that quantifies the quality of handwriting; thus, it may be used for diagnostic or forensic purposes. In a most fascinating investigation, the writing score has been used to show how King George III of England's handwriting kept deteriorating during the course of his neuropsychiatric disorder, whose exact nature is still a matter of debate (Peters, 2015).

Our present research is an attempt to further advance the quantitative analysis of handwriting. To do so, we introduced additional parameters and identified two cut-off scores to detect cognitive impairment or normal mental state with high probability.

Thus, we aimed to overcome the problem that previous research on the writing score did not yield a precise cut-off value that could reliably discriminate between normal and cognitively impaired individuals. Preliminary findings suggested that very high or very low scores on the writing score almost certainly indicate that the writer is cognitively normal or cognitively impaired, respectively (Balestrino et al., 2018). Recently, one of us conducted a proof-of-concept investigation in which we suggested that the sensitivity and specificity of the writing score could be improved by including information about how many spelling errors the writer made and how many words he/she wrote (Balestrino, 2022).

In the present investigation, we further advanced our research to identify a novel tool that, starting from the writing score, may be even more useful in identifying cognitively deteriorated people based on their handwriting. To this end, we analyzed parameters that the scientific literature suggests correlate with cognitive deterioration, such as the total number of written words (Henderson et al., 1992), the percent of spelling errors (Silveri et al., 2007), and the percent of words written (totally or partially) in capital letters (Graham, 2000).

We used only MMSE as a gage of cognitive deterioration because it is probably the most widely used test for this purpose, and it has good sensitivity and specificity in identifying cognitive deterioration (Tsoi et al., 2015). Further research is needed to investigate the relationship between the COGITAT score and specific neuropsychological domains, and additional tests investigating specific domains shall be used for this purpose. However, in a forensic validation study akin to ours, the MMSE score was found to correlate in a significant and robust way with the score obtained at a test of financial competency (Giannouli et al., 2018).

In the univariate analysis (Table 2) both age and years of education, but not sex, were different in cognitively impaired people, defined as having MMSE<24 (Measso et al., 1993; Tsoi et al., 2015). Specifically, cognitively impaired subjects were significantly older and had fewer years of education, both findings that were expected based on scientific literature data (LoGiudice and Watson, 2014; Subramaniam et al., 2015). Still in the univariate analysis (Table 2), both subsets of writing score were significantly worse, as expected, in subjects with cognitive deterioration, thus confirming the previous findings by some of us (Fontana et al., 2008; Balestrino et al., 2012). Furthermore, the total number of written words and the percent of spelling errors were significantly different between subjects with cognitive deterioration (MMSE<24) and subjects with normal cognitive status (MMSE≥24), in the sense that cognitively impaired subjects wrote significantly fewer words and made significantly more spelling mistakes, the latter parameters being significantly worse (Table 2). It was expected that those with cognitive impairment would use less words because patients with this condition are known to have lower verbal fluency (Henry et al., 2004). Similarly, the increased percentage of spelling mistakes had been previously reported in cognitively impaired people (Silveri et al., 2007). By contrast, there was no significant difference in the percent of words written in capital letters (Table 2). We offer a possible explanation for this observation by speculating that elderly individuals may employ capital letters to partially mask motor rather than cognitive dysfunction. Further research is needed to possibly confirm this hypothesis.

Then, we carried out a multivariate analysis (Table 2), which confirmed older age and percent of errors as significant predictors of cognitive impairment ($p < 0.008$ and $p < 0.001$, respectively). The "spatial orientation" subset of the writing score had borderline statistical predictive significance ($p = 0.059$), as did the total number of written words ($p = 0.076$). All these parameters were confirmed as significant at validation stage, and therefore, all of them were included for building a single score, which we called the **COG**nitive **I**mpairment assessment **T**hrough h**A**ndwri**T**ing (**COGITAT**) score.

Two different cut-off points for COGITAT score were identified and proved capable of correctly identifying cognitively impaired people and normal people, respectively, with high sensitivity and specificity.

Our study does not allow to assign with the same probability to either normal or altered cognitive status subjects having a COGITAT score between the two above cut-off scores. Further research is needed to possibly overcome this limitation. Nevertheless, we believe that the ability to judge with a statistically acceptable degree of probability a sizable number of subjects may make the COGITAT score a valuable tool in the forensic analysis of disputed wills, a field where judgment is notoriously difficult because at the time of the trial, the testator can no longer be examined, and both health records and witnesses' reports are often absent or conflicting. Moreover, and perhaps most importantly, a quantitative and statistically sound analysis of handwriting may provide valuable information about the testator's mental state right at the moment when he/she was writing the will, helping to dispel whatever uncertainty that might arise from the fact that retrospective data such as medical evaluations or testimonies are frequently far away in time from the writing of the will.

In recent years, there has been an increase in scientific interest in using handwriting in the diagnosis of Alzheimer's Disease (AD), and researchers have investigated this issue even by using machine-based approaches. Among them, Cilia and coworkers found that physical parameters of the movement carried out in either handwriting or drawing may be useful in the early diagnosis of AD (Cilia et al., 2019, 2021a, b, 2022). Those results are of great interest and relevance, however we must not forget that AD is mainly defined by a failure in cognition, while motor, sensory, or coordination deficits are less prominent early in the disease (McKhann et al., 1984, 2011; Albert et al., 2011). Thus, we believe that while machine-analyzed parameters of movement are relevant and interesting, the quantitative analysis of neuropsychological parameters of handwriting is of paramount importance in diagnosing AD based on writing characteristics. From this point of view, the COGITAT score gives utmost importance to spatial orientation of the handwriting, number of written words and percentage of errors in writing, all parameters that are relevant to cognitive deficiency.

Summing up, we suggest that handwriting analysis may be an additional tool for the diagnosis and follow up of dementia in the clinical setting. Although further research will help better defining its strengths and limitations, we believe that the COGITAT score has sufficient statistical soundness to be successfully used to help diagnose cognitive deterioration in both forensic and clinical setting.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MB: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. AB: Investigation, Writing – review & editing. NG: Investigation, Writing – review & editing. MP: Investigation, Supervision, Validation, Writing – review & editing. CR: Investigation, Writing – review & editing. PA: Writing – review & editing. LC: Investigation, Writing – review & editing. IS: Data curation, Formal analysis, Validation, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was supported by grants from the Italian Ministry of Health (Ricerca Corrente) and by the IRCCS Ospedale Policlinico San Martino (Genoa, Italy). This work was carried out within the framework of the project "RAISE - Robotics and AI for Socioeconomic Empowerment" and has been supported by European Union - NextGeneration.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1275315/full#supplementary-material

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008

Albert, M. S., and Heaton, R. K. (1988). "Intelligence testing" in *Geriatric neuropsychology*. Eds. M. S. Albert and M. B. Moss. (New York, NY, US: The Guilford Press), 13–32.

Balestrino, M. (2022). Analisi quantitativa della scrittura nella valutazione dello stato cognitivo dell'autore di un testamento. *Grafologia Medica* 1, 3–9.

Balestrino, M., Fontana, P., Terzuoli, S., Volpe, S., Inglese, M. L., and Cocito, L. (2012). Altered handwriting suggests cognitive impairment and may be relevant to posthumous evaluation. *J. Forensic Sci.* 57, 1252–1258. doi: 10.1111/j.1556-4029.2012.02131.x

Balestrino, M., Rizzetto, C., Brugnolo, A., and Girtler, N. G. (2017). Un nuovo parametro grafopatologico nella valutazione dello stato cognitivo: l'indice della scrittura. *Grafologia Medica* 1, 65–72.

Balestrino, M., Rizzetto, C., Brugnolo, A., and Girtler, N. G. (2018). Nuove ricerche sulle alterazioni della scrittura nella valutazione dello stato cognitivo. *Grafologia Medica* 1, 1–4.

Brazzelli, M., Capitani, E., Della Sala, S., Spinnler, H., and Zuffi, M. (1994). A neuropsychological instrument adding to the description of patients with suspected cortical dementia: the Milan overall dementia assessment. *J. Neurol. Neurosurg. Psychiatry* 57, 1510–1517. doi: 10.1136/jnnp.57.12.1510

Cilia, N. D., D'Alessandro, T., De Stefano, C., and Fontanella, F. (2022). Deep transfer learning algorithms applied to synthetic drawing images as a tool for supporting Alzheimer's disease prediction. *Mach. Vis. Appl.* 33:49. doi: 10.1007/s00138-022-01297-8

Cilia, N. D., D'Alessandro, T., De Stefano, C., Fontanella, F., and Molinara, M. (2021a). From online handwriting to synthetic images for Alzheimer's disease detection using a deep transfer learning approach. *IEEE J. Biomed. Health Inform.* 25, 4243–4254. doi: 10.1109/JBHI.2021.3101982

Cilia, N. D., De Stefano, C., Fontanella, F., and Freca, A. S. D. (2021b). Feature selection as a tool to support the diagnosis of cognitive impairments through handwriting analysis. *IEEE Access* 9, 78226–78240. doi: 10.1109/ACCESS.2021.3083176

Cilia, N. D., De Stefano, C., Fontanella, F., Molinara, M., and Scotto Di Freca, A. (2019). "Handwriting analysis to support Alzheimer's disease diagnosis: a preliminary study" in *Computer analysis of images and patterns lecture notes in computer science*. eds. M. Vento and G. Percannella (Cham: Springer International Publishing), 143–151.

Croisile, B. (2005). Writing, aging and Alzheimer's disease. *Psychol. Neuropsychiatr. Vieil.* 3, 183–197.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Fontana, P., Dagnino, F., Cocito, L., and Balestrino, M. (2008). Handwriting as a gauge of cognitive status: a novel forensic tool for posthumous evaluation of testamentary capacity. *Neurol. Sci.* 29, 257–261. doi: 10.1007/s10072-008-0977-3

Forbes, K. E., Shanks, M. F., and Venneri, A. (2004). The evolution of dysgraphia in Alzheimer's disease. *Brain Res. Bull.* 63, 19–24. doi: 10.1016/j.brainresbull.2003.11.005

Giannouli, V., Stamovlasis, D., and Tsolaki, M. (2018). Exploring the role of cognitive factors in a new instrument for elders' financial capacity assessment. *J. Alzheimers Dis.* 62, 1579–1594. doi: 10.3233/JAD-170812

Graham, N. L. (2000). Dysgraphia in dementia. *Neurocase* 6, 365–376. doi: 10.1080/13554790008402708

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2018). Multivariate data analysis (8th ed.). United Kingdom: Cengage Learning.

Henderson, V. W., Buckwalter, J. G., Sobel, E., Freed, D. M., and Diz, M. M. (1992). The agraphia of Alzheimer's disease. *Neurology* 42, 776–784. doi: 10.1212/wnl.42.4.776

Henry, J. D., Crawford, J. R., and Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia* 42, 1212–1222. doi: 10.1016/j.neuropsychologia.2004.02.001

LoGiudice, D., and Watson, R. (2014). Dementia in older people: an update. *Intern. Med. J.* 44, 1066–1073. doi: 10.1111/imj.12572

Luzzatti, C., Laiacona, M., and Agazzi, D. (2003). Multiple patterns of writing disorders in dementia of the Alzheimer type and their evolution. *Neuropsychologia* 41, 759–772. doi: 10.1016/s0028-3932(02)00328-7

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease. *Neurology* 34:939. doi: 10.1212/WNL.34.7.939

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005

Measso, G., Cavarzeran, F., Zappalà, G., Lebowitz, B. D., Crook, T. H., Pirozzolo, F. J., et al. (1993). The mini-mental state examination: normative study of an Italian random sample. *Dev. Neuropsychol.* 9, 77–85. doi: 10.1080/87565649109540545

Peters, T. (2015). FitzPatrick lecture: King George III and the porphyria myth - causes, consequences and re-evaluation of his mental illness with computer diagnostics. *Clin. Med. (Lond.)* 15, 168–172. doi: 10.7861/clinmedicine.15-2-168

Renier, M., Gnoato, F., Tessari, A., Formilan, M., Busonera, F., Albanese, P., et al. (2016). A correlational study between signature, writing abilities and decision-making capacity among people with initial cognitive impairment. *Aging Clin. Exp. Res.* 28, 505–511. doi: 10.1007/s40520-016-0549-y

Shenkin, S. D., Starr, J. M., Dunn, J. M., Carter, S., and Deary, I. J. (2008). Is there information contained within the sentence-writing component of the mini mental state examination? A retrospective study of community dwelling older people. *Int. J. Geriatr. Psychiatry* 23, 1283–1289. doi: 10.1002/gps.2066

Silveri, M. C., Reali, G., Jenner, C., and Puopolo, M. (2007). Attention and memory in the preclinical stage of dementia. *J. Geriatr. Psychiatry Neurol.* 20, 67–75. doi: 10.1177/0891988706297469

Subramaniam, M., Chong, S. A., Vaingankar, J. A., Abdin, E., Chua, B. Y., Chua, H. C., et al. (2015). Prevalence of dementia in people aged 60 years and above: results from the WiSE study. *J. Alzheimers Dis.* 45, 1127–1138. doi: 10.3233/JAD-142769

Tamburini, A. (1911). Questioni medico-forensi relative alla Paralisi generale progressiva specialmente in relazione alla capacità di testare. *Parte I. Rivista Sperimentale di Freniatria* 37, 456–480.

Tsoi, K. K. F., Chan, J. Y. C., Hirai, H. W., Wong, S. Y. S., and Kwok, T. C. Y. (2015). Cognitive tests to detect dementia: a systematic review and Meta-analysis. *JAMA Intern. Med.* 175, 1450–1458. doi: 10.1001/jamainternmed.2015.2152

frontiers | Frontiers in Psychology

TYPE Opinion
PUBLISHED 08 December 2022
DOI 10.3389/fpsyg.2022.1030439

OPEN ACCESS

EDITED BY
Yvonne H. A. Bouman,
Stichting Transfore, Netherlands

REVIEWED BY
Andrea Lavazza,
Centro Universitario
Internazionale, Italy

*CORRESPONDENCE
Diego Borbón
diego.borbon01@est.uexternado.edu.co

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Forensic and Legal Psychology, a section of the journal Frontiers in Psychology

RECEIVED 28 August 2022
ACCEPTED 23 November 2022
PUBLISHED 08 December 2022

CITATION
Díaz Soto JM and Borbón D (2022) Neurorights vs. neuroprediction and lie detection: The imperative limits to criminal law.
Front. Psychol. 13:1030439.
doi: 10.3389/fpsyg.2022.1030439

Neurorights vs. neuroprediction and lie detection: The imperative limits to criminal law

José Manuel Díaz Soto[1†] and Diego Borbón[2*†]

[1]Department of Criminal Law and Criminology, Universidad Externado de Colombia, Bogotá, Colombia, [2]NeuroRights Research Group, The Latin American Observatory of Human Rights and Enterprises, Universidad Externado de Colombia, Bogotá, Colombia

KEYWORDS

neurorights, neuroethics, neurolaw, neuroscience, neurotechnologies, human rights, criminology, neuromarkers

This paper analyzes the suitability of neurorights to limit the use of neuroprediction and lie detection neurotechnologies. We argue that some of their applications in criminal proceedings should be prohibited as they are severely intrusive to mental privacy and contrary to the dignity of the person. In that sense, we discuss whether neurorights can offer greater protection than current fundamental rights. We suggest that, as they have been conceived, neurorights may offer reduced protection and they should be framed to offer a true limit to the substantial barrier that is our mind and dignity. On the other hand, current human rights should be interpreted in such a way as to respect the dignity of the accused in criminal proceedings.

A brief overview on neurorights

Since 2017, an innovative discussion framework has been created to protect people from potential abusive uses of neurotechnologies. Based on neuroethics, researchers Ienca and Andorno (2017) propose to create four neuro-specific human rights: cognitive liberty, psychological continuity, mental privacy, and integrity. Likewise, Yuste et al. (2017), and nowadays the NeuroRights Foundation, promotes the creation of five NeuroRights: the right to free will, mental privacy, personal identity, fair access to mental augmentation, and protection from bias (NeuroRights Foundation., 2022).

Furthermore, the reception of those initiatives has been such that on April 12, 2021, the Chilean Congress approved a Constitutional Reform endorsing the rights to physical and mental integrity (IACHR., 2022a).[1]

1 On September 4, 2022, the people of Chile voted to deny the constitutional referendum that sought to replace the 1980 Constitution. The denial of the referendum did not affect the reform of neurorights, since it was incorporated into article 19 of the old Constitution, which remains effective as of October 2022. At the moment, the Chilean Government is preparing a new constitutional convention whose results will again be voted on by the Chilean people.

More recently, in the United Nations, resolution A/HRC/51/L.3 on neurotechnology and human rights was adopted (Human Rights Council HCR-UN, 2022).[2]

Neuroprediction and lie detection

Research in neurotechnology is allowing us to have a better understanding of the brain and enabling technology for new treatment options and better quality of life (Stieglitz, 2021). On the other hand, in clinical translational science, neuroscientists are looking to apply this technology to assess, treat, and better understand complex socioemotional processes that underlie many forms of psychopathology (White et al., 2015).

In that direction, criminal justice systems are not far behind as criminal law cares about human behavior and specially the mind (Greely and Farahany, 2018). Some authors argue that neuroscience has the power to change the criminal justice systems, and that society would benefit from active collaboration between sciences (Altimus, 2017). Although neurotechnologies could provide useful tools for the judicial system, some pose numerous ethical challenges that hinder their implementation (Coronado, 2021; Borbón and Borbón, 2022).

Neuroprediction and lie detection neurotechnologies are a clear example of why it has become so essential to discuss neurorights. Neuroprediction comprises the use of structural or functional variables of the brain for medical and behavioral predictions (Morse, 2015). In recent years, artificial intelligence (A.I.) and neurotechnologies are being used to improve the accuracy of risk assessment tools (Kehl et al., 2017; Kiehl et al., 2018; Tortora et al., 2020), using neuroimaging data to predict recidivism and criminal behavior (Aharoni et al., 2013; Kiehl et al., 2018; Delfin et al., 2019). In that sense, findings in neurocriminology have managed to identify structural and functional deficits in the brain and their relationship with antisocial behavior (Bellesi et al., 2019; Katzin et al., 2020; Ruiz and Muñoz, 2021; Borbón, 2022). These empirical results could be used by neuroprediction algorithms to identify the neuro markers that influence deviant behavior.

Parallel to advances in neuroprediction, the use of neurotechnologies for lie detection has been explored. Recent efforts to detect lies have focused on measures in the brain, believing that these may be more reliable than physiological responses in other parts of the body. Some studies used tools such as positron emission tomography, electroencephalography, functional near-infrared spectroscopy, and functional magnetic resonance imaging (Norman et al., 2006; Greely and Illes, 2007; Abootalebi et al., 2009; Langleben and Moriarty, 2013; Farah et al., 2014; Li et al., 2018).

However, these technologies are far from perfect and remain open to the subjectivity of those who interpret the results obtained while being hardly validated or reliable (Greely, 2009; Lowenberg, 2010; Schauer, 2010). Furthermore, the deep ignorance that we still have about the brain, given that neuroscience is a developing science, implies that we must proceed with caution, far from the current "neuro hype" (Bigenwald and Chambon, 2019; Morse, 2019). Neuroprediction and lie detection are not able to offer proof standards of certainty, but only of probability. In this sense, its use to serve as evidence for prosecution, or even to extend the length of criminal sentences, should be strictly regulated.

Neurorights: Progressivity and non-regressivity

In terms of human rights, the principle of progressivity is recognized in the preamble of the Universal Declaration of Human Rights and expressly protected by the Inter-American Human Rights System (IACHR., 1993). The principle of progressivity entails an obligation of non-regression, which implies that the progress made in the field of human rights is irreversible, it can always be expanded but never reduced (Cunego, 2016). Under that scope, the fundamental reason to create neurorights should be to offer citizens a greater scope of protection. However, some interpretations that could be extended to the initial proposal of neurorights concern us because they may end up transgressing these principles.

The initial paper by Ienca and Andorno (2017) proposes to create a right to cognitive liberty, which implies being able to reject neurotechnological applications in their negative facet. However, throughout the text they recognize that neurorights, like any of the current fundamental rights, are relative and that in certain circumstances they could be reduced substantially. Regarding the neuroright to mental privacy, they maintain that the collection, use, and disclosure of private information is permissible when the public interest is at stake (Ienca and Andorno, 2017). Also, considering the painless nature of brain scans, they suggest that there could be good reasons for thinking that their nonconsensual use would be justified, with a court warrant, under special circumstances (Ienca and Andorno, 2017). This, we believe, would include the debate on neuroprediction and lie detection. In addition, when dealing with a subject as sensitive as "moral enhancement",

2 This resolution is a great step forward, as the Human Rights Council will now be in charge of studying the opportunities and challenges of neurotechnology in relation to the promotion and protection of human rights. It will also allow to know the opinions and contributions of the countries and of the academy. To learn about the most recent academic research on neurorights, we recommend consulting the recent Research Topic published with Frontiers (García-López et al., 2021), which includes important works on this subject (Borbón and Borbón, 2021; Ienca, 2021; Inglese and Lavazza, 2021; Larrivee, 2021; Schleim, 2021; Wajnerman, 2021). See also (Collecchia, 2021; Goering et al., 2021; Fyfe et al., 2022; Herrera-Ferrá et al., 2022; Vidal, 2022).

Ienca and Andorno (2017) suggest that it is possible to argue on utilitarian grounds that violations of the right to mental integrity could be allowed for persistent violent offenders, but they prefer not to take a definitive position on that issue.

Our argumentation does not ignore that human rights in general are relative and that in the judicial practice they are weighted against other rights. On the contrary, we intend to bring the discussion closer to the insuperable principle and rule of human dignity, as well as to advocate for absolute prohibitions when neurotechnology is used against the person.

In that sense, revealing the neural correlates of individual thoughts and feelings can be seen as an intrusion into privacy. The violations of freedom would be even more evident in the uses of neuroprediction for sentencing or punitive purposes, or lie detection for prosecution. In those cases, the person is taken simply as a means or an object of a criminal proceeding. Should potential offenders of the Law be forced to undergo neuroimaging tests against their will, under the pretext of public safety? (Coppola, 2018). Faced with the risks implied by technological advances, and the increasingly intrusive mechanisms in privacy and the free decision of people, we advocate for the rigorous regulation of those coercive uses.

We think that inordinate reliance on neurotechnologies and A.I. could bias judicial decisions, and even put an end to the purpose of having a judicial system and criminal proceedings at all. We certainly agree that excessive and unreasonable reliance on those technologies should be avoided (Tortora et al., 2020). Proceeding in this way raises serious ethical implications (Nadelhoffer and Sinnott-Armstrong, 2012; Tortora et al., 2020) and would undermine the rights of the accused, the prohibition of self-incrimination, the presumption of innocence, the right to refuse medical treatment, to due process, defense and contradiction, the culpability principle and the *mens rea*. But especially, we consider that insisting on coercively implementing these technologies violates human dignity, a guiding principle in any democratic society that respects the rule of law. In the end, neurorights would not be complying with the principles of progressivity and non-regression. Instead, they would be turning into ambiguous clauses for the punitive power of the State.

Between neurorights and human rights

Even when criminal law has limits, such as the weights imposed by fundamental rights, it always retains intrinsic brutality, which makes its moral legitimacy problematic and uncertain (Ferrajoli, 1995). This brutality would be exacerbated if the State acquires new neurotechnological tools for punitive purposes (Borbón and Borbón, 2022). In this direction, there is no doubt that neuroscientific progress must be regulated. However, what is not so clear is whether neurorights are the best alternative.

Bublitz (2022) has criticized the inflation of rights and their resulting devaluation. This author affirms that there has not been a real academic debate, nor has it been explained why the current rights are insufficient. Borbón and Borbón (2021) have presented arguments in that same direction affirming that the current human rights already protect freedom, consent, equality, integrity, privacy, and others. In this sense, they propose that it would be much more necessary to propose legal and conventional regulations that are much clearer, precise, and extensive (Borbón and Borbón, 2021). Likewise, Ienca (2021) asserts that the relatively sporadic presence of neurorights in the academic literature poses a risk of semantic-normative ambiguity and conceptual confusion; López and Madrid (2021) argue that the legal consequences would be disastrous if neurorights are normatively manifested in a frivolous or imprecise way; and Fins (2022) states that the current Chilean neurorights reforms are vague and premature.

Conclusion and proposal

We propose further academic and political deliberation to reach a consensus on the legal instruments necessary to effectively regulate the advancement of neuroscience. Neurotechnologies used coercively for neuroprediction or lie detection should be extensively regulated or even prohibited if they are used for punitive purposes, criminal prosecution, and the limitation of freedom.

In this sense, the direction proposed by Ruiz and Muñoz (2021) seems relevant to us. They re-define neuroprediction into "neuroprevention", assuming a non-reductionist position to reach an early detection of risk factors that allows timely interventions through the application of training practices in cognitive skills aimed at reducing criminogenic factors. In general, the intent is to balance public safety with a scientifically based opportunity to reintegrate the person into society (Ruiz and Muñoz, 2021).

On the other hand, we consider that human dignity is a solid foundation to build the legal regulations of neuroscience. For the Law, human dignity is a wideranging constitutional value, gathering a whole array of protections, benefits, structures, empowerments, entitlements, institutions, forms of respect, and equalizations going well beyond a list of individual rights (Waldron, 2019). On our view, human dignity is not ponderable since it will never be conventionally admissible to treat the other as a simple object or means. Rights in general can be subject to judicial interference and even be susceptible to profound limitations, except in the case of human dignity. For those same reasons, for example, it will never be valid to torture a criminal, even when States restrict other rights.

In this sense, to protect the substantive legal grounds of freedom, integrity, privacy, or equality, any legislative proposal must be based on the ever-valid principle of human dignity. We

advocate for a strong concept of human dignity, which would imply a strict regulation on the use of neurotechnologies in criminal proceedings. This could be similar to the conventional prohibition of cruel, inhuman or degrading punishment.

All things considered, we maintain that, if neurorights are considered necessary, they should be enshrined in such a way that they prevent States from using technologies without the consent and for punitive purposes. In the same way, current fundamental rights, along with new specific and clear international treaties, must be aimed at guaranteeing the principle of human dignity. Neuroscience, in this sense, can be used in ways that respect people's rights, even as valid defensive strategies in criminal proceedings, with the person's due informed consent. In the end, our call is to firmly give the battle to preserve the most intimate personal corner: our mind.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

Some ideas of this manuscript have appeared previously in Spanish in Borbón and Borbón (2022).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abootalebi, V., Moradi, M., and Khalilzadeh, M. (2009). A new approach for EEG feature extraction in P300-based lie detection. *Comput. Methods Programs Biomed.* 94, 48–57. doi: 10.1016/j.cmpb.2008.10.001

Aharoni, E., Vincent, G., Harenski, C., Calhoun, V., Armstrong, W., Gazzangia, M., et al. (2013). Neuroprediction of future rearrest. *Proc. Nat. Acad. Sci.* 110, 6223–6228. doi: 10.1073/pnas.1219302110

Altimus, C. M. (2017). Neuroscience has the power to change the criminal justice system. *eNeuro.* 3, ENEURO.0362-16.2016. doi: 10.1523/ENEURO.0362-16.2016

Bellesi, G., Barker, E. D., Brown, L., and Valmaggia, L. (2019). Pediatric traumatic brain injury and antisocial behavior: are they linked? A Systematic Review. *Brain Inj.* 33, 1272–1292. doi: 10.1080/02699052.2019.1641621

Bigenwald, A., and Chambon, V. (2019). Criminal responsibility and neuroscience: no revolution yet. *Front. Psychol.* 10, 1406. doi: 10.3389/fpsyg.2019.01406

Borbón, D. (2022). Neurosociology and penal neuroabolitionism: rethinking justice with neuroscience. *Front. Sociol.* 7, 814338. doi: 10.3389/fsoc.2022.814338

Borbón, D., and Borbón, L. (2021). A critical perspective on NeuroRights: comments regarding ethics and law. *Front. Hum. Neurosci.* 15, 703121. doi: 10.3389/fnhum.2021.703121

Borbón, D. and Borbón, L. (2022). *NeuroDerechos Humanos y Neuroabolicionismo Penal: Comentarios críticos frente a la neuropredicción y la detección de mentiras.* Cuestiones Constitucionales: Revista Mexicana de Derecho Constitucional. p. 46. doi: 10.22201/iij.24484881e.2022.46.17047

Bublitz, J. C. (2022). Novel neurorights: from nonsense to substance. *Neuroethics.* 15, 7. doi: 10.1007/s12152-022-09481-3

Collecchia, G. (2021). Neurotecnologie e neurodiritti digitali: la privacy mentale [Neurotechnologies and neurorights: mental privacy.]. *Recenti progressi in medicina.* 112, 343–346. doi: 10.1701/3608.35871

Coppola, F. (2018). Mapping the brain to predict antisocial behavior: new frontiers in neurocriminology,'new challenges for criminal justice. *U.C.L. J. Law Jurisprudence.* 1, 106–110. doi: 10.14324/111.2052-1871.124

Coronado, B. (2021). *Aportación de la neurociencia a la predicción de la conducta violenta.* España, Madrid: Universidad Pontificia Comillas. Available online at: https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/46609/TFG%20Coronado%20Gonzalez%2C%20Beatriz%20de%20Silva.pdf?sequence=1&isAllowed=y (accessed November 29, 2022).

Cunego, A. (2016). *Development Policies Evaluation Under the Human Rights-Based Approach*. Madrid: Berg Institute. doi: 10.2307/j.ctvkwnnrz

Delfin, C., Krona, H., Andine, P., Ryding, E., Wallinins, M., and Holvander, B. (2019). Prediction of recidivism in a long-term follow-up of forensic psychiatric patients: Incremental effects of neuroimaging data. *PLoS ONE.* 14, 1–21. doi: 10.1371/journal.pone.0217127

Farah, M., Hutchinson, J. B., Phelps, E., and Wagner, A. D. (2014). Functional MRI-based lie detection: scientific and societal challenges. *Nat. Rev. Neurosci.* 15, 123–131. doi: 10.1038/nrn3665

Ferrajoli, L. (1995). *Derecho y razón, teoría del garantismo penal.* Madrid, España: Editorial Trotta, S. A.

Fins, J. (2022). The unintended consequences of chile's neurorights constitutional reform: Moving beyond negative rights to capabilities. *Neuroethics.* 15, 26. doi: 10.1007/s12152-022-09504-z

Fyfe, S., Lanphier, E., and Peterson, A. (2022). Neurorights for incarcerated persons: should we curb inflation?. *AJOB Neurosci.* 13, 165–168. doi: 10.1080/21507740.2022.2082585

García-López, E., Muñoz, J., M., and Andorno, R. (2021). Editorial: neurorights and mental freedom: emerging challenges to debates on human dignity and neurotechnologies. *Front. Hum. Neurosci.* 15, 823570. doi: 10.3389/fnhum.2021.823570

Goering, S., Klein, E., Specker Sullivan, L., Wexler, A., Agüera y Arcas, B., Bi, G., et al. (2021). Recommendations for responsible development and application of neurotechnologies. *Neuroethics.* 14, 365–386. doi: 10.1007/s12152-021-09468-6

Greely, H. (2009). Law and the revolution in neuroscience: an early look at the field. *Akron Law Rev.* 42, 687–715. Available online at: https://ideaexchange.uakron.edu/akronlawreview/vol42/iss3/2 (accessed November 29, 2022).

Greely, H., and Illes, J. (2007). Neuroscience-based lie detection: the urgent need for regulation. *Am. J. Law Med.* 33, 377–431. doi: 10.1177/009885880703300211

Greely, H. T., and Farahany, N. A. (2018). Neuroscience and the criminal justice system. *Ann. Rev. Criminol.* 2, 1. doi: 10.1146/annurev-criminol-011518-024433

Herrera-Ferrá, K., Muñoz, J. M., Nicolini, H., Saruwatari Zavala, G., and Martínez Bullé Goyri, V. M. (2022). Contextual and cultural perspectives on neurorights: reflections toward an international consensus. *AJOB Neurosci.* 1–9. doi: 10.1080/21507740.2022.2048722

Human Rights Council HCR-UN. (2022). *Resolution A/HRC/51/L.3*. Available online at: https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/51/L.3

IACHR. (1993). *Informe anual: Campos en los cuales han de tomarse medidas para dar mayor vigencia a los derechos humanos, de conformidad con la Declaración Americana de los Derechos y Deberes del Hombre y la Convención Americana sobre Derechos Humanos. IACHR.* Available online at: https://www.cidh.oas.org/annualrep/93span/cap.v.htm

IACHR. (2022a). *V Informe anual de la relatoría especial sobre derechos económicos, sociales, culturales y ambientales (redesca) de la comisión interamericana de derechos humanos (cidh), 2021.* Available online at: https://www.oas.org/es/cidh/docs/anual/2021/capitulos/redesca-es.PDF

Ienca, M. (2021). On neurorights. *Front. Hum. Neurosci.* 15, 701258. doi: 10.3389/fnhum.2021.701258

Ienca, M., and Andorno, R. (2017). Towards new human rights in the age of neuroscience and neurotechnology. *Life Sci. Soc. Policy.* 13, 1–27. doi: 10.1186/s40504-017-0050-1

Inglese, S., and Lavazza, A. (2021). What should we do with people who cannot or do not want to be protected from neurotechnological threats? *Front. Hum. Neurosci.* 15, 703092. doi: 10.3389/fnhum.2021.703092

Katzin, S., Andin,é, P., Hofvander, B., Billstedt, E., and Wallinius, M. (2020). Exploring traumatic brain injuries and aggressive antisocial behaviors in young male violent offenders. *Front. Psychiatry.* 11, 507196. doi: 10.3389/fpsyt.2020.507196

Kehl, D., Guo, P., and Kessler, S. (2017). *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.* Available online at: https://cyber.harvard.edu/publications/2017/07/Algorithms.

Kiehl, K. A., Anderson, N. E., Aharoni, E., Maurer, J. M., Harenski, K. A., Rao, V., et al. (2018). Age of gray matters: neuroprediction of recidivism. *Neuroimage.* 19, 813–823. doi: 10.1016/j.nicl.2018.05.036

Langleben, D. D., and Moriarty, J. C. (2013). Using brain imaging for lie detection: where science, law and research policy collide. *Psychol Public Policy Law.* 19, 222–234. doi: 10.1037/a0028841

Larrivee, D. (2021). Values evolution in human machine relations: grounding computationalism and neural dynamics in a physical a priorism of nature. *Front. Hum. Neurosci.* 15, 649544. doi: 10.3389/fnhum.2021.649544

Li, F., Zhu, H., Xu, J., Gao, Q., Guo, H., Wu, S., et al. (2018). Lie detection using fNIRS monitoring of inhibition-related brain regions discriminates infrequent but not frequent liars. *Front. Hum. Neurosci.* 12, 71. doi: 10.3389/fnhum.2018.00071

López, P., and Madrid, R. (2021). Sobre la conveniencia de incluir los neuroderechos en la Constitución o en la ley. *Revista Chilena De Derecho Y Tecnología.* 10, 53–76. doi: 10.5354/0719-2584.2021.56317

Lowenberg, K. (2010). *fMRI Lie Detection Fails Its First Hearing on Reliability*. Stanford Law School. Available online at: https://law.stanford.edu/2010/06/01/fmri-lie-detection-fails-its-first-hearing-on-reliability/ (accessed November 29, 2022).

Morse, S. (2015). Neuroprediction: new technology, old problems. *Bioethica Forum.* 8, 128–129. doi: 10.24894/BF.2015.08039

Morse, S. (2019). *Neurohype and the Law: A Cautionary Tale*. Faculty Scholarship at Penn Law. 2006. Available online at: https://scholarship.law.upenn.edu/faculty_scholarship/2006 (accessed November 29, 2022).

Nadelhoffer, T., and Sinnott-Armstrong, W. (2012). Neurolaw and neuroprediction: potential promises and perils. *Philosophy Compass.* 7, 631–642. doi: 10.1111/j.1747-9991.2012.00494.x

NeuroRights Foundation. (2022). *The Five Ethical NeuroRights*. Available online at: https://neurorightsfoundation.org/ (accessed November 29, 2022).

Norman, K., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005

Ruiz, A., and Muñoz, J. (2021). Neuroprevention: developing legal policies in risk assessment without aspiring to predict crime. *J. Sci. Law.* 9, 1−5. doi: 10.35005/y991-wv96

Schauer, F. (2010). Neuroscience, lie-detection, and the law. *Trends Cogn. Sci.* 14, 101–103. doi: 10.1016/j.tics.2009.12.004

Schleim, S. (2021). Neurorights in history: a contemporary review of José M. R. Delgado's "Physical Control of the Mind" (1969) and Elliot S. Valenstein's "Brain Control" (1973). *Front. Hum. Neurosci.* 15, 703308. doi: 10.3389/fnhum.2021.703308

Stieglitz, T. (2021). Why Neurotechnologies? about the purposes, opportunities and limitations of neurotechnologies in clinical applications. *Neuroethics*. 14, 5–16 doi: 10.1007/s12152-019-09406-7

Tortora, L., Meynen, G., Bijlsma, J., Tronci, E., and Ferracuti, S. (2020). Neuroprediction and A.I. in forensic psychiatry and criminal justice: a neurolaw perspective. *Front. Psychol.* 11, 1–9. doi: 10.3389/fpsyg.2020.00220

Vidal, C. (2022). Neurotechnologies under the Eye of Bioethics. *eNeuro*. 9, ENEURO.0072-22.2022. doi: 10.1523/ENEURO.0072-22.2022

Wajnerman, A. P. (2021). Is mental privacy a component of personal identity? *Front. Hum. Neurosci.* 15, 773441. doi: 10.3389/fnhum.2021.773441

Waldron, J. (2019). *Human Dignity: A Pervasive Value*. NYU School of Law, Public Law Research. p. 20–46. doi: 10.2139/ssrn.3463973

White, S. W., Richey, J. A., Gracanin, D., Bell, M. A., LaConte, S., Coffman, M., et al. (2015). The promise of neurotechnology in clinical translational science. *Clini. Psychol. Sci.* 3, 797–815. doi: 10.1177/2167702614549801

Yuste, R., Georing, S., Arcas, A., Bi, G., Carmena, J., Carter, A., et al. (2017). Four ethical priorities for neurotechnologies and AI. *Nature.* 551, 159–163. doi: 10.1038/551159a

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 14 June 2024
DOI 10.3389/fpsyt.2024.1330993

OPEN ACCESS

EDITED BY
Joan E. Van Horn,
De Forensische Zorgspecialisten, Netherlands

REVIEWED BY
Yvonne H. A. Bouman,
Stichting Transfore, Netherlands
Valeria Sebri,
European Institute of Oncology (IEO), Italy

*CORRESPONDENCE
Peter C. de Looff
peterdelooff@gmail.com

†Deceased

RECEIVED 31 October 2023
ACCEPTED 02 May 2024
PUBLISHED 14 June 2024

CITATION
de Looff PC, Noordzij ML, Nijman HLI, Goedhard L, Bogaerts S and Didden R (2024) Putting the usability of wearable technology in forensic psychiatry to the test: a randomized crossover trial.
Front. Psychiatry 15:1330993.
doi: 10.3389/fpsyt.2024.1330993

Putting the usability of wearable technology in forensic psychiatry to the test: a randomized crossover trial

Peter C. de Looff[1,2,3,4]*, Matthijs L. Noordzij[5], Henk L. I. Nijman[1,2†], Laurette Goedhard[2], Stefan Bogaerts[2,4] and Robert Didden[1,6]

[1]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands, [2]Science and Treatment Innovation, Fivoor, Rotterdam, Netherlands, [3]National Expercentre Intellectual Disabilities and Severe Behavioral Problems, De Borg, Bilthoven, Netherlands, [4]Department of Developmental Psychology, Tilburg University, Tilburg, Netherlands, [5]Department of Psychology, Health and Technology, Twente University, Enschede, Netherlands, [6]Trajectum, Specialized and Forensic Care, Zwolle, Netherlands

Introduction: Forensic psychiatric patients receive treatment to address their violent and aggressive behavior with the aim of facilitating their safe reintegration into society. On average, these treatments are effective, but the magnitude of effect sizes tends to be small, even when considering more recent advancements in digital mental health innovations. Recent research indicates that wearable technology has positive effects on the physical and mental health of the general population, and may thus also be of use in forensic psychiatry, both for patients and staff members. Several applications and use cases of wearable technology hold promise, particularly for patients with mild intellectual disability or borderline intellectual functioning, as these devices are thought to be user-friendly and provide continuous daily feedback.

Method: In the current randomized crossover trial, we addressed several limitations from previous research and compared the (continuous) usability and acceptance of four selected wearable devices. Each device was worn for one week by staff members and patients, amounting to a total of four weeks. Two of the devices were general purpose fitness trackers, while the other two devices used custom made applications designed for bio-cueing and for providing insights into physiological reactivity to daily stressors and events.

Results: Our findings indicated significant differences in usability, acceptance and continuous use between devices. The highest usability scores were obtained for the two fitness trackers (Fitbit and Garmin) compared to the two devices employing custom made applications (Sense-IT and E4 dashboard). The results showed similar outcomes for patients and staff members.

Discussion: None of the devices obtained usability scores that would justify recommendation for future use considering international standards; a finding

10.3389/fpsyt.2024.1330993

that raises concerns about the adaptation and uptake of wearable technology in the context of forensic psychiatry. We suggest that improvements in gamification and motivational aspects of wearable technology might be helpful to tackle several challenges related to wearable technology.

KEYWORDS

digital mental health, forensic psychiatry, intellectual disability, wearable technology, system usability scale, technology acceptance model, extended confirmation model

1 Introduction

Current treatments for forensic psychiatric patients are generally effective, but effect sizes tend to be small to moderate across various relevant outcomes. A meta-analysis (1) on recidivism risk in violent offenders reported that treatments significantly reduced both non-violent and violent recidivism, reporting an odds reduction in (violent) reoffending of approximately 30–35%. Multimodal treatments (intensive cognitive behavioral therapy) were found to be most effective, with significant positive effects on recidivism, although the authors note that the overall effectiveness of psychological treatment on recidivism is small (1). Small effect sizes were also found for the treatment of personality disorders and aggression (2–4), even when e-health innovations were considered (5). Patients in forensic psychiatry often suffer from various (mental) health conditions that severely affect functioning. Examples include, but are not limited to, substance use disorders, depression, bipolar disorder, and schizophrenia (6).

The relatively limited effectiveness of forensic psychiatric treatments can be attributed to the multifaceted nature of violent and aggressive behavior, and complex interactions between psychological, social, environmental, biological and neurophysiological factors (2, 7–9). Special need populations, such as those with mild intellectual disability or borderline intellectual functioning, present additional challenges. Research shows that treatments need to be adjusted to their intellectual and adaptive ability and special needs (10, 11). Personalized and continuous (24/7) treatments tailored to individual needs have been proposed as a promising opportunity to increase the efficacy of current interventions, and improve treatment outcomes (5, 12, 13). Wearable devices show particular promise in transitioning from relatively brief and standardized treatments to continuous and personalized care (14).

Wearable biosensors, such as wristbands, headbands, chest straps and patches (15) provide insight and feedback on physiological signals (e.g., heart rate [variability], breathing rate, temperature, movement, skin conductance). These devices are increasingly being used in the general population and in (mental) healthcare settings to monitor and improve mental and physical health (16). The physiological signals serve as the foundation for creating composite scores or digital biomarkers (17) such as sleep, physical activity or stress indices, which are recognized as transdiagnostic markers of (mental) health and disorders (18–20). The digital biomarkers are typically being created with machine learning methods and artificial intelligence (21). Based on these biomarkers, recommender systems might provide recommendations for personalized interventions, which will have significant impact on the use of the technology in healthcare and the relationship between patients and their healthcare professional (21–23). Recent meta-analyses have resulted in small to medium effect sizes of wearable technology, including fitness trackers, activity trackers and biofeedback devices on stress, sleep, physical activity, depression, emotional and behavioral self-regulation, cardiovascular functioning, and metabolic syndrome (14, 16, 20, 24–31). However, the implementation of wearable technology in forensic psychiatry faces challenges, including limited technology readiness, acceptance, usability of the devices, continuous use of the devices, privacy concerns and data management (14, 32–35).

Three aspects related to the implementation of wearables (usability, acceptance, and continuous use) are deemed crucial for the adoption of the technology (33). Usability serves as a proxy for the ease of use of the devices, and higher scores have been associated with increased adoption and recommendation of new technology (36). Similar findings have been reported for the subjective acceptance of new technology, which consists of two main determinants that relate to the perceived usefulness and ease of use. Finally, continuous use is a proxy for user satisfaction and extended use intention following the purchase or adoption of a product (37).

Multiple use cases for both patients and staff show potential in forensic psychiatric settings. Wearables can be used for monitoring physiological signals, predicting the risk of aggressive and violent behavior (38–40), and distinguishing between different types of violent behavior (32, 41). Biosensors (integrated in wearables) can provide insight into daily-life (physiological) stress reactivity of patients and staff members in different situations (42), and can increase resilience through biofeedback or just-in-time interventions (43). Wearables might also contribute to the overall physical and mental health of patients and staff (25, 26, 30), particularly in forensic psychiatric patients with mild intellectual disability or borderline intellectual functioning (33). That is, if the device and accompanying app is tailored to their needs. However, there is currently a scarcity of research on the usability and

acceptance of wearables among special needs samples, such as in individuals with mild intellectual disability or borderline intellectual functioning (44).

Since the introduction of the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders) (45), more emphasis is placed on adaptive functioning (instead of intellectual functioning of IQ) when classifying and assessing the severity of intellectual disability. For instance, borderline intellectual functioning is a V-code, which is used if there is a reason for support or if treatment prognosis is affected (46). The main benefit of wearables for these special needs groups is that the information from the devices might be easily adjusted to their specific needs and capabilities. If implemented correctly, the cognitive load on the user is limited. In addition, individuals might benefit from the non-intrusive and passive monitoring combined with targeted interventions to stimulate physical activity, take rest, or push just-in-time notifications to make them aware of deteriorating sleep-wake patterns and increased stress levels, thereby increasing overall interoceptive awareness and self-regulation (22, 47). For staff members who often deal with challenging and aggressive behavior from the patients, wearables might be beneficial to monitor stress levels, recovery during sleep, and overall health, which might serve as indicators of exhaustion and burnout, but also provide targeted interventions to increase resilience (43, 48). The (continuous) use and acceptance of wearables, both in consumer markets and beyond, have fallen short of initial expectations, especially regarding their continuous adoption on the long term (37, 49, 50). Therefore, prospective studies are imperative to study the usability, acceptance and continuous usage, while considering various use cases, devices, and user preferences. In a previous feasibility study (33), we compared several devices and use cases in forensic psychiatry for patients with mild intellectual disability or borderline intellectual functioning, along with their caregivers. We found an association between actually wearing the device and the intention to continuously use the device. Expectations that people have prior to wearing the device played a relatively minor role in their intention to continuously use the device. One strategy to thus increase adoption and explore various use cases is to let people actually test multiple devices. The previous feasibility study included several proprietary and commercially available devices. An important dilemma that emerged is whether it is worth the effort to develop hardware and software applications for these target groups, or that we might use commercially available devices that are readily available and are optimized for technology readiness and usability, but also have several privacy, judiciary, and proprietary caveats. The feasibility study had limitations: participants wore one device each, which prevented a direct comparison between the devices. Additionally, some devices were worn more often due to the nature of the randomization resulting in an unequal number of participants wearing a particular device.

To address the need for prospective studies and overcome the limitations of the earlier feasibility study (33), we conducted a longitudinal randomized crossover study in which four devices with different functionalities were tested on usability, acceptance, and continuous use. Most previous studies only compared these aspects in separate samples where none of the participants have worn all devices (33, 49, 51). Given the potential applications in forensic psychiatry, we included two general purpose fitness trackers, one device providing real-time bio-cuing in daily life and a device that provides the raw data on multiple physiological signals. The latter is used to provide insight into moment-to-moment and day-to-day physiological reactivity to daily life (stressful) events and situations. The main goal of the current study was to evaluate the usability, acceptance, and continuous use among both patients and staff in forensic psychiatry. We hypothesize that there will be differences in usability, acceptance, and continuous use between devices and user groups (staff vs patients).

2 Method

2.1 Participants and setting

This study included participants from four Dutch medium security forensic psychiatric centers (i.e., De Borg) that collaborate and are specialized in the treatment of patients with mild intellectual disability or borderline intellectual functioning. Besides their special needs, the patients also suffer from various mental health problems or mental disorders, such as substance use disorder and personality disorder. Many of them display severe aggressive and violent behavior. A total of 32 participants were included, evenly split between patients and staff members. Separate inclusion criteria were determined for patients and staff members. Patients needed to be admitted to one of the forensic wards, had to provide written informed consent, and had to meet eligibility for inclusion as assessed by the head of the multidisciplinary treatment staff. Staff members had to work within one of the forensic centers and have daily interaction with patients. Specific exclusion criteria were acute psychotic state and/or an objection to participation, as assessed by the primary practitioner.

2.2 Procedure

The research proposal was approved by the ethics (ECSW2020033) and science committee of Radboud University and adhered to the Helsinki Declaration for research involving human participants. The research was conducted between May 2020 and October 2021. The recruitment process consisted of folders and flyers that were distributed by research coordinators within the four healthcare centers. Participants were informed about the research through these flyers and information leaflets. Upon providing informed consent, participants were enrolled in the study. Each participant wore a different wearable for one week, resulting in a total of four weeks across all four wearables (see 2.3). The order of wearing the devices was randomly determined using a research randomizer to eliminate any ordering effects. The wearables were handed out by research coordinators who provided usage instructions. Prior to wearing the device, participants filled out the usability questionnaire to assess the expected usability of the device. Following the one week wearing period, participants filled out questionnaires to assess usability, acceptance, and continuous usage of the devices. In cases where

participants lacked access to a mobile phone or chose not to use it, a research device was provided. If needed, anonymized accounts were created for the participants.

2.3 Devices

Four devices were selected in collaboration with staff members who worked on the wards. These choices were guided by considerations such as perceived ease of use, potential benefits for patients and staff members and the applicability to various use cases in forensic psychiatry, such as bio-cueing, emotion regulation, anger management, providing insights, health tracking, or behavior modification. We selected the Fitbit Charge 3, Garmin Vivosmart 4, Empatica E4 (with a custom made user interface currently in active development), and Ticwatch E3 (with a bio-cueing app that is in active development). We opted for the Empatica E4 and Ticwatch E3 with Sense-IT app due to their capacity to provide raw data and ensure anonymous storage of user information, both crucial factors in healthcare, especially in forensic psychiatry (34). While the custom made prototypes for the Empatica and Ticwatch (42, 52, 53) might have lower ease of use and technology readiness scores (54), their inclusion stemmed from being specifically designed for the mental health context. Furthermore, the current study serves as a reference against general purpose fitness trackers.

2.3.1 Fitbit Charge 3

The Fitbit Charge 3 is a fitness tracker equipped with a built-in heart rate monitor that continuously tracks users' activity levels and heart rate in real-time. Additionally, the accompanying app offers information through composite scores (i.e., digital biomarkers) derived from heart rate and movement sensors, including sleep data. Users receive daily insights on various metrics, including step count, calories burned, stairs climbed and activity metrics. The validity of Fitbit trackers in comparison with golden standard (or criterion) devices varies depending on the physiological signal being tracked and the criterion device used for comparison. For instance, a study showed that the heart rate monitoring of the Fitbit Charge 3 (on photoplethysmography; PPG) in comparison with a criterion chest strap device is relatively poor (see (55)), as indicated by the limits of agreement and poor correlation. Regarding the physical activity measurements, the Fitbit Charge 3 was found to overestimate step count in comparison with a criterion device, though correlations for mean daily step count fall in the moderate to excellent range (56).

2.3.2 Garmin Vivosmart 4

The Garmin Vivosmart 4 is a consumer grade fitness tracker equipped with heart rate monitor (based on PPG) that provides users with various indices of heart rate and accelerometry. In addition, composite scores for sleep, energy expenditure, step count, stairs climbed, and 'stress' are available. A systematic review on Garmin activity trackers showed that step accuracy was considered to be good to excellent. However there is a limited number of studies that have assessed the accuracy of sleep, speed or elevation, and these studies often lack a criterion device such as polysomnography for sleep estimation (57). A study conducted with the Garmin Vivosmart 4 in older adults indicated that the device tended to underestimate the step count at low speeds, but exhibited more accurate readings at higher speeds (58).

2.3.3 Empatica E4 with E4 dashboard

The Empatica E4 (59) is a research grade device that records sensor data in a text file (csv) format. It measures blood volume pulse from which heart rate is derived. The E4 also provides an inter-beat-interval to calculate heart rate variability (HRV). HRV indices with the E4 were extensively studied (60–63) and only validated under resting and (very) low movement conditions. Besides blood volume pulse, electrodermal activity (EDA) is recorded and serves as an index for sympathetic nervous system activation (47, 64). EDA is useful for strong and sustained stressors (63), but the reliability and validity compared to criterion devices is uncertain (61). In addition, skin temperature and accelerometer data are recorded. The current study also aimed to further develop an application called the E4 dashboard (42), designed for clinicians and patients to obtain insights into their daily and momentary physiological reactions to stressors and daily events. The dashboard is a precursor to the version described in (42), and presents physiological graphs similar to the graphs provided by Empatica. Participants can add a calendar to the graphs, visualizing physiological reactions during various activities like therapy, conversations, treatment, work or sports. The dashboard also provides clinicians with information on commonly used parameters (e.g., HR, accelerometry, EDA level) and signal quality.

2.3.4 Ticwatch E3 with Sense-IT app

The Mobvoi Ticwatch E3 is a Smartwatch with PPG sensor to measure heart rate, along with an accelerometer and oximeter. In the current study, the full functionality of the Ticwatch was not utilized, but a custom made bio-cueing app called the Sense-IT (52, 53) was installed on the Ticwatch E3. The purpose of the Sense-IT app is to provide real-time and continuous biofeedback on heart rate changes during real-life situations (14). Additionally, it aims to enhance interoceptive and emotional awareness throughout daily activities. It is important to note that usability and validity studies are still in progress with the Sense-IT as it is currently under active development. However, a recent study that used the Sense-IT app indicated that patients and caregivers had a positive attitude towards the application (14), and usability for the Sense-it app ranged from approximately 63 to 76 (65–67). One study indicated that the usability was OK for patients, but staff members generally perceived it as poor (33).

2.4 Questionnaires

In collaboration with several staff members who regularly worked with patients, we created revised acceptance and continuous use questionnaires for patients. This adjustment was

made to accommodate patients with mild intellectual disability or borderline intellectual functioning who often struggle with word comprehension. To simplify the questionnaire for patients we modified some questions. For example, the question "I would like to use this product frequently" was adjusted to "… more often". In addition, one question of the System Usability Scale (SUS) was rephrased from a positive statement in the original version to a negatively worded question in the adjusted version (question 3) as this was considered easier for patients to understand.

2.4.1 System Usability Scale

The System Usability Scale (SUS) is commonly used to rapidly assess the subjective usability of a product, including various types of technology such as fitness trackers, digital health applications, or medical devices (36, 68). Usability is the degree to which a product is fit or able to be used. Administration of the questionnaire is considered fast and easy for a plethora of users (69, 70). The SUS has high reliability (α=.85) and a meta-analysis (68) showed that usability is a quality feature depending on the ease of use of the applications (and the accompanying technology). Usability scores for physical activity apps were relatively high in the meta-analysis (68), which is of particular interest as this type of application is also used in the current study. The SUS has clear standards and benchmarks (49, 68) in which a total score of ~77 (SD = 15.12) was found across all tested digital health applications in the included studies. The 10-item SUS is scored on a five-point Likert scale ranging from strongly agree (5) to strongly disagree (1). Even numbered questions (items 2, 4, 6, 8, and 10) are negatively worded, while the uneven numbered questions (1, 3, 5, 7, and 9) are positively worded. Missing values are replaced with a 3, following recommendations (71). For positively worded questions, the score minus 1 is calculated, while for negatively worded questions, the score is subtracted from 5. All items are then summed and multiplied by 2.5, effectively yielding a maximum total SUS score of 100. As for the interpretation, a significant body of research is available (36, 69, 72, 73) indicating an average SUS score of 68 being the average among a considerable number of usability scores. People will typically recommend a system that reaches a SUS-score of 82 (73). An adjective scale developed by Bangor et al. (2009) indicated that a mean SUS-score above 35.7 is categorized as having poor usability, above 50.9 is considered to be OK, while a score above 71.4 indicates good usability. Mean scores above 85.5 are considered to be excellent usability scores. We compared the scores in the current study with these benchmarks.

2.4.2 Technology Acceptance Model

The Technology Acceptance Model (TAM) – questionnaire is an often used questionnaire to assess the acceptance of new technology (37). The TAM is based upon the theory of reasoned action (33, 74, 75). For the current study, we administered a more recent development of the TAM that is specifically tailored to smartwatches (37). This TAM version distinguishes between 10 determinants of acceptance summarized in subscales. Two determinants of acceptance are central in the model (75, 76): perceived usefulness (PU) and perceived ease of use (PEOU). These two determinants are central to users' intention for future technology use and are influenced by the 8 other determinants: mobility, perceptions of and attitudes toward technology, affective quality, subcultural appeal, relative advantage, availability, intention to use, and cost. The TAM has scoring options ranging from strongly agree to strongly disagree on a 7-point Likert scale. The TAM subscales have reliabilities above .70 and the questionnaire consists of 36 questions. For staff members, the full scale was used, whereas a short version with simplified wording was devised for patients. The full questionnaire was perceived to be burdensome to some of them. For each construct, we selected one or two questions of each determinant in cooperation with a team of staff members who frequently interact with these patients. The short TAM questionnaire had a reliability above .80 (Cronbach alpha) (33).

2.4.3 Extended Expectation Confirmation Model

The Extended Expectation Confirmation Model (EECM) is a questionnaire specifically designed to assess the intention for continuous use of smart-wearables (77). The EECM is rooted in the expectation-confirmation theory, which seeks to elucidate user satisfaction in the context of extended use following the purchase or adoption of a product. This model is based on the beliefs that users have with regard to the products' performance and the (dis) confirmation of these beliefs and expectations (77, 78). The EECM consists of 32 questions that can be scored from strongly agree to strongly disagree using a 7-point Likert scale. The EECM subscales have reliabilities above .70. The 10 subscales of the EECM encompass continuous use, hedonic motivation, battery-life concern, self-socio motivation, perceived privacy, perceived comfort, perceived usefulness, perceived accuracy with functional limitations, satisfaction, confirmation and continuous use. In line with the TAM, we developed a shortened version of the EECM for the patients in the current study. The short EECM has a reliability above .80 (Cronbach alpha) (33). One of the EECM questions was deemed quite difficult for patients to understand due to negation and complex phrasing leading us to rephrase the question into an affirmative one: "I think that the information provided by the product is correct".

2.5 Power analysis

We conducted a power analysis prior to the study, considering an estimated effect size of ~.35 derived from a previous study (33). This previous study investigated the difference between devices for both patients and staff members. To address the research question for a repeated measures-analysis of variance (RM-ANOVA; groups, a power of 95%, and a conservative .25 correlation between measurements, a sample of n=28 was needed for the current study. In order to account for drop out, 32 participants were included.

2.6 Statistical analysis

First, we calculated descriptive statistics separately for staff members and patients. Subsequently, SUS scores were computed for each device, as well for both groups of patients and staff members. To determine whether there was a difference in usability between devices, a repeated measures ANOVA was conducted with total SUS score as the dependent variable and the type of device as the within-subject factor. Additional models for between-group differences (staff and patients) and interactions were also tested. The TAM and ECCM questionnaires varied in length for patients and staff members, and were therefore analyzed separately using descriptive statistics.

3 Results

3.1 Sample description

We included 16 patients and 16 staff members who wore all four of the devices and assessed the system usability, technology acceptance and continuous use intention. The age in years of the participants ranged from 18 to 53 (see Table 1 for an overview of descriptives). Due to technical reasons, the answers of one staff member were not properly saved digitally. To address this, we assigned a value of 3 for the SUS and 4 for both the TAM and the EECM for this participant, in line with recommendations (71).

TABLE 1 Descriptive statistics of sample.

Participants	Patients (n=16), n(%)	Staff (n=16), n(%)
Education		
Primary	10(62%)	–
Secondary	6(38%)	5(31%)
Higher	–	11(69%)
Gender		
Male	13(81%)	8(50%)
Female	3(19%)	8(50%)
Age		
Mean	31.2	33.7
SD	11.5	9.82

3.2 Usability (SUS)

After wearing the device, the descriptives from the SUS scores (Table 2) indicate that only for staff members, the Fitbit and Garmin devices received a "good" usability rating (36, 72, 73). Regarding patients, the Empatica device with E4 dashboard application was assessed to have poor usability, while the Ticwatch device with Sense-IT application received an "OK" rating. The SUS scores for the Fitbit and Garmin devices increased from the pre-test (expected usability) to the post-test (experienced usability), while the opposite trend was observed for the Empatica and Ticwatch.

To determine differences in system usability between devices for patients and staff, we conducted an RM-ANOVA with the total SUS score as the dependent variable and the type of device as the within subject factor. One outlier was detected for the Fitbit, however, this was not an extreme case. The Shapiro-Wilk test indicated that the data were distributed normally ($p>.05$). However, sphericity was found to be violated based on Mauchley's test, $\chi2(5) = 14.7$, $p = .012$. Therefore, we used Huynh-Feldt correction to interpret the results, which returned a significant result, indicating a difference in system usability between devices, $F(2.631, 81.563) = 18.689$, $p < .001$, partial $\eta^2 = .38$. *Post-hoc* analyses with Bonferroni corrections revealed that the Empatica and Ticwatch devices resulted in significantly lower system usability scores in comparison with the Fitbit and Garmin devices. Conversely, there was no significant difference between the Fitbit and Garmin devices on the one hand and Empatica and the Ticwatch on the other.

We also checked (2-way-mixed-anova) differences between staff members and patients on their system usability scores. Unfortunately, the assumptions required for these tests were violated and no non-parametric alternatives were available (79). We were thus unable to formally test the differences between staff and patients. Additional comparisons were non-applicable for age

TABLE 2 Descriptive statistics of SUS scores.

participant	Product_Start	n	Start	sd	End	sdEnd	minStart	maxStart	minEnd	maxEnd
patient	empatica	16	50.47	15.09	46.09	20.82	30.00	80.00	15.00	75.00
patient	fitbit	16	68.28	14.82	71.25	19.30	42.50	97.50	40.00	100.00
patient	garmin	16	64.69	12.11	70.78	18.77	47.50	87.50	37.50	97.50
patient	ticwatch	16	60.00	13.04	52.66	22.70	40.00	90.00	15.00	87.50
staff	empatica	16	57.66	11.53	55.31	18.93	37.50	77.50	25.00	82.50
staff	fitbit	16	76.25	13.66	76.72	9.52	50.00	100.00	50.00	87.50
staff	garmin	16	64.69	18.77	72.97	13.11	30.00	92.50	47.50	87.50
staff	ticwatch	16	66.09	13.07	56.41	23.13	50.00	92.50	7.50	92.50

10.3389/fpsyt.2024.1330993

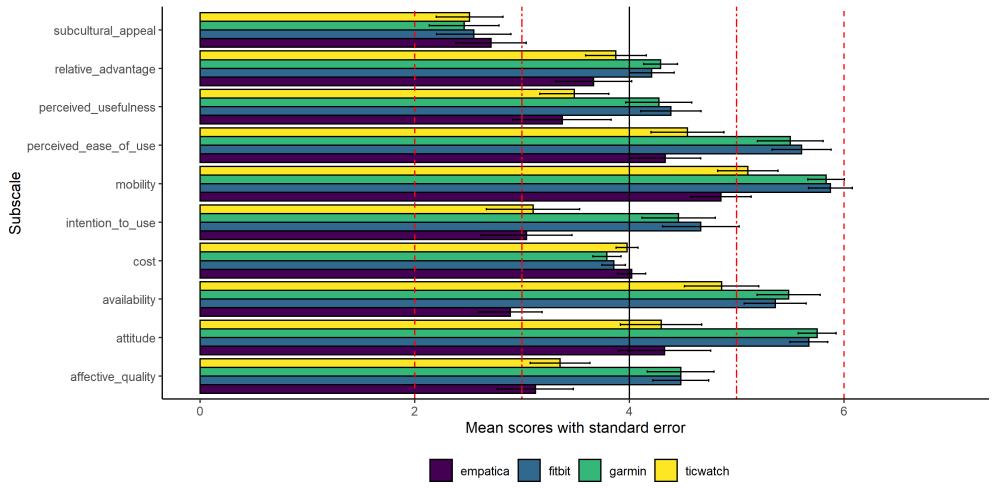


FIGURE 1
Staff members' mean scores and standard errors on the subscales of the TAM.

(as it was centered around +/- 30 years), education (we found different distributions for patients and staff members), and gender (there were not many females in the patient group).

3.3 Staff acceptance and continuous use

The TAM questionnaire is typically presented in mean or median values on the individual subscales, which differs somewhat from the approach used with the SUS. We have calculated the mean of the subscales for staff members (Figure 1), for which perceived ease of use and perceived usefulness are the central subscales. For patients we used a different approach. Since the staff members completed the full version of the TAM, and the patients a shorter version, we cannot directly compare the outcome head to head. Therefore, patients' results are described in 3.4.

As can be seen in Figure 1 for the staff members, the subcultural appeal of all devices falls considerably below the median score of the questionnaire. A mean score <3 is indicative of a score below the median, and a mean score >5 is considered well above the median. For the Fitbit and Garmin devices, the perceived ease of use, mobility, availability, and attitude all received ratings are well above the median. However, for the Ticwatch, only the mobility subscale scores are well above the median. The availability of information for the Empatica is considered well below the median, which is probably due to the fact that the wristband has no interface and was not used to provide real-time information during the current study. Information was only available after the participant had worn the device.

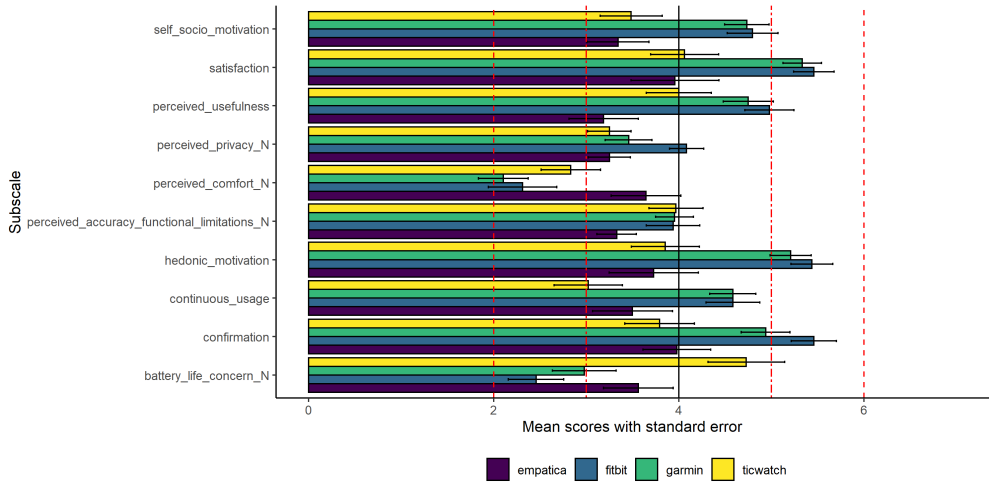


FIGURE 2
Staff members' mean scores and standard errors on the subscales of the EECM. *Note that the questions from the subscales denoted with _N were negatively phrased, while the other subscale questions were positively phrased. Similar to the TAM, the EECM also has 10 subscales.

The results for staff members on the EECM are shown in Figure 2. For the Fitbit, the mean scores are well above the median for satisfaction, hedonic motivation, and confirmation. However, scores are well below the median for perceived comfort and battery life concern (these two subscales need to be interpreted positively as the questions were phrased negatively). The Garmin has scores well above the median for satisfaction and hedonic motivation, but well below the median for perceived comfort and battery life concern. The Ticwatch has a mean score well below the median for perceived comfort. Lastly, the Empatica did not particularly stand out in terms of continuous use intention.

3.4 Patients' acceptance and continuous use

For patients, we conducted an item-level analysis of both the TAM and EECM (see Supplementary Materials). This approach was chosen because it is hard to argue that one or two questions can adequately represent an entire subscale that consists of many items. For both Fitbit and Garmin, over 75% of patients responded positively to several questions: "they liked the idea of using the Garmin and Fitbit", "found it easy to use", "thought it was attractive and pleasing", "thought it was useful for their job", "felt they could use it anywhere", "provided them with the desired information and service", "provided them with a pleasant experience, and it was better than expected", "found it to be entertaining", and "thought that the information from the product was correct". While these percentages were considerably lower for Empatica and Ticwatch, some participants did indicate positive aspects of these devices as well. Over 75% of patients believed that the information from the Empatica was correct, which may be attributed to its purposeful design as a research device, to measure physiological signals as reliably and validly as possible.

4 Discussion

4.1 Main findings

In this study, a randomized crossover design was employed to assess the usability, acceptance, and continuous use intention of four different wearable devices among both patients and staff members in forensic psychiatric settings. The findings revealed a statistically significant difference in usability between Fitbit and Garmin fitness trackers and two devices that use custom made applications (targeted at gaining insight into physiological reactivity and providing bio-cueing in daily life). Further developments and usability studies are needed to provide users with a similar usability experience as the Garmin and Fitbit fitness trackers. The E4 dashboard and Sense-IT applications were designed to address several challenges in forensic psychiatry, such as emotional self-regulation, (mental) health tracking, behavior modification, providing insight into physiological reactivity, and interoceptive awareness. Achieving comparable levels of usability, acceptance and continuous use intention as commercially available sensors is essential to improve adoption, and research has suggested that gamification and motivation boosting strategies may help to improve uptake and usability (68). In the current study, we aimed to compare multiple devices over extended periods, as earlier research indicated that hands-on experience with wearables was associated with continuous use. Consequently, we limited the scope of the study and did not investigate whether participants appreciated the tailored aspects of the custom made applications for the specific use cases in qualitative research. Rather, we used standardized questionnaires on (continuous) use and acceptance. These aspects are typically well taken care of in commercially oriented wearable technologies.

It is notable that the commercially developed Fitbit and Garmin devices were not above the usability scores that would typically lead people to recommend that technology. Previous research has shown that usability scores should be above ~82 points on the SUS for people to endorse system technology (73). The low scores in usability could seriously hinder the adoption of wearable technology in forensic psychiatry for both staff members and patients, particularly for those patients with mild intellectual disability or borderline intellectual functioning who require user friendly technology. To ensure that participants can derive benefit from the technology (e.g., to gain insight into their physiological reactivity, improve self-regulation or track elements of physical and mental health), it is imperative to develop devices tailored to the unique and personal needs of staff members and patients (12, 13).

Although no formal statistical comparison between staff members and patients was conducted due to violations of several assumptions, SUS-scores indicate a similar trend. Both patients and staff members gave higher scores for Fitbit and Garmin, but lower scores for Empatica and Ticwatch. The difference in SUS scores may be partially explained by the difference in pre-test scores as these were already lower for Empatica and Ticwatch. Although we could not directly compare acceptance and continuous use between patients and staff members due to the use of shortened versions of the questionnaires for patients, the results indicated similar trends for patients as for staff members regarding acceptance and continuous use for all devices.

4.2 Strengths and limitations

A particular strength of our current study was the use of a randomized and counterbalanced design that enabled direct comparisons on the usability of the devices, which was a limitation in our previous study (33). Additionally, the shortening of questionnaires for patients with mild intellectual disability or borderline intellectual functioning allowed for easy and time efficient assessment. Staff members found that the administration of these shortened questionnaires were more feasible and comprehensible for patients, thereby reducing the burden of participation and the time needed for questionnaire administration. However, the disadvantage of using these shorter questionnaires for patients was the inability to make direct comparisons for acceptance and continuous use intention between staff members and patients. This limitation stemmed

from the fact that patients only received a subset of the questions from the TAM and EECM. Future studies could consider employing the simplified questions for staff members as well to investigate potential differences in acceptance and continuous use intention between the two groups. Although the technology is thought to benefit both staff members and patients, it is important to recognize that the applications and use cases may differ. One aspect for which wearable technology might be useful could be the detection of stress-related problems or sleep problems among staff members. During their admission, patients may display aggression towards staff, leading to potential negative consequences such as symptoms of post-traumatic stress (48, 80, 81). Moreover, recent studies have indicated that aggressive behaviors, including threats, physical aggression, and unwanted sexual approaches, significantly contribute to absenteeism and staff turnover. Additionally, there is substantial evidence linking violence to an increased risk of anxiety, sleep problems, burnout, and depression in staff (80, 82, 83). Wearables could prove especially useful in enhancing the physical and mental health and resilience of staff members.

The four selected devices have different use cases and purposes, which is important to keep in mind when comparing the devices, especially those for which we used (first versions) of custom made software. After consulting with several clinicians and staff members, we believe that these different use cases and devices have potential value for patients and staff members in forensic psychiatric settings. Our current study provides first reference scores on usability, acceptance and continuous use intention. We expect that the potential applications of wearable technology will substantially increase for personalized and tailored use in the coming years (12, 13). The questionnaires on usability, acceptance and continuous use intention provide valuable information on the current state of the wearable technology. The questionnaires clearly indicate what additional work needs to be done before the devices can be implemented on a larger scale.

In a general sense, and as demonstrated by years of research with the SUS, we should aim for technology with usability scores above ~82. Physical activity apps reached a mean usability score of ~83 in a recent meta-analysis (68), and the authors indicated that the popularity of physical activity apps might be due to the gamified nature and motivational features built in these apps. These aspects might also be integrated into the devices and applications used in our study.

A specific limitation of the TAM was enclosed in the Cost subscale, where one question was positively phrased as "CT3: I was able to easily afford this smart watch.", while the other two were negatively phrased (e.g., CT1: this smart watch was expensive). Calculating a mean score on this subscale effectively influences the interpretation of the question. Given that other questions in the TAM are phrased positively, we would recommend to rephrase all questions in a positive manner, or adopt a similar strategy to the SUS where half of the questions are positively phrased, and half are negatively phrased (37, 74). A comparable scoring system with a maximum score of 100 could also increase comparability on usability, acceptance (and continuous use).

We did not include qualitative questions in the current study as we did in our previous study (33). This decision was made to alleviate the burden on participants who were already asked to wear four devices for of four weeks. Nevertheless, a qualitative evaluation could have provided additional insights into the specific use cases and strengths of the devices that were used. Moreover, we did not include qualitative questions regarding users' attitudes that are also fundamental aspects related to the adoption of wearable devices (84, 85).

Another limitation is the use of a liberal convenience sampling strategy. We did not include a sample with balanced age, gender, and seniority restrictions, which may have led to selection bias in the current sample. In addition, we did not select specific patient samples, so it is possible that some patients may have been more willing than others to wear the devices. A final limitation is that participants only wore the devices for one week without extensive technology use guidance. For instance, the Sense-IT app was tested in another study where participants were asked to wear the device for longer periods while receiving extensive training on the use of the devices. In those studies, the SUS scores for the Sense-IT app ranged from ~63 to ~76 (65–67). This implies that custom made applications do not readily compare with off the shelve technology (such as Fitbit and Garmin), and needs additional effort for implementation. Extensive guidance and experiential learning might be relevant for special needs populations, such as people with mild intellectual disabilities and borderline intellectual functioning.

4.3 Future research

Future research should focus on the validation of the shortened versions of the TAM and EECM in a sample of people with mild intellectual disabilities or borderline intellectual functioning. It is crucial to ensure that the questions are adapted to the needs of those who may have difficulties understanding the questions. Society is becoming increasingly complex, and some people, especially patients with mild intellectual disabilities or borderline intellectual functioning, find it difficult to keep up. We should develop technology that is easy to use, useful, valuable and which has a low cognitive load.

Several researchers have argued that the devices are to be used with caution as the validity and reliability of the algorithms are still questionable (55–57). Also, not all algorithms for artefact detection, stress detection or sleep classification can be validated due to a lack of raw data or proprietary algorithms (42). Open source algorithms and devices such as the custom made applications used in the current study might provide users with information that can be validated. Although the four devices in the current study were carefully selected for use within clinical practice of forensic psychiatric settings, there may still be a degree of subjectivity and selection bias. Future research should prioritize the comparison of additional and multiple devices.

The E4 dashboard and Sense-IT applications were used with custom made software that provides bio cueing (Ticwatch) or combines physiological reactivity information with daily-life

situations, events and circumstances (Empatica). The usability scores clearly illustrate that these custom made applications need further improvement. We did not expect that similar usability scores would be obtained as for commercial devices, but future research could explore additional tools to assess and improve usability, perceived ease of use and perceived usefulness, as these constructs are often validated, and indicative of product quality (86). Recent research has suggested that devices preferably have a clear purpose to potentially increase long-term use and user loyalty (87). Both the Empatica E4 dashboard and the Sense-IT app were designed with a very clear purpose in mind, aligning well with earlier recommendations (87).

In the current study, we only used self-report questionnaires and did not consider actual use of the devices. Future studies should consider collecting actual user data, as it might provide additional information that can better align with user needs and preferences (88).

4.4 Clinical implications

Wearables provide us with opportunities to understand how patients respond to various (stressful) daily life experiences and (treatment) situations that influence the bodily reactions, physiology and emotional well-being. Together with their therapist, patients can explore whether these technologies provide new insights related to the specific problem the patient is working on. One important implication is the (longer term) usability tailored to each patient's needs. Some use cases (e.g., tracking specific sleep-stress interactions, evaluating longer treatment effects on self-regulation) might require a different device and measurement duration than others (e.g., physiological reactivity to specific treatment situations, heart rate variability biofeedback). It is important to evaluate device usability beforehand and continuously assess usability to ensure that they remain motivated to work on a specific problem or goal.

The use of wearable technology (e.g., chest straps, ear lobe sensors, wristbands, patches) provides exiting opportunities for delivering continuous (as opposed to episodic) feedback and for interventions on outcomes relevant to the individual. Several systematic reviews showed that wearable technology can have a positive impact on sleep, stress, physical activity, depression, emotional regulation, and cardiovascular and metabolic functioning (14, 20, 24–30). Sleep, stress and physical activity are considered transdiagnostic markers of psychiatric problems (18–20), and can be monitored relatively easy with wearable technology, assuming that the algorithms are accurate and robust (35).

Besides providing information on transdiagnostic markers, wearable technology can offer insights into individual progress or decline (especially throughout treatment, and over treatment sessions), provide just-in-time interventions, or provide insight in daily person-environment interactions and their effects on physiological stress reactivity, cognition and emotion. The challenge is to integrate the wearable technology meaningfully and usefully into clinical practice. Wearable technology has potential for psychological functioning by improving insights, self-awareness, health management, and motivation, but might also have a negative impact on psychological functioning by increasing anxiety, dependency, and worrying (21–23, 42, 66, 89).

Personalized and continuous feedback opens novel opportunities to increase the efficacy of existing treatments and individual functioning (5, 12, 13). However, the current study shows that integration and implementation of this technology is not seamless, and usability, acceptance and continuous use needs improvement. Wearables might provide us with novel opportunities to develop assistive technologies that can be useful for continuous support and just-in-time warnings that can also assist clinicians in providing efficient treatment, reduce physician time, and possibly reducing the cost of healthcare (49, 90).

From a clinical healthcare perspective, Fitbit and Garmin devices are general purpose fitness trackers offering users a range of functions. People can track heart rate (variability), accelerometry, track training progress, recovery, training load, provide reminders, share data with relevant others, and provide composite scores based on physiology that estimate stress, sleep, physical activity, or energy expenditure (91). The uptake of these devices is growing, but there are concerns about their validity, reliability and precision (15, 42, 91). In contrast, the Ticwatch with Sense-IT app (92) has a single purpose function providing real-time bio-cues in the moment that provide insight into changes in heartrate during daily activities. The meaning of the information is not labelled by the device and users can adjust and personalize these settings The Sense-IT provides no interpretation of the changes in the physiology, but lets the user interpret the information. Users can adjust their thresholds for bio-cue information to their liking and add notes regarding their activities. The Sense-IT makes users aware of the bodily changes that occur under different circumstances and under different stressors and events.

On the other hand, E4 dashboard was developed to provide deeper insight into patients' physiological reactivity throughout the day and synchronizes the information with daily life stressors or events that cause increased arousal. It can be used as a talking board between clinicians and their patients giving them a better understanding of what may cause a bodily reaction of a patient. It may also provide information on under arousal or over arousal of the patient. Further developments of the E4 dashboard include adding open source algorithms for sleep, stress, and physical activity combined with evidence-based information on possible interventions. The current study aimed to establish a reference for different wearable devices and explore applications that might prove useful in forensic psychiatry and other healthcare settings, especially special need samples such as people with mild intellectual disability or borderline intellectual functioning. Different forms and types of wearable technology may prove useful in forensic psychiatry. The wearables used in our study were designed with different goals, applications, and use cases in mind. During treatment, patients have to work on several problems, ranging from trauma to violent behavior, lifestyle coaching, physical and mental health and reintegration. It is unlikely that a single wearable device or application will be useful and valuable for each use case and application. Thus, it is vital to explore different types of wearable technologies to cater the unique needs of patients in forensic psychiatry and other healthcare domains.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics (ECSW2020033) and science committee of Radboud University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

PL: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. MN: Writing – review & editing, Supervision, Methodology, Conceptualization. HN: Writing – review & editing, Resources, Supervision, Methodology, Funding acquisition, Conceptualization. LG: Writing – review & editing, Supervision, Resources, Investigation. SB: Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. RD: Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Financial support for implementation and future development of E4 dashboard was available from an implementation and innovation fund from health insurers Netherlands, and ZonMw (grant number 06360322210023). The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit for publication.

Acknowledgments

We thank Rein van den Bos, Walter Rosen, Dave Gerritsen, Jesse Neervens, Simone Vreeke, Chico Taguba, Maria de Jong, and Willem Dijs for their help in data collection. We would like to thank the project partners of the Sense-IT project for making the Sense-IT available for use in this study: University of Twente, Scelta/GGNet, VUmc, Arkin, and Pluryn.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2024.1330993/full#supplementary-material

References

1. Papalia N, Spivak B, Daffern M, Ogloff JRP. A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clin Psychol Sci Pract*. (2019) 26:e12282. doi: 10.1111/cpsp.12282

2. Gibbon S, Khalifa NR, Cheung NH-Y, Völlm BA, McCarthy L. Psychological interventions for antisocial personality disorder. *Cochrane Database Syst Rev*. (2020) 9: CD007668. doi: 10.1002/14651858

3. Iruthayarajah J, Alibrahim F, Mehta S, Janzen S, McIntyre A, Teasell R. Cognitive behavioural therapy for aggression among individuals with moderate to severe acquired brain injury: a systematic review and meta-analysis. *Brain Inj*. (2018) 32:1443–9. doi: 10.1080/02699052.2018.1496481

4. Khalifa NR, Gibbon S, Völlm BA, Cheung NH-Y, McCarthy L. Pharmacological interventions for antisocial personality disorder. *Cochrane Database Syst Rev*. (2020) 9: CD007667. doi: 10.1002/14651858.CD007667.pub3

5. Kip H, Bouman YHA, Kelders SM, van Gemert-Pijnen LJEWC. eHealth in treatment of offenders in forensic mental health: A review of the current state. Front psychiatry. *Geraadpleegd van*. (2018) 9:42. doi: 10.3389/fpsyt.2018.00042

6. Verschueren S, Jeandarme I, van den Ameele R, Buysschaert B, Bogaerts S. Patient profiles in high-security forensic psychiatry in Flanders. *Psychiatry Psychol Law*. (2023) 30:812–35. doi: 10.1080/13218719.2022.2124549

7. Kozhuharova P, Dickson H, Tully J, Blackwood N. Impaired processing of threat in psychopathy: A systematic review and meta-analysis of factorial data in male offender populations. *PloS One*. (2019) 14:e0224455. doi: 10.1371/journal.pone.0224455

8. Wong TY, Sid A, Wensing T, Eickhoff SB, Habel U, Gur RC. Neural networks of aggression: ALE meta-analyses on trait and elicited aggression. *Brain Struct Funct*. (2019) 224:133–48. doi: 10.1007/s00429-018-1765-3

9. Wong TY, Radua J, Pomarol-Clotet E, Salvador R, Albajes-Eizagirre A, Solanes A. An overlapping pattern of cerebral cortical thinning is associated with both positive symptoms and aggression in schizophrenia via the ENIGMA consortium. *Psychol Med*. (2020) 50:2034–45. doi: 10.1017/S0033291719002149

10. Didden R, Nijman H, Delforterie M, Keulen-De Vos M. *Treatment of anger and violence in individuals with intellectual disability. wiley handb what works offenders intellect dev disabil*. London: John Wiley & Sons, Ltd (2019) p. 297–309. doi: 10.1002/9781119316268.ch15

11. Edberg H, Chen Q, Andiné P, Larsson H, Hirvikoski T. Crimes and sentences in individuals with intellectual disability in a forensic psychiatric context: a register-based study. *Epidemiol Psychiatr Sci*. (2022) 31:e2. doi: 10.1017/S2045796021000718

12. Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. *Cell*. (2021) 184:1415–9. doi: 10.1016/j.cell.2021.01.015

13. Johnson KT, Picard RW. Advancing neuroscience through wearable devices. *Neuron*. (2020) 108:8–12. doi: 10.1016/j.neuron.2020.09.030

14. ter Harmsel JF, Noordzij ML, Goudriaan AE, Dekker JJM, Swinkels LTA, van der Pol TM. Biocueing and ambulatory biofeedback to enhance emotion regulation: A review of studies investigating non-psychiatric and psychiatric populations. *Int J Psychophysiol*. (2021) 159:94–106. doi: 10.1016/j.ijpsycho.2020.11.009

15. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Front Physiol*. (2018) 9. doi: 10.3389/fphys.2018.00743

16. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. *Lancet Digit Health*. (2022) 4:e615–26. doi: 10.1016/S2589-7500(22)00111-X

17. Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M. Review on psychological stress detection using biosignals. *IEEE Trans Affect Comput*. (2022) 13:440–60. doi: 10.1109/TAFFC.2019.2927337

18. Hussain I, Hossain MA, Jany R, Bari MA, Uddin M, Kamal ARM. Quantitative evaluation of EEG-biomarkers for prediction of sleep stages. *Sensors*. (2022) 22:3079. doi: 10.3390/s22083079

19. Riemann D, Spiegelhalder K, Nissen C, Hirscher V, Baglioni C, Feige B. REM sleep instability–a new pathway for insomnia? *Pharmacopsychiatry*. (2012) 45:167–76. doi: 10.1055/s-0031-1299721

20. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. *J Med Internet Res*. (2018) 20:e210. doi: 10.2196/jmir.9410

21. Triberti S, Durosini I, La Torre D, Sebri V, Savioni L, Pravettoni G. *Artificial Intelligence in Healthcare Practice: How to Tackle the "Human" Challenge*, Springer International Publishing (2022) Vol. 212. pp. 43–60.

22. Fedor S, Lewis R, Pedrelli P, Mischoulon D, Curtiss J, Picard RW. Wearable technology in clinical practice for depressive disorder. *N Engl J Med*. (2023) 389:2457–66. doi: 10.1056/NEJMra2215898

23. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. *N Engl J Med*. (2024) 390:1118–27. doi: 10.1056/NEJMra2307160

24. Clay I, Cormack F, Fedor S, Foschini L, Gentile G, van Hoof C. Measuring health-related quality of life with multimodal data: viewpoint. *J Med Internet Res*. (2022) 24:e35951. doi: 10.2196/35951

25. Daryabeygi-Khotbehsara R, Shariful Islam SM, Dunstan D, McVicar J, Abdelrazek M, Maddison R. Smartphone-based interventions to reduce sedentary behavior and promote physical activity using integrated dynamic models: systematic review. *J Med Internet Res*. (2021) 23:e26315. doi: 10.2196/26315

26. Franssen WMA, Franssen GHLM, Spaas J, Solmi F, Eijnde BO. Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials. *Int J Behav Nutr Phys Act*. (2020) 17:57. doi: 10.1186/s12966-020-00955-2

27. Hodkinson A, Kontopantelis E, Adeniji C, van Marwijk H, McMillian B, Bower P. Interventions using wearable physical activity trackers among adults with cardiometabolic conditions: A systematic review and meta-analysis. *JAMA Netw Open*. (2021) 4:e2116382. doi: 10.1001/jamanetworkopen.2021.16382

28. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. *Br J Sports Med*. (2021) 55:422–32. doi: 10.1136/bjsports-2020-102892

29. Min J, Rouanet J, Martini AC, Nashiro K, Yoo HJ, Porat S. Modulating heart rate oscillation affects plasma amyloid beta and tau levels in younger and older adults. *Sci Rep*. (2023) 13:3967. doi: 10.1038/s41598-023-30167-0

30. Pizzoli SFM, Marzorati C, Gatti D, Monzani D, Mazzocco K, Pravettoni G. A meta-analysis on heart rate variability biofeedback and depressive symptoms. *Sci Rep*. (2021) 11:6650. doi: 10.1038/s41598-021-86149-7

31. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. *Psychol Med*. (2017) 47:2578–86. doi: 10.1017/S0033291717001003

32. de Looff PC, Cornet LJM, de Kogel CH, Fernández-Castilla B, Embregts PJCM, Didden R. Heart rate and skin conductance associations with physical aggression, psychopathy, antisocial personality disorder and conduct disorder: An updated meta-analysis. *Neurosci Biobehav Rev*. (2022) 132:553–82. doi: 10.1016/j.neubiorev.2021.11.003

33. de Looff PC, Nijman HLI, Didden HCM, Noordzij ML. Usability and acceptance of wearable biosensors in forensic psychiatry: cross-sectional questionnaire study. *JMIR Form Res*. (2021) 5 e18096. doi: 10.2196/18096

34. van den Braak SW, Platje E, de Kogel CH. *Data safety and privacy for using physiological wearables in a judicial context [Dataveiligheid en privacy bij het gebruik van fysiologische wearables in de justitiële context]*. The Hague: 1ste dr. Den Haag, The Netherlands: Ministry of Justice, Scientific Research and Documentation Centre [WODC] (2021). Available at: https://repository.wodc.nl/handle/20.500.12832/3036.

35. Henry J, Lloyd H, Turner M, Kendrick C. On the robustness of machine learning models for stress and anxiety recognition from heart activity signals. *IEEE Sens J*. (2023) 23(13):1–1. doi: 10.36227/techrxiv.21688352.v2

36. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Human–Computer Interact*. (2008) 24:574–94. doi: 10.1080/10447310802205776

37. Kim KJ, Shin D-H. An acceptance model for smart watches: Implications for the adoption of future wearable technology. *Internet Res*. (2015) 25:527–41. doi: 10.1108/IntR-05-2014-0126

38. de Looff P, Noordzij ML, Moerbeek M, Nijman H, Didden R, Embregts P. Changes in heart rate and skin conductance in the 30 min preceding aggressive behavior. *Psychophysiology*. (2019) 56:UNSP e13420. doi: 10.1111/psyp.13420

39. Goodwin MS, Mazefsky CA, Ioannidis S, Erdogmus D, Siegel M. Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Res*. (2019) 12:1286–96. doi: 10.1002/aur.2151

40. Park C, Rouzi MD, Atique MMU, Finco MG, Mishra RK, Barba-Villalobos G. Machine learning-based aggression detection in children with ADHD using sensor-based physical activity monitoring. *Sensors*. (2023) 23:4949. doi: 10.3390/s23104949

41. Fanti KA, Kyranides MN, Georgiou G, Petridou M, Colins OF, Tuvblad C. Callous-unemotional, impulsive-irresponsible, and grandiose-manipulative traits: Distinct associations with heart rate, skin conductance, and startle responses to violent and erotic scenes. *Psychophysiology*. (2017) 54:663–72. doi: 10.1111/psyp.12837

42. de Looff P, Duursma R, Noordzij M, Taylor S, Jaques N, Scheepers F. Wearables: an R package with accompanying shiny application for signal analysis of a wearable device targeted at clinicians and researchers. *Front Behav Neurosci*. (2022) 16:856544. doi: 10.3389/fnbeh.2022.856544

43. Kalisch R, Baker DG, Basten U, Boks MP, Bonanno GA, Brummelman E. The resilience framework as a strategy to combat stress-related disorders. *Nat Hum Behav*. (2017) 1:784–90. doi: 10.1038/s41562-017-0200-8

44. Steinert A, Haesner M, Steinhagen-Thiessen E. Activity-tracking devices for older adults: comparison and preferences. *Univers Access Inf Soc*. (2018) 17:411–9. doi: 10.1007/s10209-017-0539-7

45. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition*. Arlington, VA: American Psychiatric Association (2013). doi: 10.1176/appi.books.9780890425596

46. Jonker F, de Looff P, van Erp S, Nijman H, Didden R. The adaptive ability performance test (ADAPT): A factor analytic study in clients with intellectual disabilities. *J Appl Res Intellect Disabil*. (2022) 36(1):3–12. doi: 10.1111/jar.13044

47. Everly GS Jr., Lating JM. *A Clinical Guide to the Treatment of the Human Stress Response*. New York, NY: Springer (2019). doi: 10.1007/978-1-4939-9098-6

48. de Looff PC, Didden HCM, Embregts PJCM, Nijman HLI. Burnout symptoms in forensic mental health nurses: Results from a longitudinal study. *Int J Ment Health Nurs*. (2019) 28:306–17. doi: 10.1111/inm.12536

49. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. *JMIR MHealth UHealth*. (2018) 6:e11066. doi: 10.2196/11066

50. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: A review of the evidence for increasing physical activity. *Front Public Health*. (2016) 4:289. doi: 10.3389/fpubh.2016.00289

51. Hofbauer LM, Rodriguez FS. How is the usability of commercial activity monitors perceived by older adults and by researchers? A cross-sectional evaluation of community-living individuals. *BMJ Open*. (2022) 12:e063135. doi: 10.1136/bmjopen-2022-063135

52. Derks YPMJ, Visser TD, Bohlmeijer ET, Noordzij ML. mHealth in Mental Health: how to efficiently and scientifically create an ambulatory biofeedback e-coaching app for patients with borderline personality disorder. *Int J Hum Factors Ergon*. (2017) 5:61–92. doi: 10.1504/IJHFE.2017.088418

53. Derks YP, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML. Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study. *JMIR MHealth UHealth*. (2019) 7:e13479. doi: 10.2196/13479

54. Bruno I, Lobo G, Covino BV, Donarelli A, Marchetti V, Panni AS. (2020). Technology readiness revisited: a proposal for extending the scope of impact assessment of European public services, in: *Proc 13th Int Conf Theory Pract Electron Gov*, . pp. 369–80. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3428502.3428552

55. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T. Measurement of heart rate using the polar OH1 and fitbit charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. *JMIR MHealth UHealth*. (2021) 9:e25313. doi: 10.2196/25313

56. Matlary RED, Holme PA, Glosli H, Rueegg CS, Grydeland M. Comparison of free-living physical activity measurements between ActiGraph GT3X-BT and Fitbit

Charge 3 in young people with haemophilia. *Haemophilia*. (2022) 28:e172–80. doi: 10.1111/hae.14624

57. Evenson KR, Spade CL. Review of validity and reliability of garmin activity trackers. *J Meas Phys Behav*. (2020) 3:170–85. doi: 10.1123/jmpb.2019-0035

58. Foster JI, Williams KL, Timmer BHB, Brauer SG. Concurrent validity of the garmin vivofit®4 to accurately record step count in older adults in challenging environments. *J Aging Phys Act*. (2022) 30:833–41. doi: 10.1123/japa.2021-0231

59. Garbarino M, Lai M, Bender D, Picard RW, Tognetti S. Empatica E3 #x2014; A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. *2014 EAI 4th Int Conf Wirel Mob Commun Healthc Mobihealth*. (2014), 39–42. doi: 10.1109/MOBIHEALTH.2014.7015904

60. Menghini L, Gianfranchi E, Cellini N, Patron E, Tagliabue M, Sarlo M. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology*. (2019) 56:e13441. doi: 10.1111/psyp.13441

61. Milstein N, Gordon I. Validating measures of electrodermal activity and heart rate variability derived from the empatica E4 utilized in research settings that involve interactive dyadic states. *Front Behav Neurosci*. (2020) 14:148. doi: 10.3389/fnbeh.2020.00148

62. Schuurmans AAT, de Looff P, Nijhof KS, Rosada C, Scholte RHJ, Popma A. Validity of the empatica E4 wristband to measure heart rate variability (HRV) parameters: a comparison to electrocardiography (ECG). *J Med Syst*. (2020) 44:190. doi: 10.1007/s10916-020-01648-w

63. van Lier HG, Pieterse ME, Garde A, Postel MG, de Haan HA, Vollenbroek-Hutten MMR. A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. *Behav Res Methods*. (2019) 52(2):607–29. doi: 10.3758/s13428–019-01263–9

64. Picard RW, Fedor S, Ayzenberg Y. Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emot Rev*. (2015) 8(1):1754073914565517. doi: 10.1177/1754073914565517

65. Bosch R, Chakhssi F, Noordzij ML. Acceptance and potential clinical added value of biocueing in forensic psychiatric patients with autism spectrum disorder and/or intellectual disability. *Psychiatry Res*. (2022) 313:114645. doi: 10.1016/j.psychres.2022.114645

66. Ter Harmsel JF, Smulders LM, Noordzij ML, Swinkels LTA, Goudriaan AE, Popma A. Forensic psychiatric outpatients' and therapists' Perspectives on a wearable biocueing app (Sense-IT) as an addition to aggression regulation therapy: qualitative focus group and interview study. *JMIR Form Res*. (2023) 7:e40237. doi: 10.2196/40237

67. van Doorn M, Nijhuis LA, Monsanto A, van Amelsvoort T, Popma A, Jaspers MWM. Usability, feasibility, and effect of a biocueing intervention in addition to a moderated digital social therapy-platform in young people with emerging mental health problems: A mixed-method approach. *Front Psychiatry*. (2022) 13:871813. doi: 10.3389/fpsyt.2022.871813

68. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S. System usability scale benchmarking for digital health apps: meta-analysis. *JMIR MHealth UHealth*. (2022) 10:e37290. doi: 10.2196/37290

69. Brooke J. *SUS: A 'Quick and Dirty' Usability Scale*. Jordan P, Thomas B, Weerdmeester B, editors. Usability Eval Ind. London: Taylor & Francis (1996) p. 189–94.

70. Dumas JS. User-Based Evaluations. In: Jacko JA, Sears A, editors. *Human–Computer Interact Handb*. Lawrence Erlbaum Associates, Inc, Mahwah, N.J (2003). p. 1093–117.

71. Lewis JR. Usability Testing. In: Salvendy G, editor. *Handb Hum Factors Ergon. 4de dr*. John Wiley, Hoboken, NJ (2012). p. 1267–312.

72. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud*. (2009) 4:114–23. doi: 10.5555/2835587.2835589

73. Brooke J. SUS: a retrospective. *J Usability Stud*. (2013) 8:29–40. doi: 10.5555/2817912.2817913

74. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q*. (1989) 13:319. doi: 10.2307/249008

75. Domingos C, Costa P, Santos NC, Pêgo JM. Usability, acceptability, and satisfaction of a wearable activity tracker in older adults: observational study in a real-life context in Northern Portugal. *J Med Internet Res*. (2022) 24:e26652. doi: 10.2196/26652

76. Kim T, Chiu W. Consumer acceptance of sports wearable technology: the role of technology readiness. *Int J Sports Mark Spons*. (2018) 20:109–26. doi: 10.1108/IJSMS-06-2017-0050

77. Pal D, Funilkul S, Vanijja V. The future of smartwatches: assessing the end-users' continuous usage using an extended expectation-confirmation model. *Univers Access Inf Soc*. (2020) 19:261–81. doi: 10.1007/s10209-018-0639-z

78. Oliver RL. A cognitive model of the antecedents and consequences of satisfaction decisions. *J Mark Res*. (1980) 17:460–9. doi: 10.1177/002224378001700405

79. Scheff SW. *Fundamental Statistical Principles for the Neurobiologist: A Survival Guide*. Amsterdam Boston: Academic Press (2016).

80. Health and safety executive. Work-related stress, anxiety or depression statistics in Great Britain (2021). Available at: https://www.hse.gov.uk/statistics/causdis/stress.pdf.

81. Jarczok MN, Jarczok M, Mauss D, Koenig J, Li J, Herr RM. Autonomic nervous system activity and workplace stressors—A systematic review. *Neurosci Biobehav Rev*. (2013) 37:1810–23. doi: 10.1016/j.neubiorev.2013.07.004

82. Di Nota PM, Bahji A, Groll D, Carleton RN, Anderson GS. Proactive psychological programs designed to mitigate posttraumatic stress injuries among at-risk workers: a systematic review and meta-analysis. *Syst Rev*. (2021) 10:126. doi: 10.1186/s13643-021-01677-7

83. Gerhardt C, Semmer NK, Sauter S, Walker A, de Wijn N, Kälin W. How are social stressors at work related to well-being and health? A systematic review and meta-analysis. *BMC Public Health*. (2021) 21:890. doi: 10.1186/s12889-021-10894-7

84. Oh J, Kang H. User engagement with smart wearables: Four defining factors and a process model. *Mob Media Commun*. (2021) 9:314–35. doi: 10.1177/2050157920958440

85. Reynolds LM, Cavadino A, Chin S, Little Z, Akroyd A, Tennant G. The benefits and acceptability of virtual reality interventions for women with metastatic breast cancer in their homes; a pilot randomised trial. *BMC Cancer*. (2022) 22:360. doi: 10.1186/s12885-021-09081-z

86. Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, Roldán-Jiménez C, De-Torres I, Ruiz-Muñoz M. Tools for evaluating the content, efficacy, and usability of mobile health apps according to the consensus-based standards for the selection of health measurement instruments: systematic review. *JMIR MHealth UHealth*. (2021) 9:e15433. doi: 10.2196/15433

87. Keogh A, Dorn JF, Walsh L, Calvo F, Caulfield B. Comparing the usability and acceptability of wearable sensors among older irish adults in a real-world context: observational study. *JMIR MHealth UHealth*. (2020) 8:e15704. doi: 10.2196/15704

88. Chuttur M. Overview of the Technology Acceptance Model: Origins, Developments and Future Directions. Sprouts Content (2009). Available at: https://aisel.aisnet.org/sprouts_all/290.

89. Sui A, Sui W, Liu S, Rhodes R. Ethical considerations for the use of consumer wearables in health research. *Digit Health*. (2023) 9:20552076231153740. doi: 10.1177/20552076231153740

90. Kalantari M. Consumers' adoption of wearable technologies: literature review, synthesis, and future research agenda. *Int J Technol Mark*. (2017) 12:274–307. doi: 10.1504/IJTMKT.2017.089665

91. Shei R-J, Holder IG, Oumsang AS, Paris BA, Paris HL. Wearable activity trackers-advanced technology or advanced marketing? *Eur J Appl Physiol*. (2022) 122:1975–90. doi: 10.1007/s00421-022-04951-1

92. ter Harmsel A, van der Pol T, Swinkels L, Goudriaan AE, Popma A, Noordzij ML. Development of a wearable biocueing app (Sense-IT) among forensic psychiatric outpatients with aggressive behavior: design and evaluation study. *JMIR Form Res*. (2021) 5:e29267. doi: 10.2196/29267

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 10 March 2023
DOI 10.3389/fpsyg.2023.983286

OPEN ACCESS

EDITED BY
Patrice Renaud,
University of Quebec in Outaouais, Canada

REVIEWED BY
Jakub Lickiewicz,
Jagiellonian University Medical College, Poland
Märta Wallinius,
Lund University, Sweden
Gloria Cosoli,
Marche Polytechnic University, Italy

*CORRESPONDENCE
Janna F. ter Harmsel
✉ annemieke.ter.harmsel@inforsa.nl

SPECIALTY SECTION
This article was submitted to Forensic and Legal Psychology, a section of the journal Frontiers in Psychology

RECEIVED 30 June 2022
ACCEPTED 18 January 2023
PUBLISHED 10 March 2023

CITATION
ter Harmsel JF, Noordzij ML, van der Pol TM, Swinkels LTA, Goudriaan AE and Popma A (2023) Exploring the effects of a wearable biocueing app (Sense-IT) as an addition to aggression regulation therapy in forensic psychiatric outpatients. *Front. Psychol.* 14:983286. doi: 10.3389/fpsyg.2023.983286

Exploring the effects of a wearable biocueing app (Sense-IT) as an addition to aggression regulation therapy in forensic psychiatric outpatients

Janna F. ter Harmsel[1,2]*, Matthijs L. Noordzij[3], Thimo M. van der Pol[1,2,4], Lise T. A. Swinkels[1,2], Anna E. Goudriaan[4,5,6] and Arne Popma[2]

[1]Forensic Mental Healthcare, Inforsa, Amsterdam, Netherlands, [2]Department of Child and Adolescent Psychiatry and Psychology, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [3]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands, [4]Department of Research and Quality of Care, Arkin Mental Health Institute, Amsterdam, Netherlands, [5]Department of Psychiatry, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, [6]Amsterdam Institute for Addiction Research, Amsterdam, Netherlands

Objective: Preventing and reducing violence is of high importance for both individuals and society. However, the overall efficacy of current treatment interventions aimed at reducing aggressive behavior is limited. New technological-based interventions may enhance treatment outcomes, for instance by facilitating out-of-session practice and providing just-in-time support. Therefore, the aim of this study was to assess the effects of the Sense-IT biocueing app as an addition to aggression regulation therapy (ART) on interoceptive awareness, emotion regulation, and aggressive behavior among forensic outpatients.

Methods: A combination of methods was used. Quantitatively, a pretest-posttest design was applied to explore group changes in aggression, emotion regulation, and anger bodily sensations associated with the combination of biocueing intervention and ART. Measures were assessed at pretest, after 4 weeks posttest, and after one-month follow-up. During the 4 weeks, a single-case experimental ABA design was applied for each participant. Biocueing was added in the intervention phase. During all phases anger, aggressive thoughts, aggressive behavior, behavioral control, and physical tension were assessed twice a day, and heart rate was measured continuously. Qualitative information regarding interoceptive awareness, coping, and aggression was collected at posttest. 25 forensic outpatients participated.

Results: A significant decrease in self-reported aggression was found between pre- and posttest. Furthermore, three-quarters of participants reported increased interoceptive awareness associated with the biocueing intervention. However, the repeated ambulatory measurements of the single-case experimental designs (SCEDs) did not indicate a clear effect favoring the addition of biocueing. On group level, no significant effects were found. On the individual level, effects favoring the intervention were only found for two participants. Overall, effect sizes were small.

Conclusion: Biocueing seems a helpful addition to increase interoceptive awareness among forensic outpatients. However, not all patients benefit from the current intervention and, more specifically, from its behavioral support component aimed at enhancing emotion regulation. Future studies should therefore focus on increasing usability, tailoring the intervention to individual needs, and on integration into therapy. Individual characteristics associated with effective support by a biocueing intervention

should be further investigated, as the use of personalized and technological-based treatment interventions is expected to increase in the coming years.

KEYWORDS

biocueing, biofeedback, aggression, emotion regulation, forensic psychiatry, wearable technology, mHealth

1. Introduction

Reducing aggressive behavior and criminal recidivism is an important goal in forensic psychiatry. For this purpose, several treatment interventions have been developed over the last decades. Most of these interventions are based on cognitive behavioral therapeutic principles and share elements with Aggression Replacement Training (Goldstein et al., 1998): a treatment program in which behavioral, affective, and cognitive components are combined to improve aggression regulation. However, although risk reductions in violent recidivism have been reported in several studies (Henwood et al., 2015), the overall efficacy of these treatment interventions aimed at reducing aggressive behavior has been found to be limited (Brännström et al., 2016; McIntosh et al., 2021). Since risk reductions are more pronounced among treatment completers (Henwood et al., 2015; Brännström et al., 2016), part of the limited effectivity of the current programs is probably related to low treatment adherence. Important to note is that low adherence may not only result in dropout but might also constrain the transfer of therapeutic skills into daily practice by impairing the completion of out-of-session assignments (Fletcher et al., 2011; Kazantzis et al., 2016). Furthermore, by focusing on achievement of cognitive control over emotional responses, current treatment programs might pay insufficient attention to other prerequisites of adequate anger regulation, such as awareness and recognition of psychophysiological signals associated with aggression and other challenging behaviors (McDonnell et al., 2015; Price and Hooven, 2018; Bellemans et al., 2019).

Over the last years, the number of studies focusing on the psychophysiological correlates of antisocial spectrum behavior and aggression has increased (Portnoy and Farrington, 2015; Blankenstein et al., 2021; De Looff et al., 2021; Blankenstein et al., 2022). In aggression research, psychophysiological measures such as heart rate (HR), skin conductance level (SCL), and heart rate variability (HRV) are used as indicators of, respectively, the general activity of the autonomic nervous system (ANS) and its two branches: the accelerating sympathetic nervous system (SNS) and the inhibitory parasympathetic nervous system (PNS; Branje and Koot, 2018). To understand the underlying mechanisms of aggressive behavior, ANS patterns of patients with aggression regulation difficulties have been compared to those of healthy controls, both at rest as well as in response to arousal-inducing events (i.e., reactivity measures; Blankenstein et al., 2022). Recent meta-analyses demonstrated that lower HR at rest has most consistently been found to be positively related to antisocial behavior in general and proactive aggression in particular, although the overall effect size is small (Portnoy and Farrington, 2015; De Looff et al., 2021). The research findings for reactivity measures are mixed. Regarding overall ANS reactivity, previous studies have shown increases in HR reactivity in response to emotional stimuli (Lorber, 2004; Ortiz and Raine, 2004) and provocation, associated with reactive aggression (Crozier et al., 2008). Other research results demonstrated blunted HR reactivity, suggesting diminished sensitivity to stressors such as threat or punishment associated with proactive aggression (Van Goozen et al., 2007). Furthermore, there is some evidence that reactive aggression is related to heightened SNS reactivity (Murray-Close et al., 2017; Armstrong et al., 2019; Thomson et al., 2021) and proactive aggression to blunted SNS and PNS reactivity (Patrick, 2014; Moore et al., 2018; Armstrong et al., 2019; Thomson et al., 2021). However, null findings for one or both associations have also been reported (Centifanti et al., 2013; Wagner and Abaied, 2015; Zijlmans et al., 2021; Ter Harmsel et al., 2022b). With researchers stressing the importance of studying the interaction between SNS and PNS to understand proactive and reactive aggression, instead of hypo- or hyperreactivity of the subsystems alone (Branje and Koot, 2018; Moore et al., 2018; Puhalla and McCloskey, 2020), the psychophysiological reactivity results remain largely inconclusive to date.

Psychophysiological measures are not only used to understand aggressive behavior but can also be used to predict aggressive incidents in real life. For a long time most studies aimed at identifying these physiological biomarkers were conducted in laboratory settings (Adams et al., 2017). However, in recent years first pioneering studies have been conducted in clinical settings, among inpatients with aggressive behavior. In a naturalistic study among patients with intellectual disabilities and behavioral problems, non-linear fluctuations in HRV (i.e., decreases in the first levels of increasing tension and a sudden increase when reaching extreme agitation) were found prior to outbursts (Palix et al., 2017). Studies among children and adolescents with autism spectrum disorders demonstrated that challenging or aggressive behaviors could be predicted approximately 1 min before occurrence using biosensor HR data of the preceding minutes (Goodwin et al., 2019; Nuske et al., 2019). Furthermore, aggressive incidents among forensic inpatients turned out to be preceded by significant increases in HR and SCL up to 20 min before manifestation (De Looff et al., 2019).

Some of the aforementioned challenges in treatment of forensic outpatients with aggressive behavior, such as the difficulties in recognizing physiological signals that precede aggressive incidents and the limitations in out-of-session practice, might be addressed by implementing the psychophysiological research results facilitated by the fast developments in e- and m-health technology. New interventions, such as serious gaming (Smeijers and Koole, 2019), virtual reality therapy (Klein Tuente et al., 2020), and mobile biofeedback or biocueing apps (Mackintosh et al., 2017), create opportunities to increase treatment adherence by enhancing motivation and by lowering barriers for out-of-session practice. Whereas serious gaming and virtual reality therapy are delivered on-site, at home, or in a clinical setting, biocueing could provide the patient with just-in-time behavioral support by real-time measurement in everyday life (Riley et al., 2015; Nahum-Shani et al., 2018).

This new intervention, biocueing, can be considered a derivate of traditional biofeedback, in which users are provided with real-time

physiological information and trained to influence physiological parameters, such as HRV (Lehrer, 2013) or cardiac coherence (McCraty and Zayas, 2014), by consciously alternating their (breathing) responses to the given feedback. In the process of biocueing wearable and mobile devices are used to collect and display the physiological biomarkers to the user in a direct way (Ter Harmsel J. F. et al., 2021). In contrast with traditional biofeedback, biocueing is more focused on aiding and enhancing momentary awareness of physiological sensations (i.e., interoceptive awareness) and internal emotional experiences (i.e., emotional awareness), and to a lesser extent on deliberate training of regulation techniques. In biocueing, the training component is restricted to the moments when physiological tension elevates and the user receives a just-in-time message encouraging the use of adequate coping strategies (Nahum-Shani et al., 2018). Both components of biocueing interventions – increasing interoceptive awareness and delivering just-in-time behavioral support – may be helpful to reduce and prevent aggressive incidents among forensics outpatients (Cornet et al., 2017; Ter Harmsel J. F. et al., 2021).

Given the potential of biocueing for the forensic population, we investigated the acceptability, usability, and clinical changes associated with the use of an earlier version of the Sense-IT biocueing app (Derks et al., 2019) in a two-week evaluation study among forensic outpatients (Ter Harmsel A. et al., 2021). Using the feedback of these end-users, a new version of the app was developed. The aim of the current study was to assess the effects of the new version of the Sense-IT biocueing app as an addition to aggression regulation therapy (ART) on interoceptive awareness, emotion regulation, and aggressive behavior among forensic outpatients. Quantitatively, we expected that the combination of biocueing intervention and ART would be associated with positive group changes between pretest, posttest, and follow-up on measures of aggression, emotion regulation and insight in anger bodily sensations (pretest-posttest design). Furthermore, we hypothesized group and individual increases in behavioral control and decreases in aggressive behavior as well as changes in exploratory measures anger, aggressive thoughts, physical tension and HR favoring the biocueing intervention phase (single-case experimental designs, SCEDs). For the qualitative part of this study, perceived effectivity would be indicated by patient-reported increases in interoceptive awareness, use of coping strategies, and prevention of aggressive incidents associated with the use of the Sense-IT app.

2. Materials and methods

2.1. Design

In this study, we used a combination of methods to answer the research question. Forensic outpatients receiving ART were invited to use the Sense-IT app (Derks et al., 2017, 2019) for 4 weeks. A pretest-posttest design was applied to examine changes on group level. Quantitative data were administered at the start (T0), after the 4 weeks (T1), and after one-month follow-up (T2). During the 4 week period a single-case experimental ABA design was applied for each participant, in which a baseline phase (A_1), was followed by an intervention phase (B) and a follow-up phase (A_2). In the two-week intervention phase, biocueing was added. Initially, we planned to randomize the start of the B-phase to either 5, 7, or 9 days after the start of phase A_1 for each group of three participants. However, since this procedure could not be aligned to the routines and schedules of potential participants, we had to let go of this multiple baseline aspect of the design. During all phases, the emotional state of the participants was assessed twice a day and HR was continuously measured. Qualitative data was collected at T1 *via* semi-structured interviews, enabling us to obtain a deeper understanding of patients' experiences concerning the effectivity of the Sense-IT. The study protocol and subsequent amendments were approved by the Medical Ethical Committee of Amsterdam University Medical Centre, Vrije Universiteit, Netherlands (NL63911.029.17). The study was registered in Netherlands Trial Register (NL8206).

2.2. Participants

Forensic outpatients, receiving ART at Inforsa, a forensic mental healthcare facility in the Netherlands, were recruited for participation in this study from 2020 to 2022. Potential participants were screened for eligibility by a research associate, consulting the patients' therapist. The eligibility criteria included: (1) a proven lack of anger management skills, indicated by either a recently committed violent crime and/or a high risk of committing one; (2) assignment to individual outpatient ART after multidisciplinary consultation; (3) basic understanding of mobile applications; and (4) an age of 16 years or above. The exclusion criteria included: (1) acute manic or psychotic symptoms; (2) current high risk of suicide; (3) severe addiction problems or other severe conditions requiring immediate intervention or hospitalization; and (4) insufficient understanding of the Dutch language. The first two inclusion criteria were assessed by checking the committed index defense (if applicable) and the clinical decisions recorded in the electronic patient file. The first three exclusion criteria were assessed by consulting the patients' therapist, using cut-off scores on the corresponding items of the Health of the Nations Outcome Scales (HoNOS; Wing et al., 1998). After screening and presenting the research project to eligible patients, 25 patients were willing to participate and enrolled in the study. Reasons for drop-out were: premature study termination due to COVID-19 regulations, reported stress increase related to participation in the study, other problems or obligations requiring attention, and insufficient stability. An outline of the recruitment and participation flow is displayed in Figure 1.

2.3. Procedure

Eligible patients expressing interest in the research project received a face-to-face appointment in the presence of their therapist to discuss study participation. The research associate provided the patient with a brief oral description and full written information about the study. The voluntary nature and the absence of any negative consequences refusing participation were emphasized. When the patient expressed willingness to participate, the next appointment was planned after at least 7 days, providing enough time for consideration. In this appointment written informed consent was obtained and baseline measurement (T0) was administered, which lasted approximately 60 min. After completion of the questionnaires, participants were provided with a smartwatch and mobile phone with the Sense-IT app. Participants were shown how to use the devices and were advised on charging and using the system safely. They also received a user manual. Participants used the devices independently during the following 4 weeks. They were encouraged

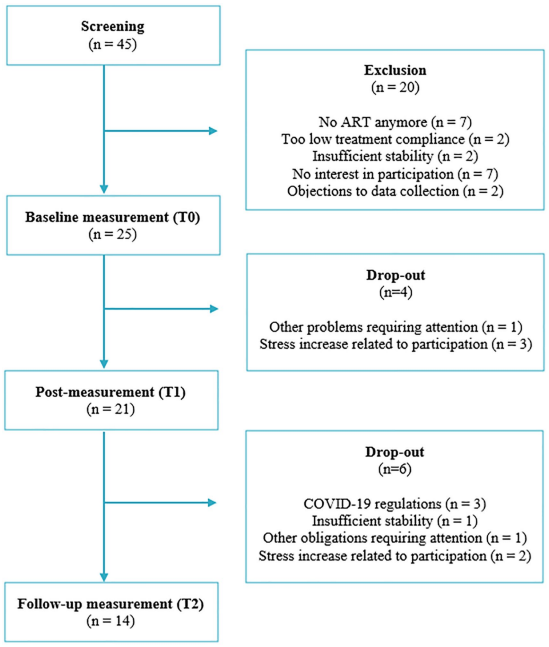


FIGURE 1
Flow chart of recruitment and participation.

to call the research associates if any problem occurred. During these weeks, the research associate met with the participants twice; once to start the intervention phase (B) and once to start the follow-up phase (A_2). In these short appointments questions were answered, if applicable, and participants were reminded to answer the daily questions. After these 4 weeks, another 60 min assessment (T1) was planned, in which both qualitative and quantitative measures were administered. One month after T1-assessment, a 30 min follow-up assessment (T2) was scheduled. An overview of outcomes and their moment of assessment is presented in Table 1.

2.4. Materials

Below, we will first introduce the studied intervention, the Sense-IT app. Next, we describe the quantitative measures used in the pretest-posttest design (change and descriptive measures) and the SCEDs (self-report and physiological measures). Finally, we describe the qualitative measures.

2.4.1. Sense-It

The newly developed version of the Sense-IT app, version 2.57 (with some minor bug fixes), was preinstalled on all smartwatches and mobile phones before distribution. The Sense-IT app was originally developed by the University of Twente and Scelta, an expert center for psychiatric patients with personality disorders (Derks et al., 2017, 2019) and modified to fit the needs of forensic outpatients assessed in an earlier study (Ter Harmsel A. et al., 2021). In the current study, we replaced the Ticwatch E, which we also used in our previous study, with the Ticwatch E2 (Mobvoi, Ltd). Compared to its predecessor the E2 is sleeker, more sophisticated, and has a slightly longer battery life. Connection with the mobile phone, the Moto C Plus or the Moto E6 Play (Google, LLC), was established *via* Bluetooth. The Sense-IT system reads the physiological data measured by the photoplethysmography (PPG) sensor and stores the data in a local database on the smartphone itself. The build-in algorithm compares the current HR to the user's mean HR at baseline and calculates a level between −3 and 5 using the standard deviation of the baseline measurement. In the current study, we further refined the baseline measurement procedure. Ultimately, baseline measurement was performed during T0, included at least 1 min of sitting in quiet, 1 min of social interaction, and 1 min of walking activity to account for sufficient variation, and lasted until the PPG sensor received 500 reliable HR measures. Our starting values for HR and heart rate variance thresholds were in line with published norms indicating a mean HR around 80 (SD ~ 7; Umetani et al., 1998). More information on the baseline procedure can be found in the Supplementary material. After baseline measurement, the real-time HR level is visually displayed on the smartwatch and changes when the HR level decreases or increases more than one level. After every three participants, we checked whether we had to refine the settings to improve usability, for example accounting for feedback about receiving too many notifications. Ultimately, the sensitivity of the app was set to low (expanding the ranges between levels by multiplying the standard deviation with a 1.5 factor) and the notifying vibrations were given at levels 4 and 5 above baseline. The Sense-IT app also detects (physical) activity categories using the accelerometer and Google activity recognition algorithms, allowing the user to receive notifications for certain activity profiles. In this study, we ended up offering notifications for low activity profiles (i.e., sitting still, walking) only. In the user interface on the smartphone, users can turn the app on and off, and open a timeline of all measurement events and level changes detected by the system. Users can add notifications to events in the timeline and report their subjective level of arousal, which might particularly be useful when tension increases. Users can also define a personalized message that is displayed when their physiological arousal exceeds a predefined level. In this study, this supportive message and an accompanying question to rate subjective stress were displayed at levels 4 and 5. The user interface also presented information about connection and synchronization, as well as a settings page which was protected by a password to prevent unwarranted changes. Screenshots of the Sense-IT app are displayed in Figure 2.

2.4.2. Quantitative measures

2.4.2.1. Pretest-posttest design

2.4.2.1.1. Change measures

At T0, T1 and T2 primary and secondary measures were administered to explore relevant changes on group level. The Aggression Questionnaire-Short Form (AQ-SF; Buss and Perry, 1992), a 12-item 5-point Likert scale self-report questionnaire, was used to assess changes in different types of aggressive behavior over the past 4 weeks: physical aggression, verbal aggression, anger, and hostility. Therapists evaluated aggressive behavior of their patients during the same period with the Modified Overt Aggression Scale (MOAS; Knoedler, 1989), a 4-item observation scale differentiating verbal aggression, aggression against property, auto-aggression, and physical aggression. The therapists based their scores on observed incidents (if applicable), information from others (if applicable), and on patients' retrospective reports of aggressive incidents during the weekly sessions. Another self-report measure, the Anger Bodily Sensations Questionnaire (ABSQ; Zwets et al., 2014), consisting of 18 items with a 5-point Likert scale, was administered to assess changes in psychophysiological awareness.

TABLE 1 Overview of outcomes, measures, and moment of assessment.

Outcome	Measure	T0	Tx (during SCED)	T1	T2
Aggressive behavior	AQ-SF	+		+	+
	MOAS	+		+	+
HR measures	Biosensor		+		
Emotional state	EMA		+		
Emotion regulation	DERS	+		+	+
Anger regulation	STAXI-2	+			
Anger bodily sensations	ABSQ	+		+	+
Aggression type	RPQ	+			
Psychopathy	YPI-s	+			
Judicial history	File information	+			
Demographic information	DQ	+			
Evaluation of the Sense-IT	Qualitative interview			+	+
System usability	SUS			+	

AQ-SF, Aggression Questionnaire-Short Form; MOAS, Modified Overt Aggression Scale; HR, Heart Rate; EMA, Ecological Momentary Assessment; DERS, Difficulties in Emotion Regulation Scale; STAXI-2, State–Trait Anger Expression Inventory; ABSQ, Anger Bodily Sensations Questionnaire; RPQ, Reactive Proactive Questionnaire; YPI-s, Youth Psychopathic Traits Inventory-Short Version; DQ, Demographic Questionnaire; SUS, System Usability Scale.

Furthermore, the Difficulties in Emotion Regulation Scale (DERS; Gratz and Roemer, 2004), a 36-item 5-point Likert scale self-report questionnaire, was administered. This questionnaire is used to assess six dimensions of emotional processing: non-acceptance of emotional responses, difficulty engaging in goal-directed behavior, impulse control difficulties, lack of emotional awareness, limited access to emotion regulation strategies, and lack of emotional clarity.

2.4.2.1.2. Descriptive measures

At T0, several other secondary measures were administered to describe the sample. A self-developed demographic questionnaire was used to gather information regarding age, gender, ethnicity, education, and past offenses. The most recent DSM-5 main psychiatric diagnosis for each participant was retrieved from the electronic patient record. To gain a better understanding of the type and nature of aggressive and antisocial behavior, three other self-report measures were administered at baseline: the Reactive Proactive Questionnaire (RPQ; Raine et al., 2006), Youth Psychopathic Traits Inventory-Short Version (YPI-s; van Baardewijk et al., 2010), and State–Trait Anger Expression Inventory (STAXI-2; Spielberger et al., 1999).

2.4.2.2. Single-case experimental designs

2.4.2.2.1. Self-report measures

During the 4 weeks of the ABA designs, Ecological Momentary Assessment (EMA) was used to assess the emotional state of the participants. For this reason, six questions were designed based on the items of the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). Participants received prompts to answer these questions twice a day, at predetermined times fitting into the daily schedule of the particular participant. They were asked to rate the extent to which they experienced behavioral control and aggressive behavior (primary measures) as well as anger, aggressive thoughts and physical tension (exploratory measures) during the preceding part of the day. A question investigating feelings of happiness was added to balance the questions. A 5-point Likert scale was used for each question, reaching from 'very slightly/not at all' to 'extremely'.

2.4.2.2.2. Physiological measures

During the 4 weeks of the ABA designs, HR was continuously measured while the Sense-IT app was used. The Sense-IT app also registered the baseline settings, kept track of the levels (i.e., a value between −3 and 5) and the activity profiles (i.e., running, cycling, and sitting still), and whether biocueing was active (phase B) or not (phases A_1 and A_2).

2.4.3. Qualitative measures

At T1, a semi-structured interview was conducted. This interview included questions about feasibility and usability of the devices, advantages and disadvantages of the Sense-IT app, and recommendations for further improvement. In this article, we focused on three questions regarding the perceived effectivity of the Sense-IT app on interoceptive awareness, use of coping strategies, and prevention of aggressive incidents. A more in-depth analysis of patients' perspectives on use and implementation of the Sense-IT app is presented elsewhere (Ter Harmsel et al., 2022a).

2.5. Data analysis

2.5.1. Quantitative data analyses

2.5.1.1. Pretest-posttest design

The quantitative data (AQ-SF, MOAS, ABSQ, and DERS) were analyzed using SPSS (version 27, IBM Corp). After checking the normality assumptions for main scales and subscales and given the small sample size (particularly for the comparisons with T2), we decided to use the nonparametric equivalent of the paired *t*-test, the Wilcoxon Matched Pairs Test. To make efficient use of the available data two missing items on the DERS, for two different participants, were replaced

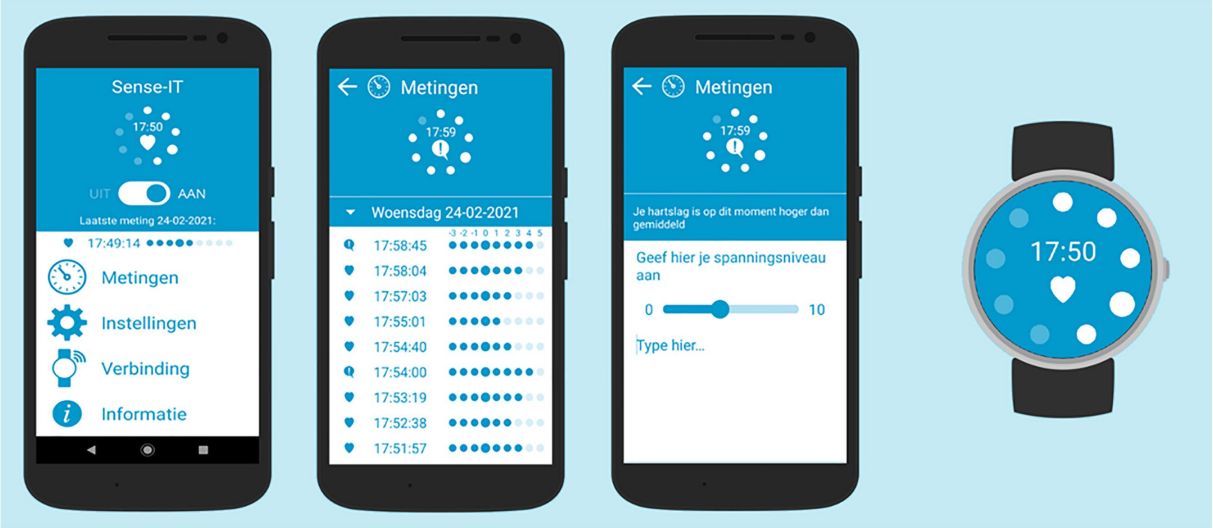


FIGURE 2
Screenshots of the Sense-IT app (version 2.57) with the main screen (presenting four menus: measurements, settings, connectivity, and information), measurement screen (presenting HR levels), notes screen (presenting default message, question to rate subjective level of arousal, and space for personal notes) and one of the watch faces.

by imputing the individual mean score on this questionnaire at that moment of assessment.

2.5.1.2. Single-case experimental designs

In order to analyze the SCED data, all EMA and HR measures were divided into the three phases (A_1, B, and A_2), using the track record of the Sense-IT app. For EMA, responses were considered as belonging to the last preceding prompt, unless the response was given less than 30 min before the next prompt. In case of multiple responses within 30 min, the response that deviated the most from the specified prompting time was discarded. In case of phase ambiguities, EMA responses were assigned to the phase to which the majority (>50%) of the period over which they reported (i.e., the time between prompts) belonged. For HR, measurements with specific activity profiles (i.e., running, cycling, car driving) were disregarded from the measurements to focus on HR data in no (i.e., sitting) to limited (i.e., walking) movement scenarios. Furthermore, the HR data was corrected for very low and high values (< 50 bpm and > 190 bpm). To calculate mean and standard deviation per day part, HR data was split into daytime (08:00 AM – 04:59 PM) and evening measures (05:00 PM - 01:59 AM). A day part was considered missing when less than 500 HR measures were present. When participants had no access to the Sense-IT app and its associated devices for at least three days (e.g., due to vacation), the corresponding period was not included in the analysis.

After data preprocessing a visual analysis, considered as the primary method in SCED research (Kazdin, 2019), was performed on the selected EMA variables (i.e., anger, physiological stress, aggressive thoughts, aggressive behavior and behavioral control) and HR variables (i.e., mean and standard deviation). For participants with at least 5 data points per phase (Bolger and Laurenceau, 2013), we graphically compared the direction and rate of change (i.e., the slopes of the regression lines) between the different phases for each variable. We made plots for each participant separately as well as for the entire group of eligible participants. First, we used (R Core Team, 2017) lme function from the nlme package (Pinheiro and Bates, 2000) to apply a multilevel (two-level) piecewise regression approach analyzing the effects between phases per variable, for all the eligible participants on group level. Second, we performed (one-level) piecewise regression analyses per variable for each participant using the R-based MultiSCED web application (Declercq et al., 2020). The unstandardized parameter estimates of each variable at the start of the study (intercept, B0), the developmental effect in the variable over time in a particular phase (time, B1), an immediate variable change when transitioning into the intervention phase (phase, B2) and a comparison of variable change over time in the intervention phase compared to the baseline phase (time x phase, B3) were calculated. For results, we reported the estimates B1 (for both baseline and intervention phase), B2, and B3. To account for Type-I errors, we used a Bonferroni-corrected value of p ($p \leq 0.01$) by dividing the critical value of p ($\alpha = 0.05$) by the number of comparisons (five). In addition, we assessed effect sizes on group and individual level for the EMA variables for which we expected a specific direction of change (i.e., behavioral control and aggressive behavior) by calculating the Improvement Rate Index (IRD; Parker et al., 2009). We calculated this nonparametric overlapping index, comparing the improvement rate between two phases, using an online single-case effect size calculator (Pustejovsky et al., 2021).

2.5.2. Qualitative data analyses

We organized and analyzed the data of the qualitative interview using Microsoft Excel. Dichotomous responses were described as relative results. Two researchers (JtH and LS) independently ranked the three most informative textual responses regarding interoceptive awareness, use of coping strategies, and prevention of aggressive incidents. The final quotations were selected by discussion between two researchers (JtH and LS), until consensus was reached, and translated from Dutch into English.

10.3389/fpsyg.2023.983286

3. Results

3.1. Descriptive characteristics

The majority of the 25 forensic outpatients who participated in this study was male (92%) and born in Netherlands (92%). For most of them, treatment was a mandatory part of their conditional sentence (64%), mainly imposed because of a violent index offense (94%). A large proportion (73%) reported problems in the family of origin: domestic violence and substance abuse were most frequently reported, but criminal behavior and psychological problems were also mentioned. All descriptive characteristics are summarized in Table 2.

TABLE 2 Demographic characteristics (N=25).

Outcome	Mean (SD)	*n* (%)
Age	29.88 (10.51)	
Gender		
Male		23 (92%)
Female		2 (8%)
Migration background		
First-generation migration background		2 (8%)
Second-generation migration background		14 (56%)
Dutch background		9 (36%)
Educational background		
None		1 (4%)
Primary education		4 (16%)
Junior secondary education		14 (56%)
Senior secondary education		6 (24%)
Indication of mild intellectual disability (SCIL)		9 (36%)
Main psychiatric classification according to DSM-5		
Disruptive disorder		10 (40%)
Substance use disorder		2 (8%)
Posttraumatic stress disorder		2 (8%)
Borderline personality disorder		2 (8%)
Other specified personality disorder		7 (28%)
Other disorder		2 (8%)
Mandatory treatment		16 (64%)
Past offenses (official records)		
0		8 (32%)
1 or 2		6 (24%)
3 to 5		4 (16%)
6 to 10		3 (12%)
More than 10		4 (16%)
Aggression type (RPQ)		
Reactive aggression	14.72 (4.39)	
Proactive aggression	7.53 (4.61)	
Anger and anger regulation (STAXI-2)		
State anger	18.80 (6.85)	
Trait anger	23.76 (6.97)	
Anger expression index	50.72 (10.81)	
Psychopathy (YPI-s)		
Interpersonal dimension	12.40 (4.73)	
Affective dimension	11.48 (4.39)	
Behavioral dimension	15.56 (3.35)	

3.2. Pretest-posttest results

First, we analyzed the results of the quasi-experimental designs with pre-, post-, and follow-up measurements. The mean scores and standard deviations on clinical outcomes aggression (AQ-SF and MOAS), insight in anger bodily sensations (ABSQ), and emotion regulation difficulties (DERS), for each moment of assessment, are presented in Table 3. For statistical testing, data of 20 participants could be used to explore the difference between T0 and T1; and data of 14 participants for the differences between T0 and T2, and T1 and T2. The results of Wilcoxon Matched Pairs tests indicated that self-reported aggression decreased significantly between T0 (Median = 35.5) and T1 (Median = 31.5); $Z = -2.043$, $p = 0.041$. No significant decreases in aggression were found between the other moments of assessment. For therapist reported aggression level, emotion regulation difficulties, and insight in anger bodily sensations no significant differences were found between pre-, post- and follow-up assessment. For this sample, three of the outcome measures changed in the expected direction between T0 and T1, and one measure (anger bodily sensations) changed in the opposite direction.

3.3. Single-case experimental design results

Next, we analyzed the results of the ABA designs. In order to select the participants with sufficient data points, we started by investigating data availability. One participant did not start using the Sense-IT app, one participant quit after phase A_1 and seven participants stopped using the app after phase B. The compliance to EMA, defined as the ratio of the number of completed EMA questions in relation to the total number of EMA prompts per phase, ranged from 43.7% in Phase A_1, to 24.7% in Phase B and 16.0% in Phase A_2. For EMA, only 3 participants met the criterion of at least 5 data points for all phases. The compliance to HR measurement, the ratio of available daytime or evening measures in relation to the maximum amount of these measures per phase, ranged from 38.5% in Phase A_1, to 29.9% in Phase B and 13.5% in Phase A_2. For HR, none of the participants had sufficient measurements in all phases. Therefore, only the results of the baseline phase (A_1) and intervention phase (B) were used for further analysis: for EMA, 9 participants had sufficient data in phase A_1 and B; for HR this applied to 10 participants.

First, we applied a multilevel piecewise regression approach (two-level) and visual analyses to analyze the effects between phases for five EMA variables and two HR variables on group level, using the data of all eligible participants. We inspected B1 (for both baseline and intervention phase), B2 and B3. On group level, we found no significant developmental effects (neither for baseline nor for intervention phase), no significant immediate changes when transitioning into the intervention phase, and no significant interaction effects on any of the variables. All group-level parameter estimates are presented in the Supplementary material. The individual and group-level results of the two primary EMA measures, behavioral control, and aggressive behavior, are illustrated in Figures 3, 4. For exploratory measures, see Supplementary material. Improvement rate differences (IRDs) for these outcomes on group level were.29 for behavioral control

TABLE 3 Overview of clinical outcomes at pre-, post- and follow-up measurement.

Outcome	T0 (*N*=25) Mean (SD)	T1 (*N*=20) Mean (SD)	T2 (*N*=14) Mean (SD)
Aggression, self-report (AQ-SF)	32.44 (9.51)	30.80 (8.68)	28.07 (10.40)
Aggression, therapist-report (MOAS)	5.24 (5.46)	3.48 (3.30)	4.92 (6.46)
Emotion regulation difficulties (DERS)	100.40 (25.55)	93.14 (25.52)	89.20 (24.98)
Anger bodily sensations (ABSQ)	49.96 (16.12)	45.70 (13.74)	42.07 (14.11)

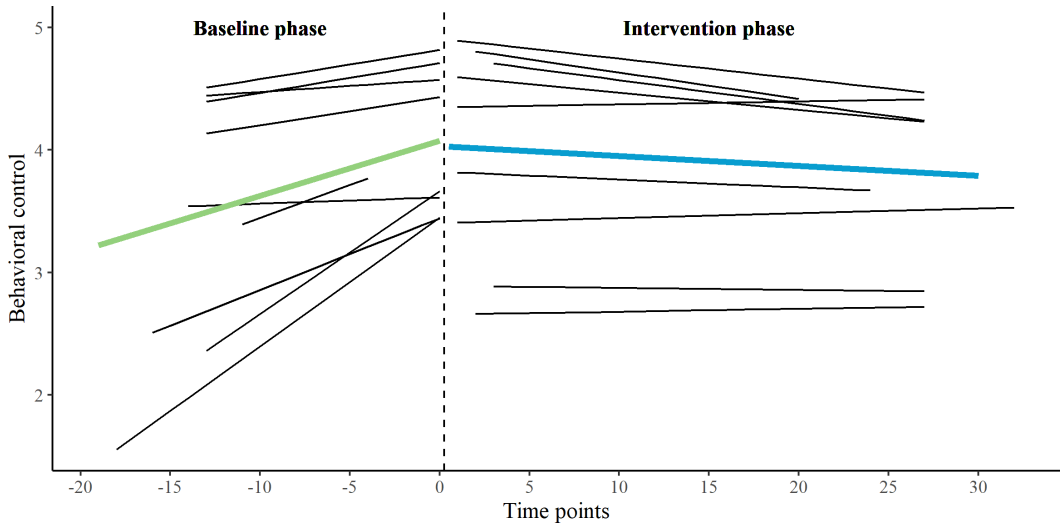


FIGURE 3
Combination of one- and two-level regression results for primary EMA measure behavioral control, measured twice a day on a 5-point Likert scale, in baseline phase A_1 and intervention phase B.

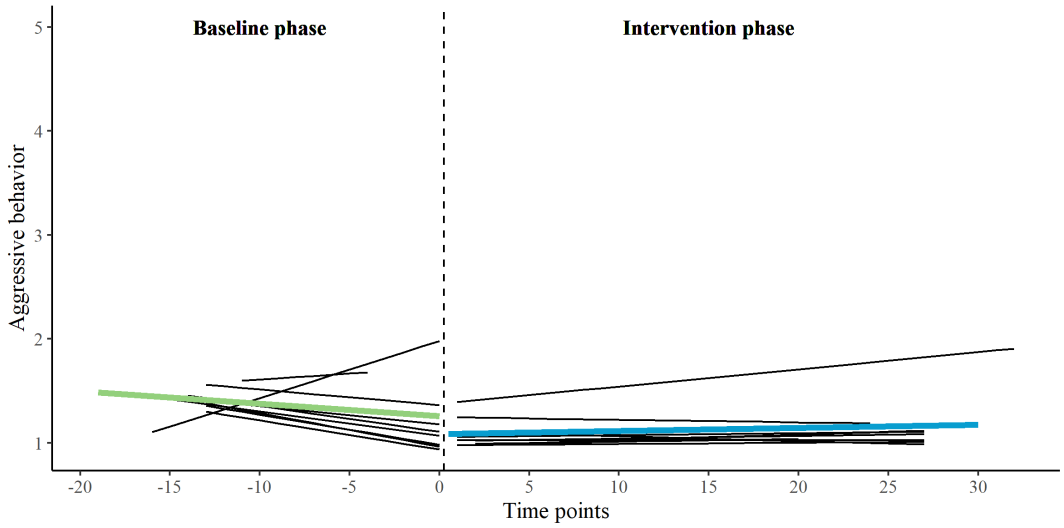


FIGURE 4
Combination of one- and two-level regression results for primary EMA measure aggressive behavior, measured twice a day on a 5-point Likert scale, in baseline phase A_1 and intervention phase B.

(increasing direction) and 0.26 for aggressive behavior (decreasing direction), indicating a small effect size of the combination of biocueing intervention and ART on these outcome measures (Parker et al., 2009).

Subsequently, piecewise regression analyses and visual analyses were conducted for each of the eligible participants separately (one-level), using MultiSCED. Significant developmental effects (for both baseline and intervention phase), immediate changes when transitioning into the intervention phase, and interaction effects are reported in the Supplementary material. An overview of all unstandardized parameter estimates for each participant is available upon request from the first author.

Improvement rate differences (IRDs) for the two primary EMA outcomes on the individual level ranged from 0.05 to 0.51 for aggressive behavior (decreasing direction) and from.05 to.55 for behavioral control (increasing direction), indicating small effect sizes with some exceptions to moderate (Parker et al., 2009).

Finally, we zoomed in on the three participants in which we found visually interesting patterns and significant interaction effects to enhance clinical understanding of the results. The names of the participants are fictitious.

Ryan

This participant, aged between 30 and 35, had severe aggression regulation problems. At baseline (T0), Ryan achieved a high score (9th decile) on the AQ-SF compared to other outpatients with violent behavior. He reported predominantly reactive aggression on the RPQ. Furthermore, Ryan experienced many emotion regulation difficulties (DERS). His anger expression index on the STAXI-2 (95th percentile) shows that he tended to express his emotions more outward than inward, and that his ability to regulate his emotions was very low. Using the piecewise regression results and visual analysis (see Figure 5), his feelings of anger, aggressive thoughts, aggressive behavior and mean HR all seem to decrease during the baseline phase (A_1). Most remarkable in the intervention phase (B) is the increase in aggressive thoughts. Both the decrease in aggressive thoughts during the baseline phase and the increase in the intervention phase reached statistical significance. However, Ryan reported that he did not express these thoughts in aggressive behavior, as indicated by the flat line. When the patterns in both phases were compared, his outcomes regarding aggressive thoughts were significantly in favor of the baseline phase. No significant differences between phases were found on other variables. At post-test (T1) Ryan reported a substantially lower score on the AQ-SF (6th decile) compared to baseline, but a higher score on the DERS. He reported that the Sense-IT biocueing app did not work for him. He noticed no effect of biocueing on his awareness of physiological signals of tension, use of adequate coping, or prevention of aggressive behavior. Ryan mentioned that the Sense-IT app signaled tension when there was none and did not signal tension when there was; questioning the accuracy of the feedback.

Eric

This participant, aged between 40 and 45, also struggled with aggression regulation problems. At baseline, he scored average on the AQ-SF compared to the norm group (6th decile). Eric also reported predominantly reactive aggression on the RPQ. He experienced emotion regulation difficulties (DERS) to an average degree. The anger expression index (81st percentile) indicates that Eric also tended to direct his anger more outward than inside, and that his regulation skills were quite low. The piecewise regression results and visual analysis (see Figure 6) illustrate that anger and aggressive behavior seem to increase in the baseline phase (A_1), while aggressive thoughts seem to decrease. The increase in aggressive behavior reached statistical significance. In the intervention phase (B), these variables all seem to change in decreasing direction. None of these decreases reached statistical significance. Mean HR seems stable. When the patterns in aggressive behavior were compared between both phases, his outcomes did significantly favor the intervention phase. No significant differences between phases were found on other variables. Compared to baseline, Eric also achieved a lower score on the AQ-SF (4th decile) at post-test. His score on the DERS remained the same. Eric reported no effect of biocueing on his awareness of physiological signals of tension, use of adequate coping, or prevention of aggressive behavior. He stated that he was not inclined to act upon the physiological feedback he received.

Joshua

This participant, aged between 20 and 25, also had aggression regulation problems. At baseline, Joshua scored average compared to other outpatients with violent behavior (6th decile). He reported similar

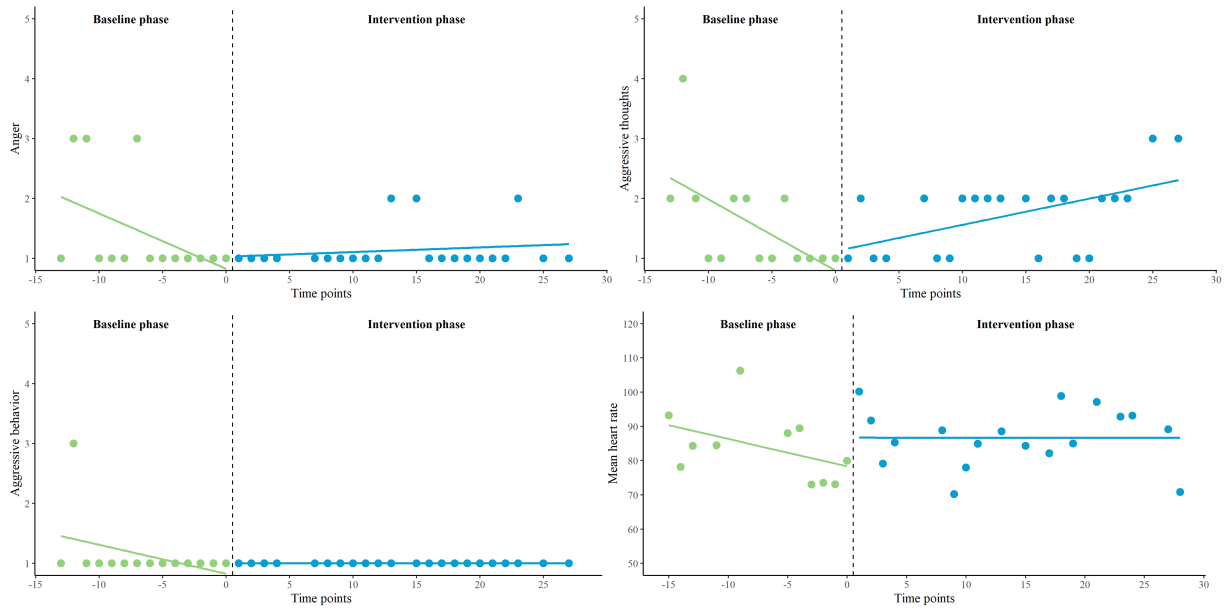


FIGURE 5
Representation of Ryan's EMA measures (anger, aggressive thoughts and aggressive behavior) and HR (average), measured twice a day on a 5-point Likert scale, in baseline (A_1) and intervention phase (B).

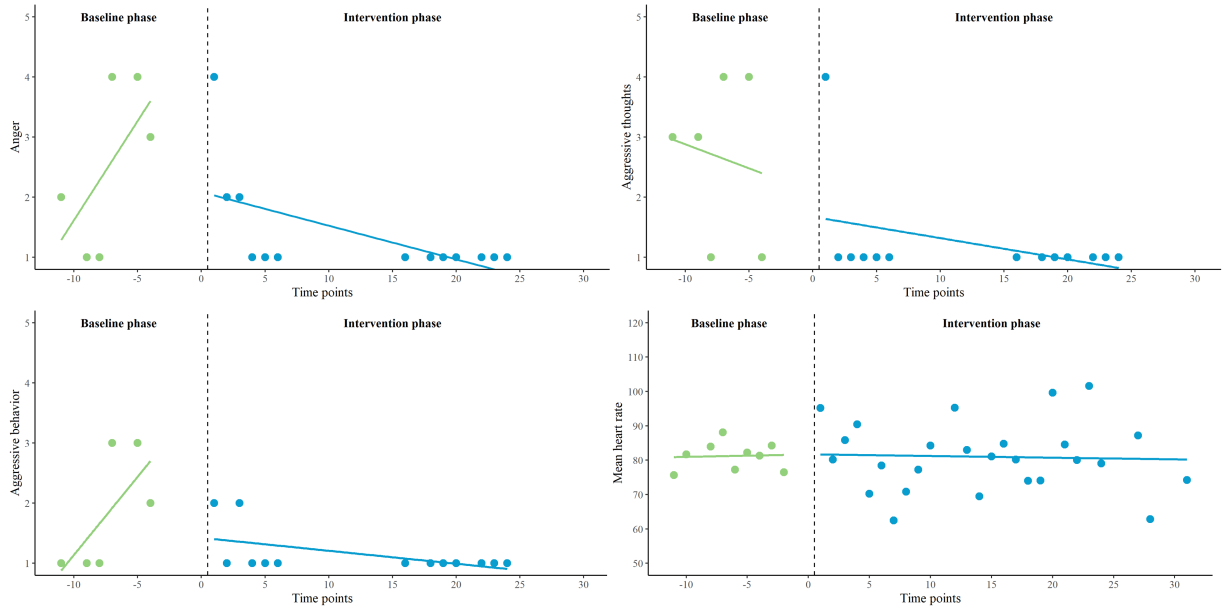


FIGURE 6
Representation of Eric's EMA measures (anger, aggressive thoughts and aggressive behavior) and HR (average), measured twice a day on a 5-point Likert scale, in baseline (A_1) and intervention phase (B).

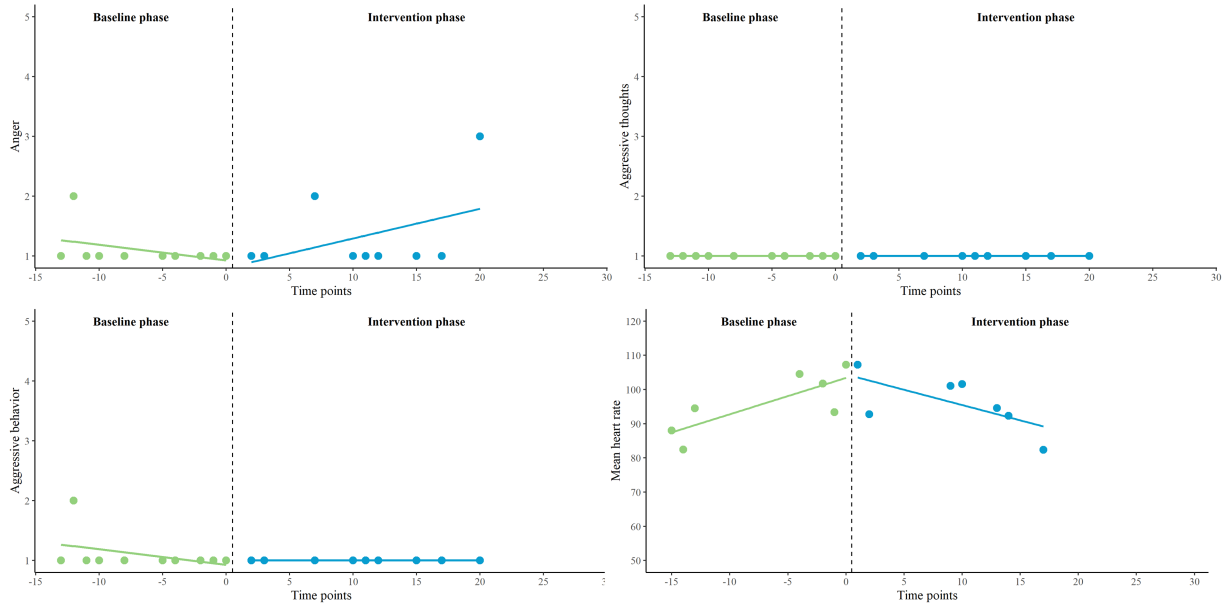


FIGURE 7
Representation of Joshua's EMA measures (anger, aggressive thoughts and aggressive behavior) and HR (average), measured twice a day on a 5-point Likert scale, in baseline (A_1) and intervention phase (B).

levels of reactive and proactive aggression on the RPQ. Joshua experienced emotion regulation problems (DERS) to an average degree. His anger expression index (60th percentile) indicated that he expressed his anger in both directions, and more inward compared to the other Ryan and Eric. His regulation skills were average to good. The piecewise regression results and visual analysis (see Figure 7) revealed that he had experienced little anger, aggressive thoughts, and aggressive behavior. No significant in- or decreases in phases and no significantly different patterns between phases were found for these variables. However, the decrease in mean HR in the intervention phase (B) significantly differed from the increase in the baseline phase (A_1), which might favor the intervention phase. Compared to baseline, Joshua scored slightly lower on the AQ-SF (5th decile) at post-test. The score on the DERS also decreased slightly, but was still in the same range. Joshua did notice a positive effect of biocueing on his awareness of physiological signals of tension. He explained that the app helped him not to get stuck in anger

by using adequate coping strategies, such as seeking distraction and clearing his mind. He did not notice an effect on prevention of aggressive behavior as he felt he had not been aggressive during the research period.

3.4. Qualitative results

Qualitatively, we focused on the perspectives of patients regarding the perceived effectivity of the Sense-IT app on interoceptive awareness, use of coping strategies, and prevention of aggressive behavior. The responses of the participants can be found in Figure 8. As shown, the majority of the forensic outpatients (76%) reported increased insight into physiological signals of tension, after using the app. For example, one of these participants mentioned: *"I felt the tension building and when it [the watch] started to vibrate I recognized that I was angry and went to do something else"* (P16) and another reported: *"That my heart rate is often high even though I do not think so myself."* (P3) However, other patients indicated: *"I was often alerted to tension when I already knew it."* (P21) and "*The app sends false notifications: it vibrates while nothing is going on and not when you are stressed out*" (P1). Approximately half of the participants (48%) felt supported by the Sense-IT app to use coping strategies in order to reduce stress. For example, one participant reported: *"When the watch vibrated I took the notification into account, for example by coming to myself and taking a breath."* (P17) and another said: *"I went to do something else and cleared my mind instead of dwelling in anger."* (P10). In some cases, the coping strategies used may have shed light on less adequate behavioral patterns:*"[When the watch vibrated] I went to smoke a joint and got calmer. I also went gaming."* (P9). Another participant mentioned: *"I will not be guided by a watch. Did look at the notification, but did not react to it. Screw it, I'll just stay angry."*(P8), indicating that at least some motivation to change behavior and openness to feedback are necessary prerequisites to benefit from the app. Moreover, it shows that it can be difficult for some patients to distinguish anger from aggression. Finally, about one-third of the participants (38%) reported that using the app might have helped them to prevent aggressive behavior in some cases. One participant responded: "*Maybe. By participating in this study, I became more aware to reduce my anger when the watch indicated, for example when stressed at work."* (P25). Other participants mentioned that they were able to calm themselves in some cases, but not in all: *"Sometimes I took it easy, but sometimes I was just angry and did not do anything."* (P9) or: *"Sometimes you do not think about the watch, then things go fast and something happens*." (P7). Another participant stressed the boundaries of an app: *"It takes more than that. When I am very aggressive, I will not be stopped by a vibrating watch or a mobile with questions*" (P16). Furthermore, several participants reported increased stress and irritation by the number of notifications and the perceived inaccuracy of the app: *"Actually, it increased my frustration and irritation. I had to actively suppress not throwing the thing away."*(P21). Other participants indicated that this last question was difficult to answer as they had not experienced aggressive outbursts in the past period.

4. Discussion

In the current study, we explored the effects of a new version of the Sense-IT biocueing app on interoceptive awareness, emotion regulation, and aggressive behavior in a forensic outpatient population. In this study, the Sense-IT app was added to regular ART. Quantitatively, we examined group changes on measures of aggression, emotion regulation and insight in anger bodily sensations between pretest, posttest, and follow-up (pretest-posttest design), as well as group and individual changes in behavioral control, aggressive behavior, anger, aggressive thoughts, physiological tension and HR in the intervention phase compared to the baseline phase (SCEDs). In addition, we qualitatively assessed patient-reported changes in interoceptive awareness, use of coping strategies, and prevention of aggressive incidents associated with the use of the Sense-IT app.

The results of the pretest-posttest design showed a significant decrease in self-reported aggression between pretest and posttest, indicating a positive effect associated with the combination of the Sense-IT biocueing intervention and ART. In addition, although on visual inspection emotion regulation difficulties decreased in this sample, no significant effects were found. Furthermore, no significant differences were found for therapist-reported aggression level and insight into anger bodily sensations. The quasi-experimental nature of this design prohibits attribution of the found effect to either the combined intervention or the regular therapy alone. However, the fact that the significant decrease in aggression was only found between pretest and posttest (the period the biocueing was added) and not for the comparison with follow-up (the period regular therapy was continued without the biocueing intervention), is interesting, and could

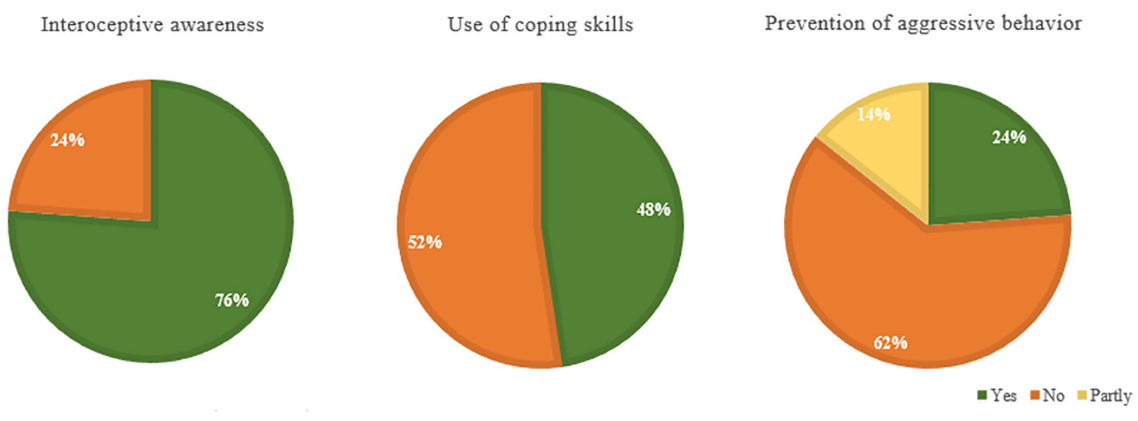


FIGURE 8
Perceived effectivity of the Sense-IT app on interoceptive awareness, use of coping skills and prevention of aggressive behavior according to forensic outpatients.

be further investigated in future studies using a controlled research design.

Due to low compliance to EMA and a low amount of sufficient HR measures per day part, data analysis of the SCEDs was limited to EMA data of 9 participants and HR data of 10 participants. Focusing on the primary measures, aggressive behavior and behavioral control, an interaction effect favoring the biocueing intervention was found for one participant (a decrease in aggressive behavior in the intervention phase compared to an increase in the baseline phase). Regarding exploratory measures, a reverse pattern was found in another participant (an increase in aggressive thoughts in the intervention phase compared to a decrease in the baseline phase). For another participant, mean HR decreased in the intervention phase compared to an increase in the baseline phase. On group level, we found no significant developmental effects in the baseline and intervention phase, no significant immediate changes when transitioning into the intervention phase, and no significant interaction effects on any of the variables. Overall, effect sizes were small, with some individual exceptions to moderate.

In contrast, the qualitative results do indicate positive changes related to the use of the Sense-IT biocueing app, such as increased interoceptive awareness among the majority of the participants, perceived support in the use of coping strategies by half of the participants, and prevention of some aggressive incidents by one-third. Although most participants reported increased insight into physiological signals of tension, results show that the step from insight toward adequate emotion regulation requires more attention. Furthermore, results seem to indicate that a certain amount of motivation to change behavior and openness to feedback are necessary to benefit from the just-in-time behavioral support delivered by the Sense-IT app. Although the usability of the Sense-IT app was acceptable (Ter Harmsel et al., 2022a), some patients reported increased stress and irritation by the number of notifications and the perceived inaccuracy of the app.

In sum, out of the subgroup of patients who qualitatively reported positive changes associated with the use of the biocueing intervention, we only found a significant quantitative change favoring the intervention phase for one patient with sufficient data for single-case analysis. We therefore conclude that, whereas the quantitative results of the pretest-posttest design and the qualitative results indicated positive changes associated with the combination of biocueing intervention and ART, the repeated ambulatory measurements of the SCEDs do not indicate a clear effect favoring the addition of a biocueing intervention.

First of all, we would like to shed some light on the findings related to interoceptive awareness. The awareness of bodily sensations has been identified as an important component in the process of emotional awareness (Lane and Schwartz, 1987; Calì et al., 2015), which is, in turn, an essential building block for adequate emotion regulation (Füstös et al., 2013; Gross and Jazaieri, 2014). In our study, the majority of patients qualitatively reported increased interoceptive awareness, although quantitively no significant difference and insight in anger bodily sensations was found. Several factors might have contributed to this finding. First, patients might have developed a more accurate view of their physiological sensations in angry interactions by using the biocueing intervention. Second, while interoceptive awareness primarily concerns perception, anger- related interoceptive awareness also entails some interpretation of bodily signals (Bellemans et al., 2018), and can therefore be considered a more demanding skill. Third, some patients had difficulty differentiating between components of the Sense-IT biocueing app (delivering a real-time visualization of their heart rate level and just-in-time behavioral support) and the integrated daily EMA questions that were used for research purposes. Since some patients reported that 'the questions' (unspecified) were really helpful to reflect on their emotional state, the qualitative measure of interoceptive awareness associated with the use of the biocueing intervention may have been somewhat diluted by perceived increases in emotional awareness by responding to the EMA questions across all phases. Fourth, recent research states that (Garfinkel et al., 2015; Forkmann et al., 2016) interoceptive awareness should be considered as a metacognitive process, integrating both interoceptive sensibility (i.e., self-evaluated assessment of subjective interoception) and interoceptive accuracy (i.e., performance on objective tests of heartbeat detection). According to this model of interoception, our study focused on the facet of interoceptive sensibility and thereby lacked information regarding interoceptive accuracy and interoceptive awareness as a metacognitive process. Since the feedback was perceived as inaccurate by a substantial part of the patients, which may be partly related to technical issues but may also be explained by limited interoceptive capabilities, understanding the role of interoception in using biocueing among forensic outpatients could have been enhanced if we had included an interoceptive accuracy measure and had used a clearer definition of interoceptive terms (Khoury et al., 2018).

Second, we would like to focus on the findings regarding emotion regulation. Although half of the participants qualitatively reported that they felt supported by the app to use coping strategies in order to reduce stress, and on visual inspection emotion regulation difficulties decreased in this sample, no significant effects associated with the combination of biocueing intervention and ART were found in the pretest-posttest design. We also found no group or individual increases in behavioral control favoring the biocueing intervention in the SCEDs. Several factors might have contributed to these findings. First, motivation to change and feedback receptivity varied among the participants. Although these factors might not be required to take advantage of the component of the Sense-IT app aimed at increasing interoceptive awareness, feedback receptivity turned out to be quite essential to benefit from the behavioral support component of the Sense-IT app, even among non-psychiatric samples without motivational difficulties (Lentferink et al., 2021). For future use, it is therefore important to assess whether patients are open to receiving feedback and willing to try out different emotion regulation strategies. Related to this, some patients had (very) high expectations of what the app should deliver, which might have led to disappointment when subjectively experienced tension was not noticed or when they received behavioral support messages while they felt relaxed and did not notice tension. Although biocueing interventions can identify substantial increases in arousal by measuring HR, they are unable to provide a flawless recognition of subjectively experienced stress and cannot determine valence in order to specify emotion categories (Siegel et al., 2018). As suggested in recent research in which patients' reported similar feedback (Bosch et al., 2022), a more detailed explanation of biocueing might help to let patients realize that additional appraisal has to be exerted by themselves. Since some patients reported less adaptive coping behaviors in response to behavioral support messages, discussing and drafting the personalized message should be integrated into therapy and given more attention. Furthermore, some patients might benefit more from integrated relaxation exercises or gamified interventions instead of a text message (Bakker et al., 2016). This all emphasizes the need to integrate new interventions in regular treatment and to tailor these interventions to patient-specific needs (Kip et al., 2018).

Third, we would like to reflect on the findings regarding aggressive behavior. First of all, as patients reported in the qualitative part of the study, it is hard to determine if (hypothetical) aggressive incidents had been prevented and if so, whether that could be associated with the use of the biocueing intervention. Regarding aggression, the pretest-posttest results did indicate a significant decrease in aggression associated with the combination of biocueing intervention and ART. A significant decline in aggressive behavior favoring the biocueing intervention was found in one participant in the SCEDs However, no group effects were found using the repeated ambulatory measurements of the SCEDs, and findings were not supported by therapist-reported aggression level. Several factors are worth noting. First, the EMA responses showed a low prevalence of aggressive behavior in most patients. This created a bottom effect in some patients. Whether this is related to social desirability, lack of concept clarity, or an actual low incidence rate remains unclear. Second, the added value of therapist-reported aggression levels should be considered limited as these scores were not only based on actually observed aggressive incidents but also on patients' reports thereof during the weekly therapy sessions. As some social desirability might have been at play in both therapist- and patient-reported aggression (Barry et al., 2017), patients' reports may have rendered even more accurate information on aggressive behavior in this study given the perceived anonymity of these reports (Lobbestael, 2015). All these factors emphasize the challenges of aggression assessment among forensic outpatients, especially in outpatient settings (Lobbestael, 2015). Furthermore, some patients seemed to mix up anger and aggressive behavior, as if their treatment goal was never to experience anger again. This again stresses the need for integration of the app into therapy and the importance of psycho-education, problematizing aggressive behavior but normalizing feelings of anger. Finally, as the findings of the case studies provide insufficient support for the idea that biocueing interventions might be particularly beneficial for patients with predominant reactive aggression, this topic needs further investigation.

4.1. Strengths and limitations

Several strengths and limitations could be applied to this study. A noteworthy strength is that this study is one of the first in which a biocueing intervention, using psychophysiological measures, is used as an addition to regular treatment in a complex forensic outpatient sample with anger regulation difficulties. As main end-users, the forensic outpatients were involved throughout the developmental process, delivering us with valuable feedback and recommendations for future use of the app. Another strength is the use of a mixed-methods design. Integrating quantitative and qualitative data, on group and individual level, enabled us to explore the effects of the combination of a biocueing intervention and ART and to extract relevant information for further development and implementation in clinical practice.

Our study also had several limitations. First of all, we did not use a control group. Although we did use control and experimental phases in the SCEDs, the lack of a control group prohibits attribution of the found pretest-posttest effects to either the combined intervention or the regular therapy alone. The second limitation is related to compliance. For example, not all patients started with the biocueing intervention and several participants had difficulty answering the EMA questions twice a day. Since we expected these challenges, given the characteristics of forensic outpatient populations, we tried to enhance compliance by sending reminders and contingency management (doubling the amount on the gift card when more than 75% of the EMA questions was answered), as suggested in experience sampling literature (Myin-Germeys and Kuppens, 2021). Despite these efforts, compliance to the intervention and the SCEDs remained low, particularly in the follow-up phase (A_2). Therefore, only the results of the baseline phase (A_1) and intervention phase (B) could be analyzed, for a select group of patients. Fortunately, most patients, including those who prematurely stopped using the app, still participated in the post-measurements of the pretest-posttest design. Although we thereby reduced the risk of bias in the quantitative and qualitative group results, some patients with negative experiences did still just return their devices and reported that they did not want to participate anymore. Another limitation related to the SCEDs is the fact that we were unable to (randomly) assign participants to different lengths of the baseline phase. Since we had already difficulty engaging participants, we had to let go of specific days that would match the research design but would not fit in the schedule of the participants. Since we only found small effects the impact of the missing multiple baseline analysis seems negligible. Fourth, limitations in the usability of the Sense-IT app may have overshadowed the effects of this additive intervention. Connectivity issues, other design preferences, restricted ability to customize the settings, use of a research-owned smartphone, and limited battery life of the smartwatch are disadvantages that are extensively discussed in another study (Ter Harmsel et al., 2022a). For now, we highlight the fact that participants kept reporting that they received too many notifications, even after we adapted the sensitivity, the levels at which notifying vibrations were given, and the activity categories in which notifications were provided. Some participants reported that they received many notifications when they engaged in only minor physical activities. Others reported that they received many notifications in intense physical activities (e.g., sports, intensive work), indicating that these were not recognized by the activity recognition algorithms. Furthermore, displayed activity profiles not always corresponded with their actual activity. In some patients, these notifications increased stress, led to frustration, and may have resulted in early termination of the research project. Important to note is that, outside of a strict research setting, patients would be able to adapt the sensitivity, levels, and activity profiles themselves, as well as to use the app on their own smartphone, which is expected to increase usability. Furthermore, the presentation of activity profiles, as recognized by the smartwatch using Google algorithms, has been modified in a new version of the Sense-IT app. As the number of notifications is directly related to the standard deviation of the baseline measurement, further refinement (e.g., a longer measurement with increased heart rate variety) of the here introduced baseline measurement procedure should be part of future biocueing studies. Moreover, although wrist based PPG measurements generally produce accurate heart rate estimations in various age groups (Chow and Yang, 2020), it would be advisable for future studies to have access to independent validation of available wearables on the basis of standardized validation protocols (e.g., Van Lier et al., 2019). Fifth, both the therapist-reported aggression levels and patients' self-reported aggressive behavior might have been prone to memory and social desirability biases, which stresses the difficulty of assessing aggressive behavior in outpatient settings. Sixth, the use of EMA questions may not only have had an impact on the qualitative measure of interoceptive awareness, but might also have impacted the entire ABA design. More specifically, the fact that forensic outpatients who often have difficulty reflecting on their emotions and behaviors were facilitated by the daily questions to do so, may have increased awareness of emotions, which may have had therapeutic effects as well.

The interpretation of the effects of the biocueing intervention may therefore have been complicated by the use of experienced sampling in this research design. Seventh, although patients and therapists were involved in the entire developmental process of the Sense-IT app, the app was not an integral part of therapy in this study. This limited integration in therapy may have had a negative impact on the results.

4.2. Implications for research and clinical practice

Since it is known that a lot of end-users stop using a mental health app if their goals and preferences are not met (Torous et al., 2019), it is important to create more realistic expectations by providing patients with a more detailed explanation of biocueing as well as to improve the usability of the Sense-IT app. A substantial amount of recommendations have yet already been implemented in a new version of the Sense-IT app. Further refinement of the baseline measurement procedure is an important and necessary step, both to increase usability and to facilitate therapists and patients. Some other usability issues, e.g., the limited battery life of the smartwatches and imperfections in activity recognition, might get solved by technological advances in the future. For future research, it would be relevant to further investigate which patients benefit from a biocueing intervention that is integrated in therapy. Important characteristics to be considered are for instance aggression type, feedback receptivity, and mandatory or voluntary treatment. In addition, it should be assessed when and for how long the intervention should preferably be used. These research directions are in line with the shift toward developing and delivering personalized interventions precisely at moments of need (Bidargaddi et al., 2020). Collaboration between research groups and implementation of multicenter trials are encouraged to increase the sample size. In forensic populations, where demanding traditional research methods are often not feasible, SCEDs should be considered. Given our experiences, we recommended selecting measures that are less sensitive to floor or ceiling effects. When EMA is used, our advice would be to clearly distinguish the research component from the studied intervention. Furthermore, to gain a deeper understanding of the role of interoception in biocueing, we suggest using a combination of measures related to different facets of this concept. Finally, we cannot stress enough the importance of integration of the intervention in therapy. In line with the feedback of the forensic outpatients indicating a need for more personalized use (i.e., on their own smartphone, only in specific circumstances, for longer or shorter periods, with the ability to customize the settings themselves), we encourage therapists and patients to use and evaluate the Sense-IT biocueing app in everyday practice.

4.3. Conclusion

In sum, the qualitative results indicate that the use of a biocueing intervention as an addition to regular ART could be considered a helpful means to increase interoceptive awareness among forensic outpatients. Furthermore, during the combination of this new intervention and regular ART significant decreases in self-reported aggressive behavior were observed. However, results of the repeated ambulatory measurements of the SCEDs do not indicate a clear effect favoring the addition of a biocueing intervention. On group level, no significant effects were found. On the individual level, effects favoring the intervention condition were only found for two participants. Decreasing compliance to the demanding research design, the possible therapeutic effects of the daily EMA questions, and limitations in both usability and integration in therapy, might have impacted the results and hampered interpretability. Future research and development should focus on increasing usability, tailoring the intervention to individual needs, and on integration into therapy. Furthermore, research should further investigate the individual characteristics (i.e., aggression type, feedback receptivity) associated with effective support by the Sense-IT app, as the use of personalized treatment interventions in clinical practice, including new technological interventions, is only expected to increase in the coming years.

Data availability statement

The datasets presented in this article are not readily available because of participants privacy. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by METC, VUmc: NL63911.029.17. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JH, MN, AP, and TP conceived the entire study and developed the study designs. JH and LS performed the qualitative analysis. JH coordinated the data collection and processing, wrote the manuscript, and analyzed the quantitative data in close cooperation with MN, who provided an expert assessment and contributed R code. MN, LS, TP, AG, and AP critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was performed within a larger research project, financially funded by the Dutch Ministry of Justice and Security (no grant code available).

Acknowledgments

We would like to thank the study participants and the therapists at Inforsa, Forensic Mental Healthcare in Amsterdam, Netherlands. Furthermore, we would also like to thank Eline Middelhoven, Lisette Stegeman, Wiesje van Overbeek, Lisa Kisters, Roos Brakeboer, Ashira Oude Lenferink, Vera Heugen, Christine van der Beek, Cylia Hendriks, Melanie Haeck, Joris Hagenbeek, and Margreet Flipse for their assistance in data collection. Finally, we would like to thank the organizations involved in the development of Sense-IT for use in this study project: University of Twente, Scelta/GGNet, Amsterdam University Medical Center, Inforsa/Arkin, and Pluryn.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.983286/full#supplementary-material

References

Adams, Z. W., McClure, E. A., Gray, K. M., Danielson, C. K., Treiber, F. A., and Ruggiero, K. J. (2017). Mobile devices for the remote acquisition of physiological and behavioral biomarkers in psychiatric clinical research. *J. Psychiatr. Res.* 85, 1–14. doi: 10.1016/j.jpsychires.2016.10.019

Armstrong, T., Wells, J., Boisvert, D. L., Lewis, R., Cooke, E. M., Woeckener, M., et al. (2019). Skin conductance, heart rate and aggressive behavior type. *Biol. Psychol.* 141, 44–51. doi: 10.1016/j.biopsycho.2018.12.012

Bakker, D., Kazantzis, N., Rickwood, D., and Rickard, N. (2016). Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Ment. Health* 3:e7. doi: 10.2196/mental.4984

Barry, C. T., Lui, J. H., and Anderson, A. C. (2017). Adolescent narcissism, aggression, and prosocial behavior: the relevance of socially desirable responding. *J. Pers. Assess.* 99, 46–55. doi: 10.1080/00223891.2016.1193812

Bellemans, T., Didden, R., Van Busschbach, J. T., Hoek, P. T., Scheffers, M. W., Lang, R. B., et al. (2019). Psychomotor therapy targeting anger and aggressive behavior in individuals with mild or borderline intellectual disabilities: a systematic review. *J. Intellect. Develop. Disabil.* 44, 121–130. doi: 10.3109/13668250.2017.1326590

Bellemans, T., Didden, R., Visser, R., Schaafsma, D., Totsika, V., and Van Busschbach, J. T. (2018). Psychomotor therapy for anger and aggression in mild intellectual disability or borderline intellectual functioning: an intervention mapping approach. *Body Mov. Dance Psychother.* 13, 1–17. doi: 10.1080/17432979.2018.1471006

Bidargaddi, N., Schrader, G., Klasnja, P., Licinio, J., and Murphy, S. (2020). Designing m-health interventions for precision mental health support. *Transl. Psychiatry* 10, 222–228. doi: 10.1038/s41398-020-00895-2

Blankenstein, N., De Rooij, M., Van Ginkel, J., Wilderjans, T., De Ruigh, E., Oldenhof, H., et al. (2021). Neurobiological correlates of anti-sociality across adolescence and young adulthood: a multi-sample, multi-method study. *Psychol. Med.* 1-16, 1–16. doi: 10.1017/S0033291721003457

Blankenstein, N. E., Vandenbroucke, A. R. E., De Vries, R., Swaab, H., Popma, A., and Jansen, L. M. C. (2022). Understanding aggression in adolescence by studying the neurobiological stress system: a systematic review. *Motiv. Sci.* 8, 133–149. doi: 10.1037/mot0000259

Bolger, N., and Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford Press.

Bosch, R., Chakhssi, F., and Noordzij, M. L. (2022). Acceptance and potential clinical added value of biocueing in forensic psychiatric patients with autism spectrum disorder and/or intellectual disability. *Psychiatry Res.* 313:114645. doi: 10.1016/j.psychres.2022.114645

Branje, S., and Koot, H. M. (2018). "Psychophysiology of aggression," in *Handbook of child and adolescent aggression*. eds. T. Malti and K. H. Rubin (New York: Guilford Press), 84–106.

Brännström, L., Kaunitz, C., Andershed, A.-K., South, S., and Smedslund, G. (2016). Aggression replacement training (ART) for reducing antisocial behavior in adolescents and adults: a systematic review. *Aggress. Violent Behav.* 27, 30–41. doi: 10.1016/j.avb.2016.02.006

Buss, A. H., and Perry, M. (1992). The aggression questionnaire. *J. Pers. Soc. Psychol.* 63:452. doi: 10.1037/0022-3514.63.3.452

Calì, G., Ambrosini, E., Picconi, L., Mehling, W., and Committeri, G. (2015). Investigating the relationship between interoceptive accuracy, interoceptive awareness, and emotional susceptibility. *Front. Psychol.* 6:1202. doi: 10.3389/fpsyg.2015.01202

Centifanti, L. C. M., Kimonis, E. R., Frick, P. J., and Aucoin, K. J. (2013). Emotional reactivity and the association between psychopathy-linked narcissism and aggression in detained adolescent boys. *Dev. Psychopathol.* 25, 473–485. doi: 10.1017/S0954579412001186

Chow, H.-W., and Yang, C.-C. (2020). Accuracy of optical heart rate sensing technology in wearable fitness trackers for young and older adults: validation and comparison study. *JMIR Mhealth Uhealth* 8:e14707. doi: 10.2196/14707

Cornet, L., Mandersloot, M., Pool, R., and De Kogel, C. (2017). De "zelfmetende" justitiabele: Een verkennend onderzoek naar technologische zelfmeetmethoden binnen justitiële context. *Cahiers* 2017:17. Available at: https://repository.wodc.nl/handle/20.500.12832/195

Crozier, J. C., Dodge, K. A., Fontaine, R. G., Lansford, J. E., Bates, J. E., Pettit, G. S., et al. (2008). Social information processing and cardiac predictors of adolescent antisocial behavior. *J. Abnorm. Psychol.* 117, 253–267. doi: 10.1037/0021-843X.117.2.253

De Looff, P. C., Cornet, L. J., De Kogel, C. H., Fernández-Castilla, B., Embregts, P. J., Didden, R., et al. (2021). Heart rate and skin conductance associations with physical aggression, psychopathy, anti-social personality disorder and conduct disorder: an updated meta-analysis. *Neurosci. Biobehav. Rev.* 132, 553–582. doi: 10.1016/j.neubiorev.2021.11.003

De Looff, P. C., Noordzij, M. L., Moerbeek, M., Nijman, H., Didden, R., and Embregts, P. (2019). Changes in heart rate and skin conductance in the 30 min preceding aggressive behavior. *Psychophysiology* 56:e13420. doi: 10.1111/psyp.13420

Declercq, L., Cools, W., Beretvas, S. N., Moeyaert, M., Ferron, J. M., and Van den Noortgate, W. (2020). Multi-SCED: a tool for (meta-) analyzing single-case experimental data with multilevel modeling. *Behav. Res. Methods* 52, 177–192. doi: 10.3758/s13428-019-01216-2

Derks, Y. P., Klaassen, R., Westerhof, G. J., Bohlmeijer, E. T., and Noordzij, M. L. (2019). Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multi-cycle usability testing study. *JMIR Mhealth Uhealth* 7:e13479. doi: 10.2196/13479

Derks, Y. P., Visser, T. D., Bohlmeijer, E. T., and Noordzij, M. L. (2017). mHealth in mental health: how to efficiently and scientifically create an ambulatory biofeedback e-coaching app for patients with borderline personality disorder. *Int. J. Hum. Factors Ergon.* 5, 61–92. doi: 10.2196/13479

Fletcher, R. R., Tam, S., Omojola, O., Redemske, R., and Kwan, J. (2011). "Wearable sensor platform and mobile application for use in cognitive behavioral therapy for drug addiction and PTSD," in *Annual international conference of the IEEE engineering n medicine and biology society, 2011*. (Boston, MA: IEEE), 1802–1805.

Forkmann, T., Scherer, A., Meessen, J., Michal, M., Schächinger, H., Vögele, C., et al. (2016). Making sense of what you sense: disentangling interoceptive awareness, sensibility and accuracy. *Int. J. Psychophysiol.* 109, 71–80. doi: 10.1016/j.ijpsycho.2016.09.019

Füstös, J., Gramann, K., Herbert, B. M., and Pollatos, O. (2013). On the embodiment of emotion regulation: interoceptive awareness facilitates reappraisal. *Soc. Cogn. Affect. Neurosci.* 8, 911–917. doi: 10.1093/scan/nss089

Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K., and Critchley, H. D. (2015). Knowing your own heart: distinguishing interoceptive accuracy from interoceptive awareness. *Biol. Psychol.* 104, 65–74. doi: 10.1016/j.biopsycho.2014.11.004

Goldstein, A. P., Glick, B., and Gibbs, J. C. (1998). *Aggression replacement training: A comprehensive intervention for aggressive youth*. (*Rev. ed.*). Champaign, IL: Research Press.

Goodwin, M. S., Mazefsky, C. A., Ioannidis, S., Erdogmus, D., and Siegel, M. (2019). Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Res.* 12, 1286–1296. doi: 10.1002/aur.2151

Gratz, K. L., and Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J. Psychopathol. Behav. Assess.* 26, 41–54. doi: 10.1007/s10862-008-9102-4

Gross, J. J., and Jazaieri, H. (2014). Emotion, emotion regulation, and psychopathology: an affective science perspective. *Clin. Psychol. Sci.* 2, 387–401. doi: 10.1177/2167702614536164

Henwood, K. S., Chou, S., and Browne, K. D. (2015). A systematic review and meta-analysis on the effectiveness of CBT informed anger management. *Aggress. Violent Behav.* 25, 280–292. doi: 10.1016/j.avb.2015.09.011

Kazantzis, N., Whittington, C., Zelencich, L., Kyrios, M., Norton, P. J., and Hofmann, S. G. (2016). Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. *Behav. Ther.* 47, 755–772. doi: 10.1016/j.beth.2016.05.002

Kazdin, A. E. (2019). Single-case experimental designs. Evaluating interventions in research and clinical practice. *Behav. Res. Ther.* 117, 3–17. doi: 10.1016/j.brat.2018.11.015

Khoury, N. M., Lutz, J., and Schuman-Olivier, Z. (2018). Interoception in psychiatric disorders: a review of randomized controlled trials with interoception-based interventions. *Harv. Rev. Psychiatry* 26, 250–263. doi: 10.1097/HRP.0000000000000170

Kip, H., Bouman, Y. H., Kelders, S. M., and Van Gemert-Pijnen, L. J. (2018). eHealth in treatment of offenders in forensic mental health: a review of the current state. *Front. Psych.* 9:42. doi: 10.3389/fpsyt.2018.00042

Klein Tuente, S., Bogaerts, S., Bulten, E., Keulen-de Vos, M., Vos, M., Bokern, H., et al. (2020). Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *J. Clin. Med.* 9:2258. doi: 10.3390/jcm9072258

Knoedler, D. W. (1989). The modified overt aggression scale. *Am. J. Psychiatry* 146, 1081b–1082b. doi: 10.1176/ajp.146.8.1081b

Lane, R. D., and Schwartz, G. E. (1987). Levels of emotional awareness: a cognitive-developmental theory and its application to psychopathology. *Am. J. Psychiatry* 144, 133–143. doi: 10.1176/ajp.144.2.133

Lehrer, P. (2013). How does heart variability biofeedback work? Resonance, the baroreflex, and other mechanism. *Biofeedback* 41, 26–31. doi: 10.5298/1081-5937-41.1.02

Lentferink, A., Noordzij, M. L., Burgler, A., Klaassen, R., Derks, Y., Oldenhuis, H., et al. (2021). On the receptivity of employees to just-in-time self-tracking and eCoaching for stress management: a mixed-methods approach. *Behav. Inf. Technol.* 41, 1398–1424. doi: 10.1080/0144929X.2021.1876764

Lobbestael, J. (2015). Challenges in aggression assessment: the gap between self-report and behavior, and a call for new valid behavioral paradigms. *J. Socialomics* 05, 5–6. doi: 10.4172/2167-0358.1000141

Lorber, M. F. (2004). Psychophysiology of aggression, psychopathy, and conduct problems: a meta-analysis. *Psychol. Bull.* 130, 531–552. doi: 10.1037/0033-2909.130.4.531

Mackintosh, M. A., Niehaus, J., Taft, C. T., Marx, B. P., Grubbs, K., and Morland, L. A. (2017). Using a mobile application in the treatment of dysregulated anger among veterans. *Mil. Med.* 182, e1941–e1949. doi: 10.7205/MILMED-D-17-00063

McCraty, R., and Zayas, M. (2014). Cardiac coherence, self-regulation, autonomic stability, and psychosocial well-being. *Front. Psychol.* 5:1090. doi: 10.3389/fpsyg.2014.01090

McDonnell, A., McCreadie, M., Mills, R., Deveau, R., Anker, R., and Hayden, J. (2015). The role of physiological arousal in the management of challenging behaviors in individuals with autistic spectrum disorders. *Res. Dev. Disabil.* 36, 311–322. doi: 10.1016/j.ridd.2014.09.012

McIntosh, L. G., Janes, S., O'Rourke, S., and Thomson, L. D. (2021). Effectiveness of psychological and psychosocial interventions for forensic mental health inpatients: a meta-analysis. *Aggress. Violent Behav.* 58:101551. doi: 10.1016/j.avb.2021.101551

Moore, C. C., Hubbard, J., Morrow, M. T., Barhight, L. R., Lines, M. M., Sallee, M., et al. (2018). The simultaneous assessment of and relations between children's sympathetic and parasympathetic psychophysiology and their reactive and proactive aggression. *Aggress. Behav.* 44, 614–623. doi: 10.1002/ab.21786

Murray-Close, D., Holterman, L. A., Breslend, N. L., and Sullivan, A. (2017). Psychophysiology of proactive and reactive relational aggression. *Biol. Psychol.* 130, 77–85. doi: 10.1016/j.biopsycho.2017.10.005

Myin-Germeys, I., and Kuppens, P. (2021). The open handbook of experience sampling methodology: a step-by-step guide to designing, conducting, and analyzing ESM studies. *Leuven.* 33–101.

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., et al. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* 52, 446–462. doi: 10.1007/s12160-016-9830-8

Nuske, H. J., Finkel, E., Hedley, D., Parma, V., Tomczuk, L., Pellecchia, M., et al. (2019). Heart rate increase predicts challenging behavior episodes in preschoolers with autism. *Stress* 22, 303–311. doi: 10.1080/10253890.2019.1572744

Ortiz, J., and Raine, A. (2004). Heart rate level and antisocial behavior in children and adolescents: a meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* 43, 154–162. doi: 10.1097/00004583-200402000-00010

Palix, J., Akselrod, M., Cungi, C., Giuliani, F., and Favrod, J. (2017). Changes in heart rate variability recorded in natural situation with T-shirt integrated sensors and level of observed behavioral excitation: a pilot study of patients with intellectual disabilities and psychiatric disorders. *Front. Psych.* 8:4. doi: 10.3389/fpsyt.2017.00004

Parker, R. I., Vannest, K. J., and Brown, L. (2009). The improvement rate difference for single-case research. *Except. Child.* 75, 135–150. doi: 10.1177/001440290907500201

Patrick, C. J. (2014). Physiological correlates of psychopathy, antisocial personality disorder, habitual aggression, and violence. *Curr. Top. Behav. Neurosci.* 21, 197–227. doi: 10.1007/7854_2014_345

Pinheiro, J. C., and Bates, D. M. (2000). "Linear mixed-effects models: basic concepts and examples," in *Mixed-effects models in S and S-plus*. eds. J. C. Pinheiro and D. M. Bates (New York: Springer), 3–56. doi: 10.1007/978-1-4419-0318-1_1

Portnoy, J., and Farrington, D. P. (2015). Resting heart rate and antisocial behavior: an updated systematic review and meta-analysis. *Aggress. Violent Behav.* 22, 33–45. doi: 10.1016/j.avb.2015.02.004

Price, C. J., and Hooven, C. (2018). Interoceptive awareness skills for emotion regulation: theory and approach of mindful awareness in body-oriented therapy (MABT). *Front. Psychol.* 9:798. doi: 10.3389/fpsyg.2018.00798

Puhalla, A. A., and McCloskey, M. S. (2020). The relationship between physiological reactivity to provocation and emotion dysregulation with proactive and reactive aggression. *Biol. Psychol.* 155:107931. doi: 10.1016/j.biopsycho.2020.107931

Pustejovsky, J., Chen, M., and Swan, D. (2021). *Single-case effect size calculator* (Version 0.5.2). Avalable at: https://jepusto.shinyapps.io/SCD-effect-sizes/ (Accessed June 28, 2022).

R Core Team (2017). *R: A language and environment for statistical computing*. Available at: http://www.R-project.org/ (Accessed June 28, 2022).

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., et al. (2006). The reactive–proactive aggression questionnaire: differential correlates of reactive and proactive aggression in adolescent boys. *Aggress. Behav.* 32, 159–171. doi: 10.1002/ab.20115

Riley, W. T., Serrano, K. J., Nilsen, W., and Atienza, A. A. (2015). Mobile and wireless technologies in health behavior and the potential for intensively adaptive interventions. *Curr. Opin. Psychol.* 5, 67–71. doi: 10.1016/j.copsyc.2015.03.024

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., et al. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychol. Bull.* 144, 343–393. doi: 10.1037/bul0000128

Smeijers, D., and Koole, S. L. (2019). Testing the effects of a virtual reality game for aggressive impulse management (VR-GAIME): study protocol. *Front. Psych.* 10:83. doi: 10.3389/fpsyt.2019.00083

Spielberger, C. D., Sydeman, S. J., Owen, A. E., and Marsh, B. J. (1999). *Measuring anxiety and anger with the state-trait anxiety inventory (STAI) and the state-trait anger expression inventory (STAXI)*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Ter Harmsel, J. F., Noordzij, M. L., Goudriaan, A. E., Dekker, J. J. M., Swinkels, L. T. A., Van der Pol, T. M., et al. (2021). Biocueing and ambulatory biofeedback to enhance emotion regulation: a review of studies investigating non-psychiatric and psychiatric populations. *Int. J. Psychophysiol.* 159, 94–106. doi: 10.1016/j.ijpsycho.2020.11.009

Ter Harmsel, J. F., Smulders, L. M., Noordzij, M. L., Swinkels, L. T. A., Goudriaan, A. E., Popma, A., et al. (2022a). Forensic psychiatric outpatients' and therapists' perspectives on a wearable biocueing app (Sense-IT) as an addition to aggression regulation therapy: qualitative focus group and interview study. *JMIR Form. Res.* 7:e40237. doi: 10.2196/40237

Ter Harmsel, A., Van der Pol, T., Swinkels, L., Goudriaan, A. E., Popma, A., and Noordzij, M. L. (2021). Development of a wearable biocueing app (Sense-IT) among forensic psychiatric outpatients with aggressive behavior: design and evaluation study. *JMIR Form. Res.* 5:e29267. doi: 10.2196/29267

Ter Harmsel, J. F., Van Dongen, J. D. M., Zijlmans, J., Van der Pol, T. M., Marhe, R., and Popma, A. (2022b). Neurobiological responses towards stimuli depicting aggressive interactions in delinquent young adults and controls: no relation to reactive and proactive aggression. *Brain Sci.* 12:124. doi: 10.3390/brainsci12020124

Thomson, N. D., Kevorkian, S., Blair, J., Farrell, A., West, S. J., and Bjork, J. M. (2021). Psychophysiological underpinnings of proactive and reactive aggression in young men and women. *Physiol. Behav.* 242:113601. doi: 10.1016/j.physbeh.2021.113601

Torous, J., Andersson, G., Bertagnoli, A., Christensen, H., Cuijpers, P., Firth, J., et al. (2019). Towards a consensus around standards for smartphone apps and digital mental health. *World Psychiatry* 18, 97–98. doi: 10.1002/wps.20592

Umetani, K., Singer, D. H., McCraty, R., and Atkinson, M. (1998). Twenty-four hour time domain heart rate variability and heart rate: relations to age and gender over nine decades. *J. Am. Coll. Cardiol.* 31, 593–601. doi: 10.1016/S0735-1097(97)00554-8

Van Baardewijk, Y., Andershed, H., Stegge, H., Nilsson, K. W., Scholte, E., and Vermeiren, R. (2010). Development and tests of short versions of the youth psychopathic traits inventory and the youth psychopathic traits inventory-child version. *Eur. J. Psychol. Assess.* 26, 122–128. doi: 10.1027/1015-5759/a000017

Van Goozen, S. H., Fairchild, G., Snoek, H., and Harold, G. T. (2007). The evidence for a neurobiological model of childhood antisocial behavior. *Psychol. Bull.* 133, 149–182. doi: 10.1037/0033-2909.133.1.149

Van Lier, H. G., Pieterse, M. E., Garde, A., Postel, M. G., De Haan, H. A., Vollenbroek-Hutten, M. M. R., et al. (2019). A standardized validity assessment protocol for physiological signals from wearable technology: methodological underpinnings and an application to the E4 biosensor. *Behav. Res. Methods* 52, 607–629. doi: 10.3758/s13428-019-01263-9

Wagner, C. R., and Abaied, J. L. (2015). Relational victimization and proactive versus reactive relational aggression: the moderating effects of respiratory sinus arrhythmia and skin conductance. *Aggress. Behav.* 41, 566–579. doi: 10.1002/ab.21596

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Wing, J., Beevor, A., Curtis, R., Park, S., Hadden, J., and Burns, A. (1998). Health of the nation outcome scales (HoNOS). *Br. J. Psychiatry* 172, 11–18. doi: 10.1192/bjp.172.1.11

Zijlmans, J., Marhe, R., van Duin, L., Luijks, M.-J., Bevaart, F., and Popma, A. (2021). No association between autonomic functioning and psychopathy and aggression in multi-problem young adults. *Front. Psychol.* 12:645089. doi: 10.3389/fpsyg.2021.645089

Zwets, A. J., Hornsveld, R., Kraaimaat, F. W., Kanters, T., Muris, P., and Van Marle, H. (2014). The psychometric properties of the anger bodily sensations questionnaire (ABSQ). *J. Forens. Psychiatry Psychol.* 25, 432–450. doi: 10.1080/14789949.2014.925957

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 16 May 2024
DOI 10.3389/fpsyt.2024.1307633

OPEN ACCESS

EDITED BY
Birgit Angela Völlm,
University of Rostock, Germany

REVIEWED BY
Filippo Aschieri,
Catholic University of the Sacred Heart, Italy
Carlo Garofalo,
University of Perugia, Italy

*CORRESPONDENCE
Fernando Renee González Moraga
fernando_renee.gonzalez_moraga@med.lu.se

RECEIVED 06 October 2023
ACCEPTED 15 April 2024
PUBLISHED 16 May 2024

CITATION
González Moraga FR, Enebrink P, Perrin S, Sygel K, Veling W and Wallinius M (2024) VR-assisted aggression treatment in forensic psychiatry: a qualitative study in patients with severe mental disorders.
Front. Psychiatry 15:1307633.
doi: 10.3389/fpsyt.2024.1307633

VR-assisted aggression treatment in forensic psychiatry: a qualitative study in patients with severe mental disorders

Fernando Renee González Moraga[1,2,3]*, Pia Enebrink[4], Sean Perrin[5], Kristina Sygel[3,6], Wim Veling[7] and Märta Wallinius[1,2,3]

[1]Evidence-based Forensic Psychiatry, Department of Clinical Sciences Lund, Psychiatry, Lund University, Lund, Sweden, [2]Research Department, Regional Forensic Psychiatric Clinic, Växjö, Sweden, [3]Centre for Ethics, Law and Mental Health, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden, [4]Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden, [5]Department of Psychology, Lund University, Lund, Sweden, [6]Department of Forensic Psychiatry, National Board of Forensic Medicine, Stockholm, Sweden, [7]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Introduction: Improvements in virtual reality (VR) have made it possible to create realistic, virtual settings for behavioral assessment and skills training that cannot otherwise be accessed in a safe way in forensic psychiatric settings. VR interventions are under development but little is known how forensic psychiatric patients with severe mental disorders experience VR-assisted assessments or treatments.

Methods: The present study aimed to help fill this knowledge gap via qualitative interviews with seven patients with severe mental disorders at a high-security forensic psychiatric clinic who had completed the newly revised Virtual Reality Aggression Prevention Training (VRAPT). All participants were interviewed 12 weeks after the VRAPT intervention, and interview data analyzed with manifest inductive content analysis.

Results: Six manifest content categories were identified: 1. Therapeutic process, 2. VRAPT method, 3. VR technology, 4. Previous treatment experiences, 5. Challenges to treatment of aggression, and 6. Unexpected experiences. The participants had diverse experiences related to both the VRAPT intervention and forensic psychiatric care. Participants described a mixture of positive experiences in relation to VR-assisted role-plays, and less positive in relation to motivation for aggression-focused treatment and technological limitations.

Discussion: The present findings suggest further studies are needed on how to best implement VR-assisted treatments for aggression in forensic settings, and potentially further modification of treatment content in interventions like VRAPT.

KEYWORDS

virtual reality, VR, aggression, forensic psychiatry, VRAPT, experiences, treatment, content analysis

1 Introduction

Forensic psychiatric patients have been described as a heterogeneous, challenging, and vulnerable group in both society and clinical settings (1–3); their behaviors and clinical status can be traced to a complex constellation of severe mental disorders, antisocial lifestyle, substance use, and a high degree of impulsivity, and/or lack of empathy (4–6). Patient aggression is a known problem in forensic psychiatry and is considered central to patient management (7, 8), yet there is a scarcity of evidence-based treatments for aggression. There is a strong need for more research to develop and evaluate aggression treatments with all the challenges this implies (9–14). Recently, immersive virtual reality (VR) has been identified as a potential tool in assessment and treatment interventions in forensic settings (14–16), also with possibilities for treatment of aggression (9).

Three core concepts in VR have been described: immersion, presence and embodiment (17–20). VR is an immersive technology in that immersion can be regarded as a quality of the system; the system presents a vivid virtual environment while shutting out physical reality (21). Thus, VR technologies can be more or less immersive depending on the quality of the system used (22). Presence, on the other hand, is a perceptual illusion of *being there*; this is not only a cognitive illusion since your body reacts to the illusion even though you are aware that you are not there – your brain cannot overcome the illusion (20). When VR technology works as intended, the user accepts, interacts and is physically, socially and emotionally engaged in the virtual world (23, 24). Embodiment, or sense of embodiment in VR, is how we incorporate and experience our body in a virtual body in VR and is an important factor to the sense of being in a virtual environment (25). Three components have been associated with embodiment: sense of self-location, sense of agency, and sense of body ownership (26).

In forensic settings, VR holds the potential to increase motivation among patients, bridge the gap between real life, therapeutic and laboratory experiences, and increase ecological validity of psychological research (27–31). Understanding the individual's motivation for the aggression (32) and the function of their aggression [e.g., social recognition, emotion regulation, communication, protection; (33)] are important components in comprehending why aggression has emerged and how to treat it, and thus these concepts play an important but understudied role in the development of VR treatments for aggression. Different ways that motivation and function of aggression have been assessed is from first perspectives among youths [see Angry Aggression Scale, AAS; (34)], based on the quadripartite violence typology [QVT: (35), psychopathy (36)], and through staffs' perspectives of forensic patients' aggression (see Assessment and Classification of Function, ACF; 37). An integrated review (38) found that psychiatric inpatient's perceived aggression and violence were associated with a lack of respect by staff towards patients, coercive or controlling interventions in their care, poor staff interpersonal skills, boredom/lack of structured activity, lack of privacy/personal space in the care environment, medication issues and patients' personal characteristics.

Today, two methods employing VR in treatment of aggression in forensic settings are available: VR Game for Aggressive Impulse Management [VR-GAIME; (39)] and Virtual Reality Aggression Prevention Training [VRAPT; (40)]. Both methods were primarily developed for forensic psychiatric patients. In the first randomized controlled trial of VR-GAIME with 30 forensic psychiatric outpatients, VR-GAIME was not more successful in reducing anger and aggressive behavior relative to the control condition, but demonstrated ways in which future work may realize the unfulfilled potential of combining serious gaming and VR in creating effective aggression management interventions (41). The first randomized controlled trial on VRAPT with 128 forensic psychiatric patients with aggression problems, found positive effects at post-treatment as indexed by self-reported hostility, anger control and non-planning impulsiveness, but these effects were not maintained at follow-up (42). While VR-assisted aggression interventions applicable to forensic settings hold the promise of safer and more ecologically valid treatments, the current evidence base is insufficient to draw conclusions about their acceptability and efficacy (42–44), underlining the urgency for more research in this area (9, 31).

Patients' experiences of their treatments are important to our understanding of the responsiveness of these treatments (to the patient) which will in turn influence motivation, treatment engagement and efficacy (14, 45–48), providing directions for further treatment developments. Recently, the VRAPT intervention was revised (9) and pilot studies carried out in forensic psychiatric and prison settings (49). In its revised version [see (9)], VRAPT aims to increase the participant's understanding and management of his/her dysfunctional and reactive aggressive behaviors through a CBT approach theoretically underpinned by the General Aggression Model GAM (50–52). During the initial assessment phase (see Figure 1 for an overview of VRAPT treatment phases), a historical and functional analysis of the participant's dysfunctional aggressive behavior is made in collaboration between participant and therapist. This is then used as a guide for all continued skills training during the skills training phase. For example, a participant may practice situations that are difficult to cope with on a daily basis and that previously have caused the participant to react aggressively and/or even commit a crime because he/she has not known how to manage the situation without resorting to aggression. For instance, the therapist can simulate a stressful situation for the participant through information obtained from the participant regarding previous experiences. Examples of environments that can be simulated with the software Social Worlds ©CleVR are a park, a bus, a café, a jail, a meeting room, and a home environment. In the virtual environment, the therapist can tailor the simulated scenario in real time, e.g. facial expressions, voices, verbal communication and movements of the avatars, while simultaneously seeing what the patient is experiencing in the environment.

The current study constitutes a qualitative investigation of forensic psychiatric patients' experiences of undergoing the revised VRAPT treatment with the following research question: How do patients experience VRAPT and its consequences, opportunities and challenges in a high-security forensic psychiatric clinic?

2 Materials and methods

This is a qualitative study applying inductive content analysis (53, 54) on data from interviews with forensic psychiatric patients who were interviewed 12 weeks after completion of the VRAPT intervention, at the time of a quantitative follow-up of the treatment, between February and October 2021. The colored portion of Figure 1 below gives an overview of the structure of VRAPT [see also Table 1 in (9) for further details of the treatment], note that the proportion of sessions for each phase is not equal, with most sessions considering skills training.

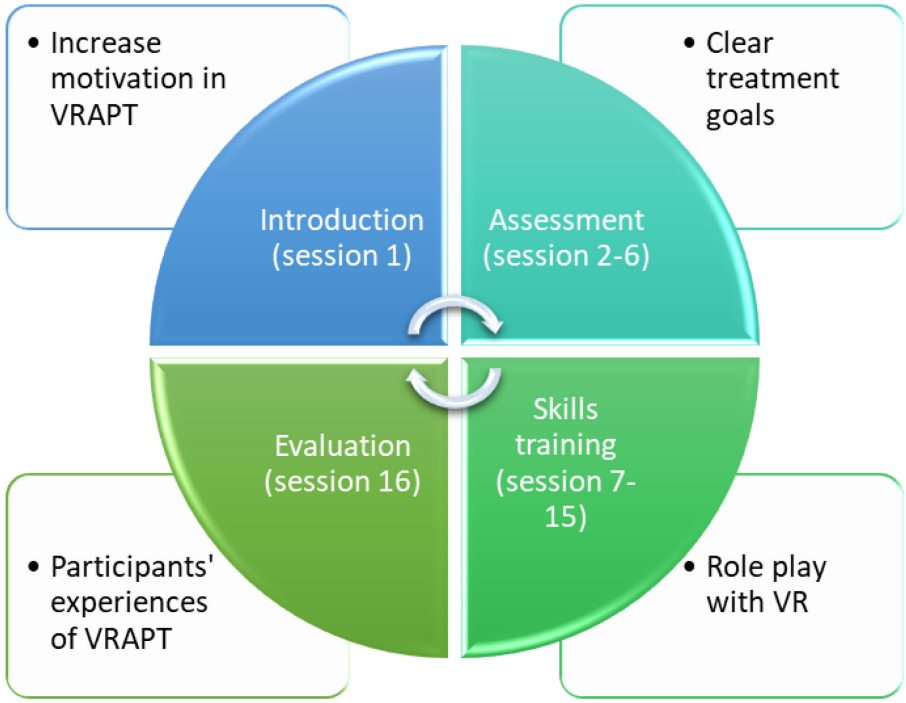


FIGURE 1
Structure of Virtual Reality Aggression Prevention Treatment (VRAPT).

TABLE 1 Content categories and subcategories of forensic psychiatric patients' experiences from the VRAPT intervention.

Category	1. Therapeutic process	2. VRAPT method	3. VR technology	4. Previous treatment experiences	5. Challenges to treatment of aggression	6. Unexpected experiences
Subcategory	1.1 Therapeutic alliance	2.1 Treatment outcomes	3.1 Novelty	4.1 Previous experiences from aggression treatment	5.1 Manipulation	6.1 Covid-19
	1.2 Treatment goals	2.2 Homework experiences	3.2 Immersion	4.2 Patients' experiences of forensic psychiatric care	5.2 Positive experiences from aggression	
	1.3 Matching of treatment to patient's needs	2.3 Staff involvement	3.3 Physiological side effects			
	1.4 Skills training in treatment	2.4 Deviations from VRAPT methodology	3.4 Limitations of technology on therapeutic work			
	1.5 Remaining treatment needs	2.5 Suggestions for improvement	3.5 Potential of VR in forensic settings			

2.1 Ethical considerations

This study was approved by the Swedish Ethical Review Authority (Dnr: 2019-02337; 2020-06317). Eligibility for participation was first assessed by the patients' treating psychiatrist, taking into account their need for aggression treatment, capacity to make an informed decision on participation (e.g., not being under the influence of acute psychosis or intellectual disability), and the possibilities to participate safely in relation to imminent risk of severe violence. Written informed consent was obtained from all participants after providing them with a verbal and written explanation of the study's purpose and procedures. Participation was voluntary, and participants were told that they could withdraw from the study at any time without having to explain why they decided to do so and without impacting any other aspects of their ongoing care. Participants received a ticket for 99 SEK (~€9) to use at the clinic kiosk after the study was completed. All data were pseudonymized before analysis and recordings deleted after the analyses were finished.

Taking into account the persuasiveness of VR technology (55) and the intrinsically coercive nature of forensic settings (56), some studies suggest that ethical priorities must account for the specific vulnerabilities (e.g., autonomy, deception, informed consent, mental liberty, moral agency, dignity) of (forensic) psychiatric patients when using new therapeutic technologies with this population (56–58). For example, therapeutic misconception (9, 59) refers to the fact that a participant may not fully differentiate between participation in clinical research and an ordinary treatment. In this study, we tried to minimize the risk for therapeutic misconception through thorough information to the participants regarding the distinction between the study and their ordinary care. However, there were occasions when participants expressed wishes that the results from their participation in the VRAPT research could be reported to the psychiatrist responsible for their treatment, as a means to affect their care process. All such requests were denied, with reference to the information they received prior to participation.

2.2 Participants

Participants were seven forensic psychiatric patients (6 males, 1 female; mean age = 36 years old; range = 22 – 46 years old), recruited from a high-security forensic psychiatric clinic in Sweden. These seven participants constituted all forensic psychiatric patients who completed the full VRAPT treatment within a quantitative pilot study. All participants had been assessed with a severe mental disorder during a forensic psychiatric investigation before being sentenced to forensic psychiatric care. Inclusion criteria were: 1) ongoing treatment under the law of forensic psychiatric care, 2) history of aggression and current problems with reactive aggression, and 3) having undergone VRAPT treatment. Exclusion criteria were: 1) inability to speak and read Swedish, 2) epilepsy, 3) intellectual disability (IQ < 70), 4) and/or severe autism spectrum disorder, 5) acute psychotic state, and 6) security risks that prohibit participation. The current psychiatric diagnoses (collected from participants' medical files) at the time of participation were schizophrenia spectrum disorders and antisocial personality disorder (both n = 4; 57%), substance use disorders (n = 3; 43%), personality disorder – trait specified (n = 2; 29%) and borderline intellectual functioning, autism, bipolar disorder and paraphilia (all n = 1; 14%).

This study was conducted as part of a larger study, with an accompanying quantitative VRAPT evaluation. All participants (N = 7) who completed the VRAPT intervention as part of the quantitative evaluation, were invited to participation in this study, of which all accepted (100% participation rate). In qualitative research, data saturation must be considered, with suggested guidelines of 9-12 interviews for data saturation (60). In this study, with a narrow focus and a specific group, we had a total sample on N = 7. However, given the data available, our consensus was that data saturation was achieved with the current sample size.

2.3 Procedure

All interviews were conducted by the main author (FGM), a licensed clinical psychologist with training in qualitative methodology and experience working in high-security forensic psychiatry. The interviews were semi-structured and based on an interview guide developed by two of the authors (FGM, MW) for this study. The interview contained open-ended questions such as "How did VRAPT contribute to your own management of aggression?", "How did you experience using VRAPT as a treatment intervention", with the opportunity for follow-up questions based on the participant's previous answers. All interviews were audio-recorded and transcribed for analysis purposes. Interviews were conducted in the clinic on a one-to-one basis except for one case where a staff member was present during the interview due to security reasons. The interviews varied in length from 13 to 34 minutes.

2.4 Analyses

This study applies an inductive approach to qualitative manifest content analysis (53) to examine patients' experiences of the VRAPT intervention. In particular, we considered recommendations from content analysis for data organization and analysis (61–63). After transcription, the interviews and transcriptions were listened to and read several times to correct the transcription and facilitate the overall comprehension of the data. Thereafter, preliminary codes were assigned. After codes were assigned, data were ordered in content categories and subcategories. A content category describes the content on a manifest level, with a low degree of interpretation and a varying degree of abstraction (63). FGM performed the initial coding, then FGM and MW independently identified content categories and subcategories. Final categories were refined and determined in consensus between FGM and MW. Thereafter, categories were

interpreted and a narrative summary of the main findings determined. No software was used in the analysis process.

3 Results

During the interviews, the following content categories emerged: 1. Therapeutic process, 2. VRAPT method, 3. VR technology, 4. Previous treatment experiences, 5. Challenges to treatment of aggression, and 6. Unexpected experiences. Table 1 provides an overview of content categories and subcategories of forensic psychiatric patients' experiences from the VRAPT intervention. Each of the categories and subcategories are described below.

3.1 Therapeutic process

This content category included participants' feelings and thoughts associated with the therapeutic process during the VRAPT intervention. There was a wide range of participants' experiences, with some comments being very specific to the participant's own experiences from the VRAPT intervention, and some comments about the general treatment experience, both VRAPT and the treatment milieu on the care unit. The following subcategories were identified: Therapeutic alliance, Treatment goals, Matching of treatment to patient's needs, Skills training in treatment, and Remaining treatment needs.

3.1.1 Therapeutic alliance

During interviews, participants described how they experienced the treatment relationship to the VRAPT therapist. Several participants perceived the relationship with their therapists in a positive way, highlighting the collaboration and emphasizing the importance of feeling safe during the session. Preparing the participant for what was to come during the session seemed to facilitate a feeling of safety for the patient.

> *"My therapist had, I think, a genuine sense of developing the patient in the flaws that they possess." (P4)*
>
> *I thought we had a good collaboration, he was clear and told me each time what was going to happen and so on. So he was good, I think … I mean, he told me exactly what was going to happen, so I was clear beforehand, I could feel safe. Then you never know how you will react." (P7)*

Another participant described the treatment relationship with the therapist as difficult and that feelings of sadness and stigmatization were triggered.

> *"(My therapist) tried to oppress me, I felt sad and bad but still thought oh I'll manage this and did it … I felt humiliated and sad, didn't want to talk to anyone, just wanted to be left alone. Why did (my therapist) do that? I am also a human, I have made a mistake and am here to pay for it, doing all the programs, treatments, attending meetings and there is nothing more. I am also human, also have feelings, think about that. Just like everyone else." (P5)*

3.1.2 Treatment goals

During the first VRAPT sessions, the participant and therapist formulate treatment goals. In some interviews, the participants had ambiguous, or negative, experiences concerning treatment goals.

> *"I cannot really answer that because no one said that we would have a goal, we just started there and so … I would probably say that it's about remembering them and trying to move forward. Putting energy into thinking about things, implementing and feeling that you know what direction to take … Trying to deal with difficult situations in a new way…" (P2)*
>
> *"If we had a goal? No we didn't have a goal." (P3)*

For other participants, treatment goals were described as an important part of the treatment, incorporating previous experiences and giving the participant directions for the continued development.

> *"My goals were to be able to de-escalate challenging situations and currently I don't remember quite how it was formulated, and I practiced that a lot, based on previous experiences I had and got during the time in VRAPT, you went back to those experiences and implemented in the goals so it would go together with the goals you had … The goals kind of gave directions you should take, and it was just to follow the road." (P4)*

3.1.3 Matching of treatment to patient's needs

A crucial component in all therapeutic interventions is matching treatment to the participant's needs. VRAPT is a new intervention, not familiar to participants or therapists in general. During interviews, descriptions by participants that VRAPT was not what they initially expected, or that the intervention did not match their own perception of treatment needs, were discerned.

> *"… I began to realize more and more that this thing with the VR study was about me doing a treatment to become less aggressive, not that I was just going to go through these things and the method. We then had some rather long conversations, several times in a row, where we concluded that I am not interested in being treated for aggression. It demanded more than I can handle." (P1)*

One participant referred to previous treatment experiences and how VRAPT failed to provide new strategies to manage difficult situations.

> *"Then it was this that you did not learn so many strategies, it was mostly breathing and such, was mostly that you went through the situation but there was not much you learned on how to handle the situation but they came during VRAPT and then you acted according to the situation but there was not so much that I learned new strategies. This about breathing and things, I have already learned that from (my previous therapist)." (P2)*

3.1.4 Skills training in treatment

The overall aim of VRAPT is to treat aggression through increased awareness and skills training. Skills training is performed in VR-assisted role-plays with participants, intended to increase participants' awareness of their own reactions, and management of reactions, during conflict resolution. This is important for treatment outcomes since participants' past dysregulated reactions have been maladaptive, generating problems for themselves and others. These skills training sessions may be challenging to participants and yet are a crucial part of the intervention. Initial unease related to skills training could, for some participants, be replaced with a sense of understanding, as continued skills training could provide insights into their own strengths and weaknesses, specifically in social situations.

> *"Very difficult, you put yourself in situations where you are challenged … Yes, it is still a bit difficult. But it will always be, to be honest. Yet it is fun at the same time because you are stimulated to be a stronger person even though it is difficult…" (P4)*

> *"In the beginning it was a bit unclear, then afterwards it just became understandable that you can use this to give rise to certain feelings or thoughts … you get used to it … It changes, it's just like with everything else in the beginning when you do things for the first time it usually feels a bit different but the more you practice, put time into it the easier it feels … The more you got into the VRAPT and the project, you got into how you are as a person. What your weaknesses are and what your strengths are in a social context…" (P4)*

> *"I thought it was okay, but it can be a bit difficult when the person I'm arguing with is arguing against me, when there's not much to say. There were certain situations where I tried to reason with the person, but it didn't go well…" (P5)*

3.1.5 Remaining treatment needs

When asked about remaining treatment needs after the completion of VRAPT, participants specifically mentioned different kinds of skills training, especially concerning communication and managing difficult situations in which you feel challenged or experience severe anger. Some participants had experienced difficulties with emotion regulation post VRAPT treatment and emphasized how they wished they had been able to access their learned skills to avoid ending up in unfavorable circumstances.

> *"I would like to work more on situations where you feel challenged as a person." (P4)*

> *"It would have been to communicate even more … Yes, to become even better at it and to stop in time. So I don't get so damn angry … I should have thought more about what the therapist said so I wouldn't have ended up in this situation I'm in now. I got so damn angry again. I had forgotten everything I had learned and thought it was tragic as hell." (P7)*

3.2 VRAPT method

Part of the interviews concerned the specific VRAPT intervention, with the following subcategories: Treatment outcomes, Suggestions for improvement, Homework experiences, Staff involvement, and Deviations from VRAPT methodology.

3.2.1 Treatment outcomes

The subjectively experienced outcomes from the treatment varied significantly between participants. For some, participation in the VRAPT intervention did not yield specific effects regarding skills training in management of aggression due to absence of aggressive impulses to manage during the sessions. Yet, an ambivalence in relation to perceived outcomes was clear in the data, where addressing aggression and learning new strategies for interpersonal interactions were described as beneficial by participants.

> *"Well, the thing is that I didn't learn anything about how to reduce aggression because I didn't get any. So the technique failed quite early on … But there was a result because we talked about aggression…" (P1)*

> *"There were both situations where I could recognize myself and those that were new, but I learned the new ones pretty well … I don't know, I didn't think it was very useful … Yes, the only thing I thought changed was that I used the strategy of being diplomatic instead of going out and punching someone … It did help, I noticed that when I tried to be calmer, the situation became easier than if you started arguing. It may help if there's a situation on the ward where a patient is arguing a lot, of course it's positive to be able to take it easy instead of getting involved." (P2)*

The opportunity to learn new communication and interaction skills and manage emotions in difficult situations was highlighted by several participants. For some, practicing emotion recognition in others was described as beneficial.

> *"Different strategies like counting slowly and breathing calmly. Have learned quite a lot actually … The advantage is that you learn to recognize facial expressions and the negative is that there are many hits you have to make." (P3)*

> *"The care has changed for me after VRAPT. I am more alert and awake to the situations I face … My own emotions are more stable … I used to find it easy to deal with anger but now it's easier. I have learned more about myself. I have learned what my weaknesses are…" (P4)*

> *"I learned to recognize and interpret facial expressions. Then I also learned to be able to handle stressful situations in a calm way … Then I learned to handle situations in a good way … There were situations that were stressful, for example when a security guard or police officer accused me of having done something, then I tried to reason and handle the situation…" (P6)*

Managing situations which give rise to tension and irritation in a constructive way while at the same time considering your own needs, was described as important. In this regard, practical skills training seemed to be crucial for the participants.

> *"What I learned the most from was when you had to relive the situation where you felt tense, irritated and angry. The role play, that's where I learned the most … Because it was a different way to digest different situations that you experienced. Normally you sit and think and reflect on things that have happened, but here you kind of got to experience it in a completely different way and also act at the same time … The most challenging thing for me was to talk about how to respect yourself as a person when dealing with other people. I was good at arguing against and respecting the other person, but respecting myself was lacking. To kind of tell how I feel in the given situation…" (P4)*

> *"It gave me many tools on how to act in many situations. I thought that … Yes, you should try to think about communicating, that's the only thing that solves everything, no violence … I learned more about stopping in time and so on … It was probably mostly when I had to communicate that I learned the most." (P7)*

3.2.2 Homework experiences

Several participants shared negative experiences of the homework assignments, some more related to repetitive aspects of the homework, and some due to associations to previous school experiences.

> *"I didn't think it was good because there were a lot of the same questions all the time. It became annoying that it didn´t differ." (P2)*

> *"It was like going to school again. You had to write homework and things like that." (P3)*

However, some participants described the homework assignments as a point of orientation and skills training during VRAPT, and that help from ward staff to remember the assignments was important.

> *"It gave structure to the whole project, that you had homework such as, okay, this week you should write down which situations you have experienced, how they have affected you, what thoughts and feelings have emerged, it became more structured to use it … It has been fun and rewarding, then you have to think, you had to do that…" (P4)*

> *"They were good but the hard part was remembering to do them. (*If the participant needed support*) Yes, I asked the staff but they didn't always remember, so there were some times when I didn't do the homework…" (P6)*

3.2.3 Staff involvement

The participants undergoing the VRAPT intervention were all within an inpatient care setting, with 24/7 staff availability. Thus, there were possibilities for ward staff to be involved in the intervention, for instance with homework assignments. Yet, ward staff involvement seemed to have been minimal. In some cases, this was preferable from the participant's perspective, but in other cases this was described as a shortcoming of VRAPT.

> *"Other staff have not been so involved in VRAPT, but they more asked "what are you doing now?" and if you explained that you are doing your homework, they saw you sitting with the notebook and doing your homework. They talked a bit about it but not more than that." (P4)*

3.2.4 Deviations from VRAPT methodology

VRAPT is a manualized treatment, with a manual complemented by therapist and participant workbooks to guide the treatment. Tailoring of the treatment content to participants' needs is possible, but the general concept should be followed for the intervention to be identified as VRAPT. During the interviews, examples of care interventions that were not part of VRAPT methodology yet took place during the intervention, were described. In some cases, this may have been related to general treatment needs in the patient — a need for someone to talk to about the symptomatology associated with

psychiatric illness. In other cases, skills training according to the manual was performed but outside of VR.

> *"Then we talked a lot, we talked a lot about things that had nothing to do with this VR study." (P1)*
>
> *"So what me and my supervisor did was that we role-played outside VR, face to face." (P4)*

3.2.5 Suggestions for improvement

During the interviews, participants were asked to describe what they believed needs to be improved in the VRAPT intervention. More time for skills training in VR was a common wish, both more time during the individual sessions to be able to penetrate the participant's management skills, but also possibilities for an extension, or booster sessions, of VRAPT.

> *"That you should have gone through the situation a little more than you did and how you reacted and so on … We had an hour but actually it would probably have taken longer than an hour for each session … felt that I would have needed a little more time each time." (P2)*
>
> *"…I would like to go to a continuation level or whatever you call it. Then I would have liked to be involved, but I guess that doesn´t exist…* (About wanting to have a continuation level) *Yes, three or four more months." (P7)*

Other specific parts that were suggested for improvements were the workbooks but also parts related to the VR technology, e.g. facial expressions of the avatars.

> (About the workbooks) *"It should have been rephrased a bit and been different, it would have been easier and more interesting, if you had varied the questions a bit more…" (P2)*
>
> (about emotion recognition) *"Yes, I wish they had improved that because they are so similar. Wait, what was it now, one was angry, one looked disgusted and those two were very similar and it is very difficult to choose which is which…" (P7)*

3.3 VR technology

Another content category evident in the data was content related to the VR technology per se. All VRAPT sessions except for the last session included time in the virtual world. That is, the participants in this study had several hours of experience in the VR environment, and the following subcategories of participants' experiences were discerned in the data: Novelty, Immersion, Physiological side-effects, and Limitations of technology on therapeutic work.

3.3.1 Novelty

For most participants, the use of VR was novel, which increased their interest in participating in the intervention. At the beginning, this involved many new experiences and could be both thrilling and uncomfortable to a certain degree. However, the majority of the participants shared positive experiences of using VR technology.

> *"My experience over time was that I thought the first times I tried it, it was fun because it was new and so on." (P1)*
>
> *"At first it felt a bit strange to use those glasses, like entering a completely different world. I'm not used to video games or anything like that. It felt a bit strange but at the same time exciting … It felt good. It was fun…" (P7)*

3.3.2 Immersion

A central concept in VR research and practice is immersion. The participants described how they experienced taking part in the VR experience, where some described a lack of immersion.

> *"That it would feel like you were in another world and lost everything around you but I didn't really get that experience…" (P1)*
>
> *"Then you would forget that it's the VR world, but you still did that to some extent … The advantage is that you are challenged to become stronger and the disadvantage is that it is virtual. By that I mean the role-plays, as I said, you couldn't immerse yourself 100%."* (P4)

When asked specifically about these experiences, participants referred to technological limitations in graphics and sound including voice distortion.

> *"It was bad graphics, the graphics engine itself was bad, for instance you couldn't sit down anywhere, it was very limited … No, it was just the psychologist's voice, but it was mixed and I was always aware of it … I was always aware that this with VR had a purpose, it became artificial in some way … It's too unrealistic." (P1)*
>
> *"In VR, you are like a robot…(Interviewer: If I understand you correctly, the voices made you lose the feeling of being there, like you disconnected from it at certain moment) Yes, sometimes I even started laughing because it feels so unreal. But you tried to focus on the task." (P4)*

For other patients, VR provided a possibility of being emotionally connected to the virtual world and, to a certain point, VR glasses became an almost non-existent accessory.

> *"It was emotional." (P5)*

> *"It's quite interesting, it feels like you end up in a different world once you put the glasses on, the headphones … It almost felt like the real world, except that it was simulated and maybe not as good graphics, but it felt very real … It can be heavy with the equipment but you forget about it once you are in the VR world, or you notice afterwards that it has been quite heavy. You don't notice it at the time…" (P6)*

3.3.3 Physiological side-effects

Some participants reported physiological symptoms when using the VR equipment, symptoms which gradually diminished with the passing of the sessions as they became accustomed to using VR.

> *"When you put on the VR glasses and have the controls in your hands and have to learn on-site, you might feel a bit dizzy, not really dizzy but you lose your balance a bit … It happened the first times but then you get used to it…" (P4)*

3.3.4 Limitations of technology on therapeutic work

During interviews, some of the previously described limitations and lack of outcome for participants were related to limitations of the technology used in the current study. This concerned both the VR technology and the wristwatch used to measure skin conductance and heart-rate variability. This affected not only the participant's experience during the session but also the participant's general view on participating in a treatment evaluation. Situations where the technology did not correspond to what is possible in reality was described as a limitation.

> *"…a lot of the equipment didn't work. For example, we should measure heart rate and sweating. It happened that I had a pulse of 30 and sometimes it was up to 250, it didn't work, something was wrong. Thus I became a bit skeptical about the study itself…" (P1)*

> *"…but you would also have been able to just walk away because it's also a strategy to just walk away and you didn't have that in those situations and I think that's a bit stupid." (P2)*

3.3.5 Potential of VR in forensic settings

When participants were asked about the possibilities of VR in forensic settings, different kinds of skills training were highlighted.

> *"Improving for patients. It might be … aggression and practicing interacting with others, practicing everyday life, going to the bank and everything possible." (P6)*

3.4 Previous treatment experiences

While the interview guide was specifically focused on VRAPT and VR, content related to participants' previous treatment experiences, both considering specific treatments but also concerning general treatment in forensic psychiatry, emerged. This was conceptualized in the subcategories: Previous experiences from aggression treatment and Patients' experiences of forensic psychiatric care.

3.4.1 Previous experiences from aggression treatment

Despite several of the participants having had experiences of different kinds of psychological treatments, none of them could describe ever having received any aggression-specific treatment. In some cases, the participants even shared experiences where therapists had refused to talk about aggression and that psychotherapy was aborted due to lack of progress. VRAPT was described as something novel, with possibilities for deeper penetration of participant's needs.

> *"Yes, I did have a few conversations with (my therapist), it's actually the first time I've gone through it with a psychologist, it never happened at the other clinic, no psychologist brought it up, they just said no…*(if the care provider talked about aggression) *No, it went so far that (previous therapist) terminated the contact with me, she couldn't understand me. Now the doctor has to take over with medication, she said, seven or eight years we had sat and talked and then she quit."* (P1)

> *"I haven't experienced so many. I've had psychological sessions and that's the only thing I can compare it to, and I'd like to say that VRAPT is more in-depth than regular sessions with a psychologist…" (P4)*

3.4.2 Patients' experiences of forensic psychiatric care

During the interviews, some participants reflected on experiences of forensic psychiatry in general, sharing negative memories and experiences of missing out on things happening in the outside world. Others described how being locked up affects them and may give rise to feelings of irritation.

> *"You know, I have been institutionalized since (several years) when I was incarcerated. It's hard. It's ward after ward, a*

corridor with an area of one meter where 15 people live … My own feelings are what I've gone through meeting after meeting for (several years) years, it takes a bit of a toll on my psyche but what can I do … I'm bitter, I'm done and ready to be treated in outpatient care. It feels like I've been here too long, the crimes were a long time ago but I haven't shown any day that I'm angry.. I also have feelings, mom, dad, siblings … I have lost my (relatives). I haven't been able to attend any funerals." (P5)
"Being imprisoned can make me irritable, and rules and restrictions." (P6)

3.5 Challenges to treatment of aggression

VRAPT was conceived as a treatment directed towards reactive aggression. In the interviews, it could be observed that aggression was more complex than that in this sample of forensic psychiatric patients. For some, instrumental aggression was the predominant form of aggression. For others, aggression stemmed from a severe mental illness, such as psychosis. Further complicating things, different types of aggression may have been present at different points in participants' lives. In the data, descriptions where participants had specific aims with their actions during therapy, or in relation to aggressive acts, that could pose significant challenges to treatment of aggression were visible. Two subcategories were discerned: Manipulation and Positive experiences from aggression.

3.5.1 Manipulation

Some participants were very aware of the therapist's intentions during treatment, so they just "played the game".

"You also noticed that some situations were made from the outside where the therapist thought what I was triggered by, so I did the opposite because I understood what (the therapist) was looking for. That (the therapist) wanted a reaction or something…" (P2)

3.5.2 Positive experiences from aggression

For other participants, aggression was perceived as something positive, and fantasies of harming others were common and fulfilling for the participant.

"Especially now after this, I had assured everyone that I would not do anything, but still it happened. Almost 10 last minutes without my own will. So for me, the aggression is not just negative, I get a positive experience from it … I've fantasized about hurting people a lot on the ward, but it's only been in the realm of fantasy then." (P1)

In some cases, the desire to hurt others was described as so intense that it was unthinkable to undergo an aggression treatment. Also, previous experiences of failed treatments paved way for a negative view on ever being able to succeed with aggression treatment.

"…I do find so many positive things in having these violent fantasies, I don't want to get rid of them … I personally don't suffer much from having this problem. I feel an enormous unwillingness to do anything about it because I've been in so much therapy that I've come to the conclusion that I'll never get rid of this, I feel hopeless when I think about dealing with the situation in a better way … I've almost given up on the idea that I could prevent myself from committing violent crimes…" (P1)

3.6 Unexpected experiences

3.6.1 Covid 19

During the course of a study, it is not uncommon for unexpected events to occur that may affect the research. In this case, as therapists began to see the participants in the VRAPT intervention, the first cases of Covid-19 appeared in Sweden. This led to a number of actions being taken to reduce patient-therapist contact, as well as actions to reduce the likelihood of patients infecting each other and, most of all, to prevent staff from entering the clinic with the virus. Given these circumstances, a specific part of the interview concerned how the pandemic may have affected the participants' experiences of taking part in the VRAPT intervention. In the data, it is clear that participants in general felt confident that the clinic took the necessary precautions to prevent them from becoming infected. For some participants, Covid-19 did not cause any worry.

"It wasn't a big deal, I don't see what the problem would be during a pandemic … Nothing changed. It didn't affect anything." (P3)

"Well, what we did was to keep our distance during sessions and have good hand hygiene and that was what mattered in those moments. Then you could see that there were corona routines with distance and hand sanitizers … The mask you wore, before you put on the VR glasses, that probably reduced the risk of getting infected." (P4)

4 Discussion

The aim of this study was to describe forensic psychiatric patients' experiences of having undergone the newly revised VRAPT intervention at a high-security forensic psychiatric clinic. The participants had very diverse experiences, and important findings not only specifically related to the VRAPT intervention itself but to the forensic psychiatric care setting emerged. Some participants had distinctly positive experiences, where specifically

skills training through role-plays in VR was emphasized as something beneficial that possibly could be extended. Challenges in the form of lack of motivation for aggression treatment among participants, and with technological limitations were described. In total, five content categories were manifest in the data: 1. Therapeutic process, 2. VRAPT method, 3. VR technology, 4. Previous treatment experiences, 5. Challenges to treatment of aggression, and 6. Unexpected experiences.

4.1 Therapeutic process

The participants described that the relationship with their therapists was an important aspect of their VRAPT treatment. Most participants described the professional relationship in a positive way, highlighting aspects such as open communication and feeling safe during the treatment. This can be related to what in psychological treatment literature is referred to as therapeutic, or working alliance (64, 65). For example, having the therapist describe clearly what was going to happen in a role-play could lead to increased feelings of safety, even in situations that could be challenging for the participant. However, there were also participants who described feelings of sadness and stigmatization, giving voice to feelings resembling hopelessness (e.g., "*I am also a human, I have made a mistake and am here to pay for it, doing all the programs, treatments, attending meetings and there is nothing more.*"). Given the context of the study, forensic psychiatry, possible impact of multiple stigmas identity (e.g., severe mental disorder, criminal history, ethnic/racial minorities) and self-stigma (66) may have affected the participants' experiences, especially regarding the therapeutic process in the study. We believe it is important to highlight here the importance of a critical, reflective and ethical perspective in the treatment of aggression with forensic patients, as we consider that the treatment of aggression (and in general, any treatment of vulnerable groups) cannot be based exclusively on following a manual. Maintaining a global perspective of the patient, and a holistic vision in relation to the treatment, is fundamental in forensic settings. Also, working with and following up the patient's experiences from the therapeutic collaboration is important, regardless of whether it is specified in a manual, something which may facilitate to discover and manage when patients have negative experiences during the treatment. In settings such as the one studied here, misunderstandings may arise between therapist and patient and the coercive nature of the care in general may affect the patient's overall experiences, which makes focused work with the therapeutic collaboration crucial. A possible paternalistic stance in the therapeutic collaboration might be seen in some quotes; *"My therapist had, I think, a genuine sense of developing the patient in the flaws that they possess."* Here, the context of the study must be considered, as high-security forensic psychiatry is a treatment context with clear coercive elements where such attitudes might occur and affect the care.

In terms of agreement on treatment goals, an aspect commonly described as a crucial part of the working alliance (64, 65), several participants found it difficult to recall the specific treatment goals. This can be related to several aspects, e.g. the fact that the interviews were conducted 12 weeks post treatment termination, possibly affecting their memory of specific treatment goals. Other aspects possibly influencing this could be the complex clinical profile of forensic psychiatric patients, with multiple difficulties including severe mental disorders (e.g., psychosis) and cognitive deficits (67, 68). In that regard, it is even more striking that one of the participants was able to elaborate on and relate the treatment goals to his own development. Interestingly, one of the subcategories that emerged, which is closely related to treatment goals, was matching of treatment to patient's needs. It seems as if some of the participants could not agree with the therapist on the overarching goal of the treatment — managing the participant's own aggression — since that would demand too much of the participant. Differences in preconceptions between participants and therapists on the nature of VR-assisted treatment of aggression, affecting the participant's perceived value of the treatment, were evident. Not surprisingly, spending initial time on a common understanding of goal and case formulation seems crucial for the participant's experience of the therapeutic process (69).

One part of the therapeutic process, which was highlighted as beneficial by several participants, was the skills training in the virtual worlds. During VR-assisted role-plays, some participants described discomfort due to them practicing managing challenging situations, i.e. it was difficult to find strategies for coping with social situations in which they were confronted with situations that had provoked aggression in the past. For some, this feeling of discomfort was the source of motivation to cope and learn new communication and self-regulation strategies. Continued skills training and practicing communication strategies were highlighted as remaining treatment needs, something which should be considered in future revisions of the VRAPT intervention.

4.2 VRAPT method

The majority of participants, albeit some with initial ambivalence, were satisfied with the outcomes of their VRAPT treatment, emphasizing the importance of talking about aggression (compared to previous experiences of not being allowed to discuss the subject), learning to recognize and interpret facial expressions and social situations, learning and practicing communication skills, and rehearse management of stressful situations through VR role-plays (14, 31). However, several participants mentioned homework assignments as demotivating, being too repetitive and reminding them of school experiences, while some participants described them as a central part of the treatment. Notably, despite staff being technically available 24/7, a lack of staff support with the homework assignments was mentioned. Homework assignments play an important role in cognitive and cognitive-behavioral psychological treatments, emphasized early by Beck and colleagues (70), in practicing and reinforcing the skills learned during treatment in the patient's everyday life (71). Here, future revisions of the VRAPT intervention need to assimilate these experiences into a renewed take on homework assignments, reducing repetitive wording and possibly directly involving ward

staff in homework assignments to increase the likelihood of completing them successfully. Hopefully, this may aid transference of learned skills to patients' everyday lives but is still to be evaluated in coming studies. Another critical variable to improve the impact of this kind of therapeutic intervention is the commitment of the patient or user (72), which should be measured in coming studies.

During interviews, deviations from the VRAPT methodology, e.g., performing face-to-face role-plays instead of VR-assisted role-plays and altering the direction of the treatment, were described. The provided descriptions suggest that these may have been necessary alterations from a clinical point of view. However, this obviously affects VRAPT treatment integrity. In larger, quantitative evaluations of the VRAPT intervention, the limit to which treatment integrity must be maintained in order for the treatment to be conceptualized as "VRAPT" must be clarified. In clinical everyday practice this may, however, prove challenging. Forensic psychiatric care has many facets, and in the course of a manualized treatment (e.g., VRAPT) an indeterminate number of events occur, which implies incorporating different institutional levels when implementing and evaluating new virtual reality treatments (73, 74).

4.3 VR technology

The majority of participants had positive experiences of using the VR technology and perceived it as "novel" and "fun", something which in itself may have been a motivating factor for participating in the treatment. Participants formulated areas within forensic psychiatry, specifically related to different kinds of skills training, where they saw VR as having potential for improving interventions. The current results are believed to add to previous descriptions of VR as a way of increasing the possibilities for bridging the "gap" between the real world and the therapeutic setting and as motivation-enhancing within forensic settings (46).

Considering main concepts in VR — presence, immersion, and embodiment (17–20) — our results suggest that the applied technology created a sense of presence and immersion for many participants. In some cases, participants described that the technology was good enough for them "to be there" (i.e., experience presence). However, some participants reported not feeling immersed enough, and attributed this to technological limitations e.g. graphics. Other technological limitations described by participants were related to a sense of embodiment, in that the participants found the lack of being able to perform their own physical movements in the VR world limiting (e.g., not being able to back away during a discussion). However, no severe or persisting cases of cybersickness were reported, with such experiences limited to the beginning of VRAPT. In previous studies on forensic psychiatric patients, cybersickness has been one of the causes of dropout (41). Overall, the technology applied in the current VRAPT intervention seems valid for a "good enough" experience without major side-effects for the majority of participants. However, the descriptions of technological limitations hindering a sense of immersion and embodiment, can be considered in future revisions of the software.

4.4 Previous treatment experiences

None of the participants described having previously received any type of treatment focused specifically on aggression problems, making comparisons of the VRAPT intervention to other aggression treatments impossible. A positive approach towards interventions such as VRAPT was noted. However, participants spontaneously shared their general experiences of forensic psychiatric care. Many of these were negative experiences of being locked up and frustration concerning the forensic psychiatric care system. For the patients, adapting to the forensic institution may often be a challenging process, and focusing on development of a sense of self and connectedness has been suggested to help enhance recovery (75).

In order to understand and treat aggression in a forensic psychiatric context, it is necessary to include those situations that incite patients to aggression within the institution. Previous studies have revealed that both social climate and sense of community predict aggressiveness in high-security hospitals (76). Other studies (77) have indicated that institutional restrictions and patients' psychopathology influence treatment course and outcome of forensic psychiatric patients. With this in mind, all treatments within the forensic psychiatric setting will undoubtedly be influenced by the patients' previous experiences from this, or similar, settings. This is a more general note than something specific to the VRAPT intervention, however.

4.5 Challenges to treatment of aggression

Some of the participants described their own attitudes associated with instrumental (or "recreational") violence, where, for some, the desire to harm was a fundamental part of how they identified themselves. Descriptions of thought-out manipulations of the therapeutic process may be in line with a more instrumental aggression, and with psychopathy in general, but it is unclear how an assessment of psychopathy could uniquely inform treatment and rehabilitation strategies (78, 79). VRAPT was not designed as a treatment for instrumental aggression, which may be why these participants in previous content categories provided negative experiences of the intervention. The established categorization of aggression into reactive, instrumental and psychotic aggression has been confirmed in long-stay public psychiatric hospitals (80), and we consider this categorization to be possibly useful for tailoring treatment for aggression in forensic settings. In its current form, VRAPT should be specifically directed for persons with predominantly reactive aggression. However, aggression should not be understood as a linear and categorizable process as it is more adequately described as complex and multidimensional but may be communicated more easily in categories. Thus, the multifaceted nature of dysfunctional aggression must be

acknowledged in interventions, while interventions such as VRAPT might need to focus on the type of aggression where a potential benefit for the participants (and the safety of society) is most obvious – in the case of VRAPT the reactive aggression. On the other hand, we believe that it is important to highlight responsivity as modeled in the Risk-Need-Responsivity (RNR) model (81) during the VR-assisted aggression treatment because of the complexity of forensic psychiatric patients, and as well as the challenges therapists must consider in clinical relationships with forensic clients (82).

4.6 Unexpected experiences

As described previously, the VRAPT intervention that the participants in this study completed was initiated at the same time as Covid-19 started to spread over the world. However, when asked about in the interviews, this was described as being handled with specific routines and not creating any feelings of being unsafe among the participants. This suggests that the routines related to Covid-19 were assimilated as part of the everyday care at the clinic. During the pandemic, many changes occurred in forensic psychiatric institutions (83). However, in this study, the general perception of the participants was that the clinic handled it adequately and that they felt confident in the actions taken. In summary, the participants described that the pandemic did not negatively influence their participation in the VRAPT treatment.

4.7 Limitations and strengths

For the research team, the entire logistical organization of the study, including the interviews, was difficult and complex, as the study was initiated at the same time that the COVID-19 pandemic incapacitated the world. Nevertheless, the patients do not report any limitations due to COVID-19 affecting their participation in a negative way. Another possible limitation was that the interviews first were conducted 12 weeks after the VRAPT intervention was completed, to not affect a quantitative follow-up that was conducted 12 weeks post-VRAPT. It may have been that conducting interviews with patients just after their completion of VRAPT could have resulted in a somewhat different description of their experiences. Given the scarcity of qualitative studies in this field, such limitations cannot be completely avoided and must be recognized. Another limitation is something which must be handled in qualitative studies in general – sometimes participants' quotes may seem to apply to different categories. For example, the subcategories "treatment outcomes" and "remaining treatment needs" in this study do show similarities. However, in the interpretive process, the proposed content categories and subcategories were deemed to be sufficiently distinct. Furthermore, the sample only consisted of seven participants, something which could be considered as a smaller sample even in qualitative research (60). However, the study was performed with a narrow focus and on a total sample from a clinical pilot study, and our assessment was that data saturation was achieved through the interviews. The inherent bias of the interviewer and researchers who conducted the analyses, all forensic psychologists with considerable experience from working in forensic psychiatry, during the study and interpretation of the data should also be considered a limitation. A final limitation of this study was the imbalance in the gender of the participants — 6 men and 1 woman. This makes it difficult to have a broad perspective of how female patients perceive undergoing the VRAPT intervention. However, this gender imbalance is in line with the gender distribution of forensic psychiatric patients in general in Sweden. In the future, it is recommended that VR-assisted treatment of aggression should involve a larger proportion of, or focus exclusively on, female participants.

A significant strength of this study was that all participants who completed VRAPT at the time of the study agreed to participate in interviews. Notably, participants described that they were proud to participate and to be part of the technological development in society. Considerations on credibility, dependability and transferability to ensure trustworthiness of outcomes (61) were made through assuring diversity in the sample with various cultural backgrounds, ages and levels of education, various psychiatric problems, different types of crimes and varying length of stay in forensic psychiatry.

5 Conclusion

This study reports forensic psychiatric patients' experiences of recently having undergone VR-assisted aggression treatment through the VRAPT intervention. Interviews demonstrated skills training through role-plays in VR as something that was perceived as beneficial by several participants and suggested that this part could be extended. Challenges in the form of lack of motivation for aggression treatment among participants, and with technological limitations were described. The categories that emerged bottom-up from interviews should be put in relation to clinical challenges within the specific context; patients' experiences from the intervention need to be understood in relation to how psychotic functioning, including paranoia, may affect their reaction to VR-assisted interventions and to aggression management in particular. Nevertheless, previous studies indicate that VR-assisted interventions, also those focusing on aggression, are feasible in these clinical settings (14, 41, 42).

The study highlights the importance of educating providers (in this case therapists) on how to use VR, its possibilities and technological limitations. Continued qualitative studies on patients' experiences is recommended, given that this can help care providers tailor treatment interventions to patients' needs.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Swedish Ethical Review Authority (Dnr: 2019-02337; 2020-06317). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FG: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing, Formal analysis. PE: Conceptualization, Methodology, Supervision, Writing – review & editing. SP: Conceptualization, Supervision, Writing – review & editing. KS: Conceptualization, Supervision, Writing – review & editing. WV: Conceptualization, Writing – review & editing. MW: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Swedish Research Council for Health, Working Life and Welfare (#2018-01409), by the Southern Healthcare Region in Sweden (#933868), and by Region Kronoberg (#936434; #989068).

Acknowledgments

Thanks to Mona Göthe for participation in transcription of interviews. Thanks to Stéphanie Klein Tuente for great help during the start-up of the study. A big thank you to all participants, for sharing their experiences.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

1. Munthe C, Radovic S, Anckarsäter H. Ethical issues in forensic psychiatric research on mentally disordered offenders. *Bioethics*. (2010) 24:35–44. doi: 10.1111/j.1467-8519.2009.01773.x

2. Shapiro DL. Ethical issues in forensic psychology and psychiatry. *Ethics Med Public Health*. (2016) 2:45–58. doi: 10.1016/j.jemep.2016.01.015

3. Völlm B, Bartlett P, McDonald R. Ethical issues of long-term forensic psychiatric care. *Ethics Med Public Health*. (2016) 2:36–44. doi: 10.1016/j.jemep.2016.01.005

4. Bogaerts S, Polak M, Spreen M, Zwets A. High and low aggressive narcissism and anti-social lifestyle in relationship to impulsivity, hostility, and empathy in a group of forensic patients in the Netherlands. *J Forensic Psychol Pract*. (2012) 12:147–62. doi: 10.1080/15228932.2012.650144

5. Lobbestael J, Cima M, Lemmens A. The relationship between personality disorder traits and reactive versus proactive motivation for aggression. *Psychiatry Res*. (2015) 229:155–60. doi: 10.1016/j.psychres.2015.07.052

6. Krona H, Anckarsäter H, Nilsson T, Hofvander B. Patterns of lifetime criminality in mentally disordered offenders–findings from a nationally representative cohort. *Front Psychiatry*. (2021) 12. doi: 10.3389/fpsyt.2021.564171

7. Gatner DT, Moulden HM, Mamak M, Chaimowitz GA. At risk of what? Understanding forensic psychiatric inpatient aggression through a violence risk scenario planning lens. *Int J Forensic Ment Health*. (2021) 20:398–407. doi: 10.1080/14999013.2021.1899343

8. Wallinius M. *Aggressive antisocial behavior-clinical, cognitive, and behavioral covariates of its persistence*. Lund: Doctoral dissertation, Lund University (2012).

9. González Moraga FR, Tuente S, Perrin S, Enebrink P, Sygel K, Veling W, et al. New developments in virtual reality-assisted treatment of aggression in forensic settings: the case of VRAPT. *Front Virtual Real*. (2022) 2: doi: 10.3389/frvir.2021.675004

10. González Moraga FR, Garcia D, Billstedt E, Wallinius M. Facets of psychopathy, intelligence and aggressive antisocial behaviors in young violent offenders. *Front Psychol*. (2019) 10:984. doi: 10.3389/fpsyg.2019.00984

11. Howner K, Andiné P, Bertilsson G, Hultcrantz M, Lindström E, Mowafi F, et al. Mapping systematic reviews on forensic psychiatric care: a systematic review identifying knowledge gaps. *Front Psychiatry*. (2018) 9:452. doi: 10.3389/fpsyt.2018.00452

12. Lee AH, DiGiuseppe R. Anger and aggression treatments: a review of meta-analyses. *Curr Opin Psychol*. (2018) 19:65–74. doi: 10.1016/j.copsyc.2017.04.004

13. Trestman RL. Treating aggression in forensic psychiatric settings. *J Am Acad Psychiatry Law*. (2017) 45:40–3.

14. Sygel K, Wallinius M. Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. *Front Psychiatry*. (2021) 12. doi: 10.3389/fpsyt.2021.673089

15. Geraets CN, van der Stouwe EC, Pot-Kolder R, Veling W. Advances in immersive virtual reality interventions for mental disorders: A new reality? *Curr Opin Psychol*. (2021) 41:40–5. doi: 10.1016/j.copsyc.2021.02.004

16. Geraets C, Wallinius M, Sygel K. Use of virtual reality in psychiatric diagnostic assessments: A systematic review. *Front Psychiatry*. (2022) 260. doi: 10.3389/fpsyt.2022.828410

17. Slater M. Place Illusion and Plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc London*. (2009) 364:74. doi: 10.3389/frobt.2016.00074

18. Bailey JO, Bailenson JN, Casasanto D. When does virtual embodiment change our minds? *Presence*. (2016) 25:222–33. doi: 10.1162/PRES_a_00263

19. Riva G. *Virtual reality in clinical psychology. In reference module in neuroscience and biobehavioral psychology*. Amsterdam: Elsevier (2020).

20. Slater M. Immersion and the illusion of presence in virtual reality. *Br J Psychol*. (2018) 109:431–3. doi: 10.1111/bjop.12305

21. Lanier J. *Dawn of the new everything: A journey through virtual reality*. London: Random House (2017).

22. Cummings JJ, Bailenson JN. How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychol*. (2016) 19:272–309. doi: 10.1080/15213269.2015.1015740

23. Lee KM. Presence, explicated. *Communication Theory*. (2004) 14:27–50. doi: 10.1093/ct/14.1.27

24. Ticknor B. *Virtual reality and the criminal justice system: exploring the possibilities for correctional rehabilitation*. Lanham: Lexington Books (2018).

25. Slater M, Spanlang B, Corominas D. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans Graphics (TOG)*. (2010) 29:1–9. doi: 10.1145/1778765.1778829

26. Kilteni K, Groten R, Slater M. The sense of embodiment in virtual reality. *Presence: Teleoperators Virtual Environments*. (2012) 21:373–87. doi: 10.1162/PRES_a_00124

27. Fromberger P, Jordan K, Müller JL. Use of virtual reality in forensic psychiatry: a new paradigm? *Der Nervenarzt*. (2014) 85:298–303. doi: 10.1007/s00115-013-3904-7

28. Kip H, Bouman YHA, Kelders SM, van Gemert-Pijnen L. J. E. W. C. eHealth in treatment of offenders in forensic mental health: a review of the current state. *Front Psychiatry*. (2018) 9:42. doi: 10.3389/fpsyt.2018.00042

29. Kisker J, Gruber T, Schöne B. Experiences in VR entail different processes of retrieval as opposed to conventional laboratory settings: A study on human memory. *Curr Psychol*. (2019) 40:3190–7. doi: 10.1007/s12144-019-00257-2

30. Spiegel B. *VRx: how virtual therapeutics will revolutionize medicine*. New York: Basic Books (2020).

31. Tereso IM, Ramos ALG, Santos BR, Costa JPM. Virtual reality and forensic mental health. In: *In digital therapies in psychosocial rehabilitation and mental health*. IGI Global (2022). p. 245–60. doi: 10.4018/978-1-7998-8634-1

32. Ireland JL. Individual assessments of aggression: Accounting for core factors. In: Ireland JL, Ireland CA, Birch P, editors. *Violent and sexual offenders: Assessment, treatment and management, 2nd ed*. Routledge, Abingdon (2018). p. 203–27.

33. Lewis M, Ireland JL. Understanding motives for aggression in forensic psychiatric patients: A preliminary study. *J Forensic Psychiatry Psychol*. (2019) 30:496–509. doi: 10.1080/14789949.2019.1570541

34. Bjørnebekk G, Howard R. Validation of a motivation-based typology of angry aggression among antisocial youths in Norway. *Behav Sci Law*. (2012) 30:167–80. doi: 10.1002/bsl.2007

35. Howard R, Howells K, Jinks M, McMurran M. A quadripartite typology of violence (QTV): Relationships with functions of aggression in violent youths. In: Needham I, Callaghan P, Palmstierna T, Nijman H, Oud N, editors. *Violence in clinical psychiatry*. Kavanah, The Netherlands (2009). p. 342–5.

36. Camp JP, Skeem JL, Barchard K, Lilienfeld SO, Poythress NG. Psychopathic predators? Getting specific about the relation between psychopathy and violence. *J consulting Clin Psychol*. (2013) 81:467. doi: 10.1037/a0031349

37. Daffern M, Howells K, Ogloff J. What's the point? Towards a methodology for assessing the function of psychiatric inpatient aggression. *Behav Res Ther*. (2007) 45:101–11. doi: 10.1016/j.brat.2006.01.011

38. Fletcher A, Crowe M, Manuel J, Foulds J. Comparison of patients' and staff's perspectives on the causes of violence and aggression in psychiatric inpatient settings: An integrative review. *J Psychiatr Ment Health Nurs*. (2021) 28:924–39. doi: 10.1111/jpm.12758

39. Smeijers D, Koole SL. Testing the effects of a virtual reality game for aggressive impulse management (VR-GAIME): study protocol. *Front Psychiatry*. (2019) 10:83. doi: 10.3389/fpsyt.2019.00083

40. Tuente S, Bogaerts S, Van Ijzendoorn S, Veling W. Effect of virtual reality aggression prevention training for forensic psychiatric patients (VRAPT): study protocol of a multi-center RCT. *BMC Psychiatry*. (2018) 18:1–9. doi: 10.1186/s12888-018-1830-8

41. Smeijers D, Bulten EH, Verkes RJ, Koole SL. Testing the effects of a virtual reality game for aggressive impulse management: A preliminary randomized controlled trial among forensic psychiatric outpatients. *Brain Sci*. (2021) 11:1484. doi: 10.3390/brainsci11111484

42. Tuente S, Bogaerts S, Bulten E, Keulen-de Vos M, Vos M, Bokern H, et al. Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *J Clin Med*. (2020) 9:2258. doi: 10.3390/jcm9072258

43. Lobbestael J, Cima MJ. Virtual reality for aggression assessment: The development and preliminary results of two virtual reality tasks to assess reactive and proactive aggression in males. *Brain Sci*. (2021) 11:1653. doi: 10.3390/brainsci11121653

44. Quarmley M, Vafiadis A, Jarcho JM. Irritability and rejection-elicited aggression in adolescents and young adults. *J Child Psychol Psychiatry*. (2023) 64.9. doi: 10.1111/jcpp.13804

45. Bailenson J. *Experience on demand: What virtual reality is, how it works, and what it can do*. New York: WW Norton & Company (2018).

46. Kip H, Kelders SM, Weerink K, Kuiper A, Brüninghoff I, Bouman YH, et al. Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. *Front Psychol*. (2019) 10:406. doi: 10.3389/fpsyg.2019.00406

47. Lutz M, Zani D, Fritz M, Dudeck M, Franke I. A review and comparative analysis of the risk-needs-responsivity, good lives, and recovery models in forensic psychiatric treatment. *Front Psychiatry*. (2022) 13:988905. doi: 10.3389/fpsyt.2022.988905

48. Söderberg A, Wallinius M, Munthe C, Rask M, Hörberg U. Patients' experiences of participation in high-security, forensic psychiatric care. *Issues Ment Health Nurs*. (2022) 43:683–92. doi: 10.1080/01612840.2022.2033894

49. Ivarsson D, Delfin C, Enebrink P, Wallinius M. Pinpointing change in virtual reality assisted treatment for violent offenders: a pilot study of Virtual Reality Aggression Prevention Training (VRAPT). *Frontiers in psychiatry*. (2023) 14, 1239066. doi: 10.3389/fpsyt.2023.1239066

50. Anderson CA, Bushman BJ. Human aggression. *Psychology*. (2002) 53:27. doi: 10.1146/annurev.psych.53.100901.135231

51. Anderson CA, Carnagey NL. Violent evil and the general aggression model. In: *The social psychology of good and evil* (2004) New York: Guilford. p. 168–92.

52. DeWall CN, Anderson CA, Bushman BJ. The general aggression model: Theoretical extensions to violence. *Psychol violence*. (2011) 1:245. doi: 10.1037/a0023842

53. Kyngäs H. Inductive content analysis. In: *In The application of content analysis in nursing science research*. Springer, Cham (2020). p. 13–21.

54. Vears DF, Gillam L. Inductive content analysis: A guide for beginning qualitative researchers. *Focus Health Prof Education: A Multi-disciplinary J*. (2022) 23:111–27. doi: 10.11157/fohpe.v23i1.544

55. Fogg BJ. *Persuasive technology: using computers to change what we think and do* Vol. 2002. Ubiquity (2002) Burlington, MA: Morgan Kaufmann. p. 2.

56. Blitz MJ. Extended reality, mental liberty, and state power in forensic settings. *AJOB Neurosci*. (2022) 13:173–6. doi: 10.1080/21507740.2022.2086647

57. Kellmeyer P, Biller-Andorno N, Meynen G. Ethical tensions of virtual reality treatment in vulnerable patients. *Nat Med*. (2019) 25:1185–8. doi: 10.1038/s41591-019-0543-y

58. Ligthart S, Meynen G, Biller-Andorno N, Kooijmans T, Kellmeyer P. Is virtually everything possible? The relevance of ethics and human rights for introducing extended reality in forensic psychiatry. *AJOB Neurosci*. (2022) 13:144–57. doi: 10.1080/21507740.2021.1898489

59. Lidz CW, Appelbaum PS. The therapeutic misconception: problems and solutions. *Med Care*. (2002) 40(9):V55–63. doi: 10.1097/00005650-200209001-00008

60. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Soc Sci Med*. (2022) 292:114523. doi: 10.1016/j.socscimed.2021.114523

61. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today*. (2004) 24:105–12. doi: 10.1016/j.nedt.2003.10.001

62. Bengtsson M. How to plan and perform a qualitative study using content analysis. *NursingPlus Open*. (2016) 2:8–14. doi: 10.1016/j.npls.2016.01.001

63. Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Educ Today*. (2017) 56:29–34. doi: 10.1016/j.nedt.2017.06.002

64. Bordin ES. "Theory and research on the therapeutic working alliance: New directions," in *The Working Alliance: Theory, Research, and Practice*, eds Horvath A. O., Greenberg L. S. Hoboken, NJ: John Wiley and Sons (1994). p. 13–37.

65. Clercx M, de Vogel V, Lancel M, Keulen-de Vos M. The influence of therapy alliance and treatment motivation in patients with Cluster B personality disorders on incidents in forensic hospitals. *J Forensic Pract*. (2021) 23:272–84. doi: 10.1108/JFP-05-2021-0022

66. West ML, Yanos PT, Mulay AL. Triple stigma of forensic psychiatric patients: Mental illness, race, and criminal history. *Int J Forensic Ment Health*. (2014) 13:75–90. doi: 10.1080/14999013.2014.885471

67. Andiné P, Bergman H. Focus on brain health to improve care, treatment, and rehabilitation in forensic psychiatry. *Front Psychiatry*. (2019) 10:840. doi: 10.3389/fpsyt.2019.00840

68. Laporte N, Ozolins A, Westling S, Westrin Å, Wallinius M. Clinical characteristics and self-harm in forensic psychiatric patients. *Frontiers in psychiatry*. (2021) 12, 698372. doi: 10.3389/fpsyt.2021.698372

69. Hart S, Sturmey P, Logan C, McMurran M. Forensic case formulation. *Int J Forensic Ment Health*. (2011) 10:118–26. doi: 10.1080/14999013.2011.577137

70. Beck AT. *Cognitive therapy of depression*. New York: Guilford press (1979).

71. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. *JMIR Ment Health*. (2017) 4: e5283. doi: 10.2196/mental.5283

72. Fleming TM, De Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. *Front Psychiatry*. (2016) 7:65. doi: 10.3389/fpsyt.2016.00065

73. Papalia N, Spivak B, Daffern M, Ogloff JR. A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clin Psychology: Sci Pract*. (2019) 26:e12282. doi: 10.1111/cpsp.12282

74. Kouijzer MM, Kip H, Bouman YH, Kelders SM. Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. *Implementation Sci Commun*. (2023) 4:1–29. doi: 10.1186/s43058-023-00442-2

75. Clarke C, Lumbard D, Sambrook S, Kerr K. What does recovery mean to a forensic mental health patient? A systematic review and narrative synthesis of the qualitative literature. *J Forensic Psychiatry Psychol*. (2016) 27:38–54. doi: 10.1080/14789949.2015.1102311

76. Puzzo I, Aldridge-Waddon L, Bush E, Farr C. The relationship between ward social climate, ward sense of community, and incidents of disruptive behavior: a study of a high secure psychiatric sample. *Int J Forensic Ment Health*. (2019) 18:153–63. doi: 10.1080/14999013.2018.1532972

77. Franke I, Büsselmann M, Streb J, Dudeck M. Perceived institutional restraint is associated with psychological distress in forensic psychiatric inpatients. *Front Psychiatry*. (2019) 10:410. doi: 10.3389/fpsyt.2019.00410

78. Salekin RT. Psychopathy and therapeutic pessimism: Clinical lore or clinical reality? (2002)) 22.1:79–112. Clinical. doi: 10.1016/S0272-7358(01)00083-6

79. Larsen RR, Jalava J, Griffiths S. Are Psychopathy Checklist (PCL) psychopaths dangerous, untreatable, and without conscience? A systematic review of the empirical evidence. *Psychology Public Policy Law*. (2020) 26:297. doi: 10.1037/law0000239

80. Quanbeck CD, McDermott BE, Lam J, Eisenstark H, Sokolov G, Scott CL. Categorization of aggressive acts committed by chronically assaultive state hospital patients. *Psychiatr Serv*. (2007) 58:521–8. doi: 10.1176/appi.ps.58.4.521

81. Skeem JL, Steadman HJ, Manchak SM. Applicability of the risk-need-responsivity model to persons with mental illness involved in the criminal justice system. *Psychiatr Serv*. (2015) 66:916–22. doi: 10.1176/appi.ps.201400448

82. Chudzik L, Aschieri F. Clinical relationships with forensic clients: A three-dimensional model. *Aggression Violent Behav*. (2013) 18:722–31. doi: 10.1016/j.avb.2013.07.027

83. Terkildsen MD, Vestergaard LK, Møllerhøj J, Sørensen LU. Forensic psychiatric patients' Perspectives on COVID-19 prevention measures: A qualitative study. *J Forensic Psychol Res Pract*. (2022) 24.2: 245–67. doi: 10.1080/24732850.2022.2118095

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 30 April 2024
DOI 10.3389/fpsyg.2024.1284983

OPEN ACCESS

EDITED BY
Gian Mauro Manzoni,
University of eCampus, Italy

REVIEWED BY
Valeria Sebri,
European Institute of Oncology (IEO), Italy
Bhoopesh Sharma,
Shree Guru Gobind Singh Tricentenary University, India

*CORRESPONDENCE
M. T. E. Kouijzer
✉ m.t.e.kouijzer@utwente.nl

RECEIVED 29 August 2023
ACCEPTED 15 April 2024
PUBLISHED 30 April 2024

CITATION
Kouijzer MTE, Kip H, Kelders SM and Bouman YHA (2024) The introduction of virtual reality in forensic mental healthcare – an interview study on the first impressions of patients and healthcare providers regarding VR in treatment.
Front. Psychol. 15:1284983.
doi: 10.3389/fpsyg.2024.1284983

The introduction of virtual reality in forensic mental healthcare – an interview study on the first impressions of patients and healthcare providers regarding VR in treatment

M. T. E. Kouijzer[1]*, H. Kip[1,2], S. M. Kelders[1,3] and Y. H. A. Bouman[2]

[1]Department of Technology, Human and Institutional Behaviour, Centre for eHealth and Wellbeing Research, University of Twente, Enschede, Netherlands, [2]Department of Research, Transfore, Deventer, Netherlands, [3]Opentia Research Unit, North-West University, Vanderbijlpark, South Africa

Background: Recognizing the potential of Virtual Reality (VR) as a powerful technology to support behavior change, the careful introduction of this technology into treatment settings is essential. This is especially important in vulnerable populations like forensic psychiatric patients. This study aims to gain insight from the impressions of both patients and healthcare providers concerning the integration of VR in practice. The study aims to contribute valuable information that guides the introduction of VR technology, ensuring its optimal use in the complex context of forensic mental healthcare.

Methods: Semi-structured interviews were conducted with healthcare providers (n = 10) working at forensic outpatient clinics and forensic psychiatric patients (n = 8). All participants experienced VR before the interview. Inductive thematic analysis was employed for analyzing the interview data.

Results: Patients valued the unique opportunity to simulate personal experiences in VR scenarios and reflect on them with healthcare providers. In addition to positive first impressions, areas for improvement were identified, including the wish for enhanced realism and reduced physical discomfort while immersed in VR. Finally, important factors contributing to the successful introduction of VR were identified. For example, taking into account psychological distress experienced by patients or supporting healthcare providers with implementation resources.

Conclusion: The integration of VR into forensic mental healthcare holds great potential for behavior change. However, its immersive characteristics also increase the chance of amplifying psychological distress. This emphasizes the need for caution when using VR– especially when a vulnerable patient group is subjected to triggering scenarios. This study advocates for a gradual introduction of the technology and provides valuable insights into essential elements for this introduction in clinical practice. It highlights that even the initial step of integrating VR into practice – the introduction phase – demands careful planning and a personalized approach. This underscores the need for ongoing refinement and a systematic approach to the overall implementation of VR. These efforts are crucial to fully realize its potential in clinical practice.

KEYWORDS
virtual reality, implementation, forensic mental healthcare, interviews, VR, treatment

1 Introduction

Forensic psychiatry, a specialized field within mental healthcare, is focused on the assessment and treatment of individuals whose behavior has led or could lead to offending, often complicated by one or more psychiatric disorders (Mullen, 2000). For forensic psychiatric patients, a notable difficulty with recognizing and regulating their emotions is often evident, presenting a key area for intervention and exploration (García-Sancho et al., 2014; Roberton et al., 2015; Garofalo et al., 2018). This opens up possibilities for innovative approaches to enhance emotion awareness and regulation. Recent studies have highlighted Virtual Reality (VR) as a potentially suitable treatment tool for addressing emotion regulation challenges (Kip et al., 2019a,b; Tuente et al., 2020; Smeijers et al., 2021). VR offers the possibility to immerse patients in a unique virtual environment that can simulate real-world scenarios through a head-mounted display and 3D graphics (Skip et al., 2018). Particularly, interactive VR shows promise in bridging the gap between the treatment room and the outside world (Botella et al., 2017; Tuente et al., 2020). In interactive VR, patients are immersed in real-world scenarios that allow them to experience a sense of presence while interacting with the virtual world as if they were physically present within it (Botella et al., 2017; Tuente et al., 2020). VR allows patients to experience a sense of belonging within a virtual body but also actively contributes to regulating emotional well-being (Slater et al., 2008, 2010; Ventura et al., 2018; So et al., 2022). It offers patients the unique opportunity to engage in therapeutic activities, providing them with a safe space to practice new behaviors and coping strategies (Botella et al., 2017; Sygel and Wallinius, 2021).

Interactive VR offers various treatment opportunities within forensic psychiatry. It can be applied to expose patients to stimuli or situations that can elicit an emotional response such as fear or anger (Botella et al., 2017). By gradually exposing patients to these scenarios in a safe and controlled environment, they can learn to better manage their anxiety, fear, or aggression by practicing coping strategies (Botella et al., 2017; Baniasadi et al., 2020; González Moraga et al., 2022). For example, a patient with emotion regulation issues could be exposed in a role-play to a relatively strict police officer and practice their relaxation and communication skills. Besides exposure, VR can be used as a tool for assessment of individuals' risk of violence or re-offending, e.g., by recreating virtual scenarios that may trigger problem behavior that resembles behavior outside of the treatment room, allowing healthcare providers to observe and evaluate patients' reactions and potential risk factors in real-world situations (Renaud et al., 2014). Despite the growing awareness of the possibilities offered by VR, the use of VR in practice remains in its infancy (Garrett et al., 2018; Smith et al., 2020).

While research has demonstrated the potential benefits of VR within forensic psychiatric patients, the practical implementation of VR in practice often lags behind (Kouijzer et al., 2023). This process of implementation, crucial for the effective use of VR, encounters obstacles because of implementation barriers like limited familiarity with the technology, resistance to change, and technological apprehension (Kouijzer et al., 2023). To conquer these challenges, a thorough introduction to VR is advised (Kouijzer et al., 2023). From an implementation perspective, introducing the new technology among the people who need to work with it, such as healthcare providers and patients. The introduction of VR refers to familiarizing healthcare providers with the technology before its actual application in treatment and letting patients gradually acclimate to VR during the initial treatment sessions (Kouijzer et al., 2023). From an ethical point of view, a careful introduction of new technology is important, especially in this unique and vulnerable target group of forensic mental healthcare patients where transgressive behavior and a variety of psychiatric disorders play an important role (Fassaert et al., 2016). The introduction of VR technology itself requires cautious consideration due to its immersive and intrusive characteristics (Baniasadi et al., 2020). Because of these characteristics, VR can elicit intense emotional and psychological responses (Fromberger et al., 2014; González Moraga et al., 2022). The virtual environments and scenarios created in VR can be highly realistic, exposing patients to situations that could trigger their problem behavior or simulate traumatic experiences. Recalling traumatic experiences or memories can be highly effective in treatment, as demonstrated in EMDR therapy (Portigliatti Pomeri et al., 2021). However, it may also lead to heightened psychological distress, which could have unintended consequences on the mental well-being and safety of patients and healthcare providers involved in the VR treatment (González Moraga et al., 2022). This indicates the importance of a balanced approach that weighs potential therapeutic gains against the potential risks and fosters a careful introduction into practice (González Moraga et al., 2022).

To determine how to introduce VR in practice, it is important to consider the perspective of stakeholders in the development and implementation phase of the technology (van Gemert-Pijnen et al., 2018). Ventura et al. (2018) underline the importance of a focus on user experience in this introduction, especially when introducing a new technology as a treatment tool. They emphasized that future studies should focus on the psychological aspects and personal feelings of participants during a full-body immersion in VR (Ventura et al., 2018). By understanding the first impressions of both healthcare providers and patients regarding the VR intervention, their initial reactions and perceptions are explored. These play an important role in shaping the overall implementation strategy. They offer insight into how stakeholders perceive the technology's potential benefits and challenges, allowing for a more informed and effective integration process and thus increasing the chances of adoption and long-term use (van Gemert-Pijnen et al., 2018). Additionally, involving end-users prevents a top-down approach in which researchers or software developers dictate how VR should be introduced. It allows for optimal fit between the needs and wishes of patients and healthcare providers and the technology, making sure that VR is of added value for them (Kip et al., 2019a,b).

1.1 The current study

Given the immersive characteristics of VR technology, the current study places a significant focus on understanding the perspectives of end-users to navigate the careful introduction of this technology in a vulnerable forensic population. The primary objective is to gain insight into patients' and healthcare providers' initial impressions and perspectives regarding the use of VR in forensic mental healthcare. By prioritizing the examination of user experience, the study aims to contribute to a balanced and informed approach that weighs therapeutic gains against potential risks for patients and healthcare providers, fostering a careful integration of VR technology in the treatment of forensic psychiatric patients. To achieve this overall aim, the following research questions will be addressed:

1a. What are the initial impressions of patients regarding their immersive experience within the VR intervention?

1b. What are the initial impressions of healthcare providers regarding the dashboard and possibilities of the VR intervention?

2. To what degree do patients report changes in their psychological distress levels during VR immersion?

3a. What critical factors should be considered when introducing VR in treatment, according to patients?

3b. What critical factors should be considered when introducing VR in treatment, according to healthcare providers?

2 Materials and methods

2.1 Study setting

This study focused on investigating the first impressions of patients and healthcare providers regarding a virtual reality intervention in two Dutch forensic mental healthcare organizations: Transfore and De Waag. Both organizations provide treatment for aggression regulation and sexually transgressive behavior to forensic psychiatric patients who either committed or are at risk of committing a criminal offense due to psychiatric problems. Transfore has multiple treatment locations in the east of The Netherlands and offers treatment to over 1,500 in-and-out patients every year. De Waag is an outpatient clinic with 12 treatment sites throughout The Netherlands. They offer treatment to around 7,000 patients a year. Ethical approval was obtained from the Ethics Committee of the University of Twente (Behavioral, Management, and Social Sciences, number 210645). This qualitative study adheres to the consolidated criteria for reporting qualitative research (COREQ) (Tong et al., 2007).

2.2 The intervention

The interactive VR intervention that was applied in the current study is called 'Triggers & Helpers'. The VR software was developed by CleVR. The patient wears a head-mounted display and noise-canceling headphones. While being immersed in a VR scenario, the patient can walk through a broad range of virtual environments such as a supermarket, a shopping street, or a home environment, using a controller. Additionally, the patient can conduct a role-play with virtual characters. This character is portrayed by the healthcare provider using a voice-morphing microphone. The healthcare provider can assume a broad range of virtual characters with different types of voices, allowing for a highly personalized experience. They can control the movements, facial expressions, and body language of the character using a dashboard (see Figure 1). Here, they can also enable changes in environments, such as increasing the number of passers-by on a street or characters that enter a virtual room during the scenario. In Figure 2, the setup of the VR technology is displayed. With this technology, a personalized VR scenario can be developed for different types of patient needs. In Figures 3, 4, screenshots of two virtual environments of the application "Triggers & Helpers" by CleVR are provided.

2.3 Participants

Interviews were conducted with patients and healthcare providers. Inclusion criteria for patients were: they had (1) to be fully informed about the study and willing to participate voluntarily, (2) no prior experience with the VR intervention "Triggers & Helpers" to elicit their first impressions, (3) followed some form of aggression regulation treatment in an out-patient setting, and (4) approval from their healthcare provider, who indicated that the immersion in VR would not be uncomfortable or damaging for the patient or their treatment goals. For safety reasons, their healthcare provider was present during the VR experience and interview. For healthcare providers, inclusion criteria were that they were (1) currently working in forensic mental healthcare, (2) involved in any type of aggression regulation treatment for forensic outpatients, and (3) potential end-users of the VR intervention "Triggers & Helpers". Patients and healthcare providers who fail to meet the inclusion criteria will be excluded from the study. For the patient group, (1) individuals with epilepsy, (2) severe dizziness, or (3) severe visual impairments, as well as those experiencing (4) active psychosis or (5) another form of crisis as assessed by their healthcare provider, will be excluded from the study. These measures have been implemented to ensure the safety and well-being of participants during the VR immersion.

The recruitment of participants was carried out by convenience sampling, a nonrandom sampling method where members of the target population meet certain practical criteria, such as easy accessibility to the researcher, availability at a given time, or willingness to participate (Dornyei, 2007; Given and Lisa, 2008). Suitable healthcare providers and patients were identified by the project team. This team consisted of two researchers, three healthcare providers, two former patients, and one policy officer working on the development and implementation of the VR intervention. The team identified a list of 10 potential healthcare providers working in forensic mental healthcare organizations that fit the inclusion criteria. The team paid attention to including a broad range of participants, working in different organizations, and working with different aggression regulation treatment groups to ensure that a variety of healthcare provider perspectives were included. These healthcare providers were approached by email by the researcher (MK) with the request to participate in an interview and all agreed. Next, these healthcare providers were asked to select suitable patients out of their caseload, inform them about the study face-to-face, and ask if they wanted to participate. Patients were intentionally selected from two distinct forensic mental healthcare organizations, representing a range of age groups, and participating in different types of aggression regulation treatment: individuals from a 'regular' aggression regulation group

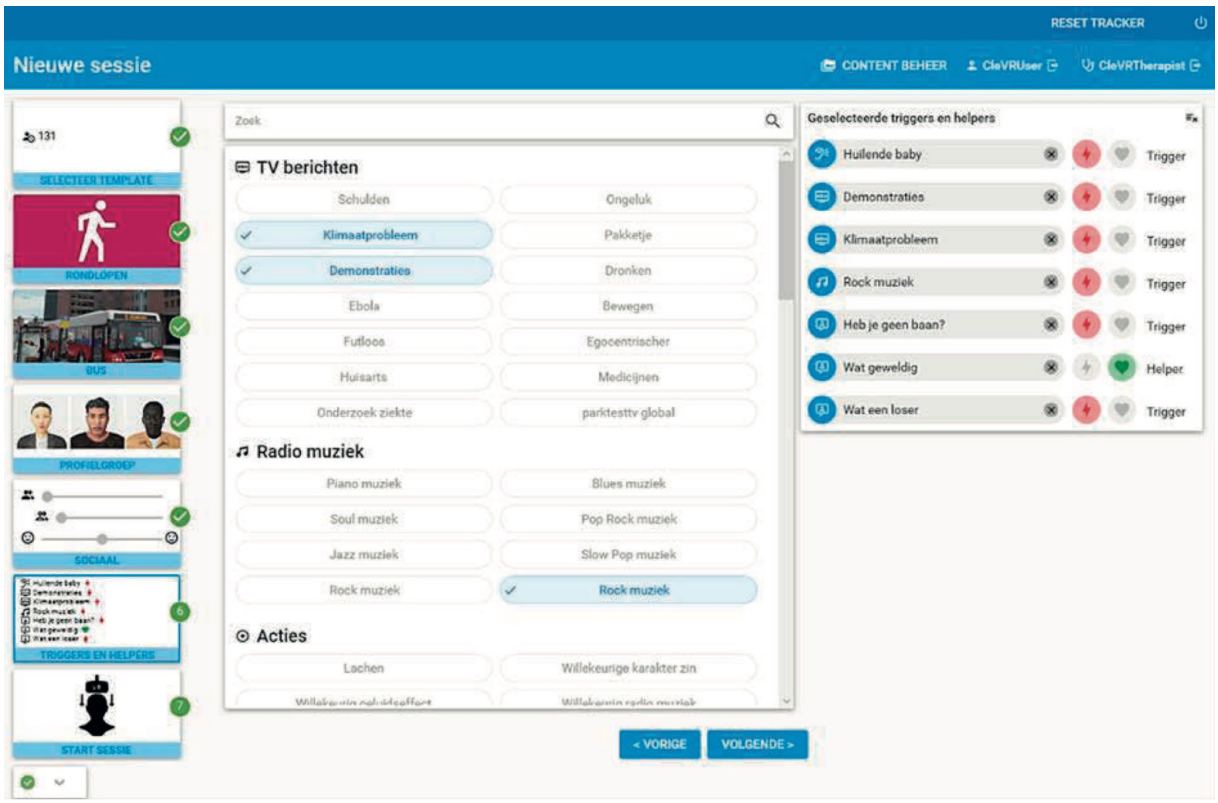


FIGURE 1
Dashboard of the VR intervention "Triggers & Helpers".

and those from a group tailored for individuals with mild intellectual disabilities. This deliberate inclusion of participants from diverse backgrounds and varying intellectual abilities ensured a comprehensive representation of potential end-users of the VR application. In total, 8 patients were approached by their healthcare provider, and all agreed to participate in an interview. No approached patients or healthcare providers declined participation or dropped out during the study. A total of 18 interviews were conducted by one researcher (MK) between March and July of 2021. Participant characteristics are presented in Table 1.

2.4 Data collection

Data were gathered through in-depth, semi-structured interviews by one researcher (MK). Semi-structured interviews refer to a qualitative research method that is conducted with one respondent at a time, employing open-ended questions, often accompanied by follow-up 'why' and 'how' probing questions. This dialogue allows to delve into unforeseen and important issues for the research topic (Adams, 2015). Semi-structured interviews are often used in exploratory research when not much is known about a specific topic and are especially suited when multiple interview questions require follow-up queries in the form of probing questions that ask about the independent thoughts of each participant (Adams, 2015).

Before the interviews took place, patients and healthcare providers were informed by the researcher (MK) about the goal of the interview, the reasons for and interests in the research topic, and signed the informed consent form. In addition, patients were immersed for the first time in different virtual environments and had the opportunity to explore these environments by walking through them or talking to virtual characters in a role-play setting. While patients were immersed in the VR intervention itself, healthcare providers were informed by the researcher about the possibilities of VR and viewed screenshots of all available virtual environments, characters, and the dashboard before the interview took place.

The interview schedule for patients consisted of 7 open-ended questions with accompanying sub- and probing questions. A pilot interview was conducted by one researcher (MK) to refine the schedule and improve questions whenever necessary. The interview questions centered around patients' first impressions of VR and any suggestions for improvement: *"What did you think of the VR experience?"*. Specific attention was focused on patients' experience of psychological distress during the VR scenarios, this was asked during the interviews; *"On a scale of 1 to 10, with 1 being completely relaxed and 10 experiencing extreme stress, how high was your level of distress?"*. During the interview, patients were also queried about their perceptions of using VR in treatment and any important considerations for its introduction into practice. Sample questions include: "*To what extent would you like to use this VR application in treatment?"* and *"How would you apply this VR application in treatment?"*. Additional probing questions were asked to elicit an extra level of detail via verbal prompts to clarify, elaborate, or explain a prior answer to an interview question that the participant had already given. The patient interviews, after the VR immersion, took place face-to-face at the forensic hospital and lasted an average of 23 min.

FIGURE 2
Setup of VR consisted of laptop **(A)**, tablet with dashboard of "Triggers & Helpers" application **(B)**, voice-morphing microphone **(C)**, VR head-mounted display **(D)**, VR controllers **(E)**, noise-canceling headphones **(F)**.

FIGURE 3
Screenshot of a virtual living room in which the patient can perform a role-play with a virtual character (©CleVR).

The interview schedule for healthcare providers consisted of 9 open-ended questions with accompanying sub- and probing questions. An additional pilot interview was conducted. The interview questions focused on the first impressions of healthcare providers of the VR dashboard and opportunities in treatment. Additionally, any points for improvement for the VR system were discussed. A sample question: *"What is your initial impression of using the VR dashboard?"*. Furthermore, healthcare providers were asked about their thoughts on

FIGURE 4
Screenshot of a virtual shopping street in which the patient can walk around using a controller or can perform a role-play with multiple virtual characters (©CleVR).

TABLE 1 Participant characteristics.

Characteristics	Patients – *N* (%)	Healthcare providers – *N* (%)
Gender		
Male	8 (100)	2 (20)
Female	0 (0)	8 (80)
Age		
20–29 y	1 (12,5)	
30–39 y	4 (50)	
40–49 y	2 (25)	
> 50 y	1 (12,5)	
Treatment type		
Aggression Regulation group	6 (75)	
Aggression Regulation group for mild intellectual disabilities	2 (25)	
Function		
GZ psychologist		6 (60)
Forensic nurse		2 (20)
Occupational therapist		1 (10)
Forensic remedial educationalist		1 (10)
Prior experience with VR		
Used VR in treatment	0 (0)	8 (80)
Never used VR in treatment	8 (100)	2 (20)

the use of VR in treatment and any points of attention for the introduction of VR in their work practice, for example: "*How would you apply this VR application in treatment?*", and "*What would that look like in practice? Can you give some examples?*". Additional probing questions were asked in the healthcare provider interviews as well. The interviews with healthcare providers took place via Zoom, an online meeting program, due to the worldwide Covid pandemic that limited treatment on site. These interviews lasted an average of 47 min. Both types of interviews were recorded with a voice recorder. The interview schedules for patients and healthcare providers are provided in Tables A1, A2.

2.5 Data analysis

The audio recordings of the semi-structured interviews were transcribed verbatim and anonymized by one researcher (MK). To this qualitative data, thematic analysis was employed. This analysis

provides an accessible and systematic procedure for generating codes and themes from the qualitative data (Clarke and Braun, 2017). A coding scheme was iteratively created by one researcher (MK), while another researcher (HK) remained consistently engaged in the process, providing continuous oversight. First, the transcripts were read carefully to familiarize with the data and identify meaningful fragments related to one of the research questions. These fragments were linked to codes that captured interesting features of the data, relevant to the research questions. These different codes were building blocks for themes, a larger pattern of meaning. These themes and related codes were summarized in a coding scheme. The codes were selected by the method of constant comparison between interview fragments (Glaser, 1965; Boeije, 2002). The coding scheme was adapted throughout this process. This first version of the coding scheme was thoroughly discussed with another researcher (HK) and an improved version was used to code the other transcripts. The transcripts were coded and compared until the saturation point, at which no new codes relevant to the research questions were identified in the data (Glaser, 1965; Boeije, 2002). One researcher (MK) coded the fragments and discussed them with another researcher (HK) in case of any doubt. Definitions of codes were adapted throughout the process.

3 Results

3.1 First impressions of VR according to patients

To answer the first research question, the initial impressions of patients regarding their immersive experience within the VR intervention are explored. This category refers to the judgments or perceptions of patients while being immersed in VR for the first time. The related codes mentioned by patients are reported and defined in Table 2.

3.1.1 Positive first impressions

All patients mentioned positive first impressions regarding VR. Their first experience was regarded as fun, increasing curiosity, creating possibilities, and overall experiencing *feelings of enjoyment*. A few patients described the VR intervention as a great, innovative technology that creates possibilities to practice new behavior in real-life situations that can be re-created in a virtual world. Patients mentioned that beforehand, they did not know what to expect from the intervention. However, they were excited and curious to try it out; *"I did not know what to expect. I never experienced it before, but I was curious, and it was awesome. (…) It's good to practice in VR how to find peace again. I would highly recommend this."* (P8)

Most patients were positively surprised by how easy it was to recreate situations with the healthcare provider and experience them in VR, in contrast to only talking about an experience during treatment. To illustrate; "*It's great that you can make things clearer with this. Explaining experiences [face-to-face] is more difficult than showing it [in VR]."* (P5)

Regarding the *sense of presence*, the feeling of being fully engaged and immersed in the virtual environment, most patients described that they felt as if they were walking through the virtual environment and talking to a virtual character. Some patients described feeling present as if being present in a game; *"Physically you are in the [treatment] room, but mentally you are in VR. It's a strange feeling. It really can be compared to a game"* (P3). Additionally, the vast majority of patients described heightened alertness and an increased sense of situational awareness. Patients shared that the unfamiliar environment prompted them to prepare themselves to act upon unexpected situations. For instance, there was a sense of curiosity among patients about the potential outcomes when they encountered virtual characters that initiated conversations. One patient illustrated:

> *"I have to be alert to all the people walking by. For example, when suddenly a man comes very close to me and unexpectedly takes his phone out of his pocket, then I notice for myself that I become very alert to these small movements. That's something from my past."* (P8)

A factor that influenced the feeling of presence was the level of experienced realism. Aspects that contributed to this feeling were details in the environment; *"I thought the environment was portrayed quite realistically. I once saw a garbage can or an air conditioner hanging*

TABLE 2 First impressions of VR according to patients.

Code	Definition	N*int*[a]	N*tot*[b]
Positive first impressions			
Feelings of enjoyment	Situations or interactions that bring emotions of joy, satisfaction, curiosity, or excitement to patients while using VR.	8	18
Sense of presence	Subjective experiences of the extent to which patients were fully engaged and immersed in a virtual environment.	7	16
Points of improvement			
Lack of sense of presence	Lack of subjective experiences of the extent to which patients we not fully engaged and immersed in a virtual environment.	7	10
Unnatural movement	Aspects that could be improved related to navigating within a virtual environment, allowing users to explore and interact with this environment.	7	9
Feelings of discomfort	Unpleasant sensations experienced by patients while being immersed in VR, such as dizziness or increased body temperature.	8	8

[a]The number of interviews the quote was mentioned in (N*int*).
[b]The number of times the code was mentioned in all interviews with patients (N*tot*).

in the corner at the shopping street. Those are details I pay attention to." (P1). In addition, according to some patients, the feeling of realism increased when the virtual characters started an interaction with the patient. Patients had the feeling that they had to react, had to shake their hand, or had to step aside when a character approached. As explained by some patients; They know that it is a simulated environment, however, when immersed in VR, they are forced to act upon the virtual situation; *"It feels different. You know it's fake, but it looks realistic. Your brain will believe that you are in VR. It looks real."* (P5)

3.1.2 Points of improvement

Patients provided both, positive and negative first impressions of VR during the study. While some patients had predominantly positive experiences regarding their sense of presence, others highlighted areas for improvement in this matter.

A frequently mentioned concern among patients centered on the *lack of a sense of presence* during instances where they focused on the details in the virtual environment. For instance, the sense of presence decreased when patients directed their attention toward virtual characters. One patient expressed, "*I do not experience it as real; it feels unrealistic. There is still very little feeling or emotion [while conversing with a virtual character]. So, I cannot see if he means what he says. It is now very fake and superficial.*" (P7). Another aspect of VR that decreased the sense of presence was related to the appearance and movement of virtual characters. Some patients found the virtual characters to be lacking in emotional expression, describing their body language, appearance, and physical movements as robotic or unnatural. One patient elaborated, *"Now the person facing you is still robotic. It must seem somewhat real to use for people with aggression problems. The movements and appearance feel very unnatural."* (P2). This unrealistic portrayal of characters made it challenging for patients to fully engage in interactions with them. Despite this, some patients acknowledged that these VR characters could still be used to simulate realistic situations since they felt like they needed to respond in a conversation when a character talked to them. Finally, patients expressed a desire for more realistic details in both the characters and environment to increase the sense of presence. One patient wished for a higher level of fidelity and accuracy in the visual and auditory aspects of the virtual environment. They suggested, *"It would be nice if you could also look in the store through the window or hear some more background noises. Maybe you can walk everywhere and hear sounds from the houses above or birds flying and chirping above you."* (P6)

An additional negative first impression was related to the *unnatural movement* in VR that was experienced by patients while walking through an environment. They experienced the slow walking pace as irritating and unnatural. The slow movements did not match the physical movement patients would be able to execute in real life. They described this as hindering the feeling of being completely immersed in the virtual world.

> *"It's different. You stand still yourself [in real-life], but it feels like I'm walking [in VR]. Everything I see has to be processed by my brain. So basically, my brain is being fooled and that's why it feels so weird in my body. The reality does not match and that makes me dizzy for a while."* (P7)

The unnatural movement experienced by patients increased the *feelings of discomfort* for four patients in VR. They mentioned that they experienced dizziness during VR because of it. As one patient illustrated:

> *"It's different. You stand still yourself [in real-life], but it feels like I'm walking [in VR]. Everything I see has to be processed by my brain. So basically, my brain is being fooled and that's why it feels so weird in my body. The reality doesn't match and that makes me dizzy for a while."* (P7)

Additional aspects that increased discomfort were increased body temperature during the VR session. This was mentioned by one patient who explained that this was because of the feeling of being "enclosed" by the head-mounted display and the noise-canceling headphone.

Most patients agreed on these points of improvement, however, two patients did not have any specific suggestions for improvement for the current version of the VR intervention.

3.2 First impressions of VR according to healthcare providers

To answer the first research question regarding the first impressions of healthcare providers, their initial judgments and perceptions regarding the current version of the VR dashboard were explored. The codes mentioned by healthcare providers are reported and defined in Table 3.

TABLE 3 First impressions of VR according to healthcare providers.

Code	Definition	N*int*[a]	N*tot*[b]
Personalization	The customization and adaptation of VR scenarios to suit the preferences, needs, and characteristics of individual patients.	5	7
User-friendly	The ease of use, accessibility, and intuitive nature of the VR software.	3	3
Preview of VR scenario	A brief demonstration or glimpse that is provided to the users that shows the VR environment and virtual characters.	2	3
Variety in environments	The presence of diverse and suitable settings in which a scenario can take place within the VR software.	1	1
Wish for more realistic detail	The desire to have a higher level of fidelity and accuracy in the visual and auditory aspects of the virtual environment.	1	1

[a]The number of interviews the quote was mentioned in (N*int*).
[b]The number of times the code was mentioned in all interviews with healthcare providers (N*tot*).

3.2.1 Positive first impressions

Overall, healthcare providers positively evaluated the great variety of options to create and *personalize* a VR scenario, such as the extended list of VR characters and environments. The software was reviewed as *user-friendly*, providing the possibility to create a VR scenario with a structured, step-by-step approach. As one healthcare provider illustrated; *"I find it useful that you can see on the left [of the dashboard] at which step you are and what the next step is. All the expansions are nice. This allows more variations to be made in the VR characters. That is nice for patients to be able to personalize it. I find it user-friendly. It's easy and even for me it's doable [to set up a VR scenario]."* (H8). Healthcare providers appreciated the option to go back and forward between the steps to adjust the VR scenario to fit with the treatment goal. To illustrate:

> *"My first impression is that it is user-friendly. It is useful and nice that it [setting up a VR scenario] goes step by step. I am surprised about all the options you can choose from. It's super comprehensive. I don't miss anything in terms of environments."* (H9)

3.2.2 Points of improvement

In addition to positive first impressions, healthcare providers mentioned some points of improvement for the VR intervention. Some mentioned that it would be valuable to have the option to see a *preview of the VR scenario* before actually immersing patients in the VR environment. They mentioned that it can be difficult to imagine what the scenario would look like for the patient when being submersed in VR;

> *"It would be nice if you as a healthcare provider could see a concept of the VR scenario, perhaps by clicking on a special button at the last step, to see what that session will really look like in VR. Then you can see, for example, that the police officer is placed over there and the cashier is really behind the cash register. That you can see a preview of the session that you created before you click 'play'. Then you can easily adjust if something isn't right yet."* (H6)

Regarding the VR environments, most healthcare providers appreciated the variety of options. However, a healthcare provider did mention that it would be great if there would be more *variety in environments*, besides an office space, since most patients work in more physical workplaces, such as construction sites. A variety in characters was already achieved and appreciated. However, healthcare providers mentioned that a filter to this extensive list of VR characters would make it easier to personalize a VR scenario. For example, find a fitting character for a role play, such as a police officer or an older gentleman. Adding a filter for the characters' profession, age, gender, or length was mentioned.

Additionally, the virtual living room and kitchen were regarded as too clean and neat. According to one healthcare provider, this would not match the realistic living situation of most patients. They would prefer *more realistic details*, such as a living room that is messier and less clean. As someone illustrated; *"What I did notice is that the VR environments look too neat. It should be a bit messier to be realistic."* (H10)

3.3 Subjective psychological distress

To answer the second research question, the experienced psychological distress of patients during a VR session is explored.

TABLE 4 Subjective psychological distress on a scale from 1 to 10[a].

Patient nr.	Task 1	Task 2	Task 3	Task 4
1	2	2	2	3
2	1	1	2	2
3	2	2	2	2
4	1	1	1	1
5	6	5	5	6
6	6	7	8	8
7	1	2	2	4
8	2	1	6	6

[a]A score of 1 indicates that the patient was being completely relaxed and a score of 10 indicates extreme stress experienced by the patient.

While patients were immersed in different VR scenarios, they were asked to rate their subjective psychological distress on a scale from 1 to 10, with 1 being completely relaxed and 10 experiencing extreme stress. For an overview of their experienced subjective psychological distress during the VR session, see Table 4.

In general, patients mentioned that their experienced distress was relatively low during the start of the VR session. Six out of eight patients mentioned that their distress level was 2 out of 10 or even lower at the start of the VR session. The other two patients rated their distress level with a 6 as a starting point. Most patients mentioned that this psychological distress did not increase further during the VR session. However, two patients reported a notable increase in distress during interactions with the virtual environment. These patients mentioned an increase from level 1 to level 4 or 6 during this interaction. For example, patients mentioned an increase in distress when a virtual character stood in front of them and started moving or talking. They explained this increase by having the feeling that they had to be more alert and were forced to react and deal with the situation. As illustrated by a patient; *"The stress is a bit higher because someone [a virtual character] is facing me. I know it is not real, but I still have the feeling that I have to deal with him."* (P7)

Additionally, patients did not know what to expect during the VR session. This experience of uncertainty increased distress and forced patients to be extremely attentive to the situation and their behavior. One patient illustrated the consequence of this increase:

> *"It forces me to react differently than I normally would. Normally I would run away, but now I have to stay and keep calm. I have to go along with a situation that is unfamiliar to me, that I actually don't feel comfortable with at the time. Someone who suddenly stands so close to me and makes unexpected movements. Those kinds of moments actually only occur in a bad dream, when you completely freeze. That's what actually happens."* (P8)

3.4 Possibilities of VR in treatment according to patients

To answer the third research question on critical factors to consider when introducing VR in treatment, the opinions and preferences of patients on if and how VR could be applied in treatment are provided. The viewpoint of patients regarding their willingness to use VR and the potential applications of VR play an important role in the introduction of VR in practice. By exploring their opinion, it

TABLE 5 Possibilities of VR in treatment according to patients.

Code	Definition	N*int*[a]	N*tot*[b]
Willingness to use VR	Indication of the extent to which patients are ready and willing to use VR in their own treatment.	5	7
How to use VR in treatment			
Possibility of reflection	The capacity to analyze and evaluate one's behavior and responses in VR scenarios that recreate situations experienced by patients in the real world.	8	10
Triggering fear	The intentional use of VR scenarios to elicit fear or a sense of anxiety in individuals by exposing them to simulated situations or stimuli that evoke a fearful response.	3	3
Triggering aggression	The intentional use of VR scenarios to elicit aggressive thoughts, emotions, or behaviors in individuals by exposing them to simulated situations or stimulate that evoke aggression.	3	3

[a]The number of interviews the quote was mentioned in (N*int*).
[b]The number of times the code was mentioned in all interviews with patients (N*tot*).

becomes possible to identify key areas where VR can be most effectively applied, as a first step in the integration into treatment. The codes mentioned by patients are reported and defined in Table 5.

3.4.1 Willingness to use VR in treatment

The opinions of patients regarding their willingness to use VR in their treatment differed. Five out of eight patients were willing to use VR in their own treatment. They were excited to use new, innovative technology. Some patients mentioned that they would also recommend this technology as a treatment tool to other patients; *"Yes, I would definitely like to use this. I would really recommend it to others as well. (…) You can actually see the effects immediately. Thus, treatment-oriented, I think VR could be more effective [than traditional treatment]"* (Skip et al., 2018).

However, three out of eight patients mentioned that they would not prefer to use VR. They also mentioned that this form of technology would not be suitable for all patients. Firstly, a patient mentioned not being willing to use VR because he felt a lack of technical skills and experience with innovative technology created a barrier. He expressed a clear preference for practicing behavior in real life; *"I prefer to have a real person in front of me. I cannot imagine talking to a virtual person. It just does not feel real. I know this is fake. I find digital and virtual communication harder to understand."* (P7). Another patient agreed, he described difficulty in acting in VR as if it were a real situation, feeling 'insensitive' to the virtual scenarios. However, the patient expressed that others might feel more receptive toward VR.

3.4.2 Treatment possibilities with VR

While three out of eight patients did not prefer to use VR in their treatment, all patients mentioned several opportunities on how VR could be applied in treatment. Patients had a clear idea of the possibility of VR to recreate real-life experiences and reflect on these experiences together with a healthcare provider. They explained that VR provides the possibility for healthcare providers to see what they normally cannot, creating an important *possibility of reflection* in real-time. One patient illustrated:

> *"It's nice that it's possible to recreate my experiences or what I'm going through in daily life [in VR] and then reflect on that. If you simulate my experiences in the [VR] system, the healthcare providers can see what they normally don't see. You can tell that you fought with someone yesterday, and tell them exactly what happened, but then it is still a guess for that person how it really went. [In VR] you can simulate that situation together and reflect on it."* (P7)

Patients mentioned two concrete examples using VR as a tool for exposure in treatment. Firstly, VR can be implemented in the treatment of patients with agoraphobia, patients who feel insecure when interacting with others, or any other kind of *fear that patients can be exposed to* in VR. They expressed that a triggering or fear-inducing situation will be easier to experience step-by-step in the safe environment of the virtual world because VR allows patients to remove themselves from the virtual scenario whenever their distress rises too high. They can be willing to practice this in the controlled environment of VR, however, they might be too afraid to practice this in real life. One patient illustrated;

> *"You could use it for role play. For example, if someone is afraid of something, you can expose them to it in a controlled way. I do think that could be useful. You can make someone take a step forward that they might be too afraid of or too shy to do in real-life."* (P6)

Second, according to patients, VR can be used for patients who display aggressive behavior toward other people. Patients might be *exposed to situations that trigger their aggression* and, they can react to this in a safe, controlled, virtual environment. As a patient expressed: "*If a patient wants to hit someone, at least in VR they hit the air.*" (P4).

3.5 Possibilities of VR in treatment according to healthcare providers

To answer the third research question on critical factors to consider when introducing VR in treatment, the opinions and preferences of healthcare providers on how VR could be applied in treatment are provided. The viewpoint of healthcare providers regarding potential applications of VR plays a crucial role in the introduction of VR in practice. By exploring the possibilities and added value of VR, these key areas can be used as a starting point for the introduction of the technology in practice. The codes mentioned by healthcare providers are reported and defined in Table 6.

TABLE 6 Possibilities of VR in treatment according to healthcare providers.

Code	Definition	N*int*[a]	N*tot*[b]
Treatment possibilities with VR			
Observation and assessment	The application of VR to evaluate potential risk factors of patients related to criminal behavior or mental health disorders.	4	4
Practice new behavior and copings skills	The application of VR as a tool for patients to simulate and engage in realistic scenarios where they can practice and refine desired behavior and coping strategies.	3	3
Exposure	The application of VR to simulate and expose individuals to fear- or aggression provoking situations in a controlled and safe environment.	2	2
Integrating VR into daily practice			
Expectation management	The process of informing and setting realistic expectations for patients undergoing VR treatment.	3	3
Existing treatment protocols	Established and standardized procedures, guidelines, or frameworks that healthcare providers follow when providing treatment to patients and that could benefit from VR.	3	3
Part of existing treatment	Integrating VR as a complementary tool within existing treatment frameworks.	1	2
Suitable for all	VR treatment is appropriate and safe for a wide range of patients, without specific exclusion criteria based on their characteristics or conditions.	2	2
Added value	Potential benefits, or positive outcomes that VR can offer patient's treatment or the therapeutic process, that need to be discussed before the decision on whether a patient can use VR or not.	1	1
Deliberate choice	Conscious and thoughtful decision made about using VR in treatment after careful consideration and discussion within the team of healthcare providers and with the patient.	1	1
Different roles	Distinct responsibilities and contributions that healthcare providers assume within VR treatment.	1	1
Implementation materials and activities			
Materials	Resources, documents, or tools that are developed ant utilized to support the integration of VR into treatment.	6	7
Attention to the possibility of VR	Deliberately considering and exploring the potential applications, benefits and implications of incorporating VR in treatment.	3	3
Intervision groups	Structured and collaborative peer support groups where healthcare providers engage in discussion, reflection, and learning from each other's experiences in their clinical practice.	2	2
Training sessions	Organized and structured activities designed to provide healthcare providers with knowledge, skills, and competencies necessary to effectively and safely use VR technology in their practice.	1	2
Available time	Allocated timeframe or duration in which healthcare providers need to learn how to use VR and practice with the technology.	1	1
Templates	Pre-designed formats or exercises that serve as a starting point or framework for creating VR scenarios.	1	1
Indication criteria			
Suitable for all	VR treatment is appropriate and safe for a wide range of patients, without specific exclusion criteria based on their characteristics or conditions.	2	2
Added value	Potential benefits, or positive outcomes that VR can offer patient's treatment or the therapeutic process, that need to be discussed before the decision on whether a patient can use VR or not.	1	1

[a]The number of interviews the quote was mentioned in (N*int*).
[b]The number of times the code was mentioned in all interviews with healthcare providers (N*tot*).

3.5.1 Treatment possibilities with VR

Healthcare providers described three ways to use VR in treatment. Firstly, VR can be used as an *observation- and assessment tool*. With VR, insight into patients' experiences and behavior can be gained. While exposing patients to a virtual scenario, personal triggers or risk factors can be discovered; *"I think you can use it for the initial phase and discover personal triggers."* (H7)

Secondly, VR can be used to *practice new behaviors and coping strategies*. Healthcare providers mentioned specific applications. For example, when patients are exposed to a challenging situation, they

can practice applying stress reduction techniques, and relaxation exercises; *"I am working with a client on stress reduction. We are looking at what causes him stress, for example, big crowds or loud music around him. With all those triggers I can simulate a scenario in which we can practice relaxation exercises."* (H4). In addition, patients can practice conflict management skills in a role-play that triggers their aggressive behavior. As illustrated by a healthcare provider:

> *"You could use it for aggression problems. I think most people will start role-playing. For example, when clients feel threatened on the subway if someone stares at them. With aggression, it is often useful to practice management skills, conflict management."* (H8)

Finally, VR can be used for *exposure*. Patients can be exposed to situations that trigger their aggressive behavior or personal fear. Within VR, a behavioral experiment can be set up and the healthcare providers and patient can reflect on how the patient acts and feels during the experiment. One healthcare provider illustrated; *"You can use it in the exploratory phase, so to discover signals and triggers. After, you can use it as a replacement for the exposure. You can move much faster in the VR world [than in real-life] to apply exposure and set up more behavioral experiments as a result."* (H1).

3.5.2 Integrating VR into treatment practice

As can be seen in Table 6, healthcare providers advised on how to embed new VR technology into daily work practice. Several points of attention were mentioned. First, a focus on *expectation management*. According to healthcare providers, patients should be informed about the use of VR early on in their treatment. They need information about the technology itself, the goal, the possibilities, and the working mechanisms. In addition, patients must be informed that they must provide input on personal experiences, triggers, and situations for VR sessions. They should have the opportunity to think along with their healthcare provider and discuss their expectations; *"I think they should slowly get used to VR. I think you should explain very well that you first practice with them and go through it together: What can you expect? That they have a bit of an idea."* (H5).

Second, healthcare providers mentioned cognitive behavioral therapy and aggression regulation treatment specifically as *existing treatment protocols* to which VR could be a valuable addition. Healthcare providers mentioned that it is important to see VR sessions as an intervention that is used as *part of the existing treatment* plan, not as an independent form of treatment. As one healthcare provider illustrated:

> *"It is best to embed it in an existing treatment. I think patients can get used to it slowly. You should explain it very well and go through it together: What it is, what is possible. That they [patients] have an idea and that they can also come up with a situation that you can practice [in VR]".* (H5)

Healthcare providers addressed indication criteria; patient-related factors that need to be adhered to before patients can use VR in their treatment. Two healthcare providers mentioned that this VR technology could be *suitable for all patients*, since all patients have their own triggers, e.g., a difficult home situation, complex family relationships, or problems in a work environment. Because of the wide variety of possibilities in VR, it applies to most patients' treatment goals. However, multiple healthcare providers mentioned that VR should not be applied as a standard treatment tool. The *added value* of VR for each and every patient must be discussed with the team of healthcare providers and with the patient. Only if the added value for the patient's treatment goals is to be expected and it fits with the current treatment plan, VR should be part of their treatment. One healthcare provider illustrated: *"I would not necessarily say in advance that we are going to use VR as standard for everyone. I think that it is very important to look at each and every patient to see whether it fits the treatment goal. Because, for example, with schema therapy treatments you may not need it."* (H1)

They emphasized that the use of VR should be a *deliberate choice* between the patient and the healthcare provider. The option to use VR should be discussed before treatment starts, at the beginning of the treatment process. The option to use VR and integrate it into a treatment plan should be discussed within a team meeting before being offered as an option to the patient. To illustrate:

> *"Pay attention to the option to use VR from the start of the treatment. It is important that it is discussed in a team meeting, while the treatment plan is discussed. I think that it is important that patients are informed about this early on in the treatment. That there is a possibility to use VR. Then you can think about it together with the patient."* (H2)

Lastly, healthcare providers described that it is important to pay attention to the new skills that are needed to implement VR in the existing workflow. Healthcare providers need to gain skills in *different roles* in order to use VR. They have to create and control the VR session, they have to pay attention to the patient, their emotions, and behavior, all while keeping the treatment goals in mind during the session. As one healthcare provider illustrated, this can increase the threshold for use; *"What I find a barrier in use, is that you have to do a lot. You have to pay attention to the client, control the VR session and you also have to think about what we are going to do."* (H10).

3.5.3 Implementation materials and activities

A thorough implementation was identified by healthcare providers as an important starting point for working with VR. To help with the integration of VR into existing treatment, healthcare providers expressed the need for several *implementation materials*, activities, or strategies. An introductory video, information brochure, or an online psycho-education module were mentioned by a few healthcare professionals to inform patients and themselves about VR, its possibilities, and its added value. To illustrate; *"Yes, I think it would be a good idea to take a more creative approach, like a video or folder because then they [patients] might also be more stimulated and fascinated to use it. If they have to read a lot, then I think half of them will not do it. Maybe more than half."* (H4).

In addition to implementation materials, three healthcare providers mentioned that the threshold for use would lower whenever their colleagues and managers would pay more *attention to the possibility of using VR* in treatment. One healthcare provider illustrated; *"The managers have to motivate colleagues and say: Well, this is VR therapy and we think this is an important development. People need to be referred to VR as an option for treatment. It needs more attention so that people think; 'Oh, there is a VR set? Nice, I'm using it in my treatment plan."* (H3).

According to the participants, the most important aspect of the implementation of VR is not the materials that are available for patients or healthcare providers; it is the available time they have to practice together with colleagues on how to use VR technology and apply it to treatment. Healthcare providers mentioned that this would increase their self-confidence regarding the use of VR and lower the threshold for actual use:

> *"It just takes a lot of time, training, and practice. You will need to practice this in an intervision with colleagues, because it is mainly a lot of 'doing, doing, doing', before you have the self-confidence to say: 'Oh yes, it works, I can do it.', and actually apply it to treatment with patients. So, I think this is the biggest investment we have to make."* (H1)

As illustrated above, *intervision groups*, frequent meetings with experienced colleagues, were valued to discuss knowledge on VR, its potential, exchange ideas, address barriers, and share experiences in working with VR technology in treatment with patients. Investing time to practice and organize intervision sessions is seen as important.

In addition to practicing together with colleagues, official *training sessions* on the use of VR and how to set up scenarios were mentioned. In the training, example exercises or *templates* of frequently experienced triggers or situations could be discussed. For example, a role-play script on having an argument with your boss. These prompts could help healthcare providers in setting up a realistic VR scenario. As one healthcare provider mentioned; *"Certain templates that you make available, which are common situations [written down as a VR exercise] which therapists can then perform. That also gives something to hold on to during treatment."* (H2). To support healthcare providers in the use of VR, VR *experts* or coordinators were mentioned. These experienced colleagues can support healthcare providers who want to apply VR in their treatment by sharing experiences, supervising VR sessions, answering questions, or playing a motivational role.

4 Discussion

4.1 Principal findings

The study aimed to provide insight into patients' and healthcare providers' initial impressions and perspectives regarding the introduction of VR in forensic mental healthcare, to further guide its integration in clinical practice. Patients' first impressions of VR were predominantly positive, with feelings of enjoyment and curiosity being prominent. They often felt a strong sense of presence, feeling fully immersed in the virtual environment. However, there were also points for improvement noted. The patient's sense of presence decreased when they focused on unrealistic details in the environment, unnatural movement of the virtual characters, and feelings of discomfort, including dizziness or increased body temperature. Healthcare providers highlighted the importance of personalization in VR scenarios, allowing customization to fit individual patients' needs and treatment goals. They found the VR software user-friendly and intuitive, emphasizing the ease of use and various possibilities. Some providers expressed a desire for a preview of VR scenarios before immersion and wished for more realistic details in the visual and auditory aspects of the VR environment.

The psychological distress experienced by patients during the VR sessions varied, with most reporting low distress levels at the outset. However, some experienced increased distress during interactions with the virtual environment, particularly when virtual characters engaged with them, leading to feelings of alertness and the need to react. Additionally, the uncertainty of not knowing what to expect during the VR session heightened distress levels, requiring patients to be highly attentive.

For the introduction of VR in practice, patients and healthcare providers highlighted some key points of attention. Emphasis should be placed on the careful introduction of VR to patients, particularly regarding triggering scenarios. Additionally, they stressed the importance of expectation management and shared decision-making between patients and healthcare providers while discussing the option to use VR in treatment. Attention to implementation resources and activities, such as training sessions, intervision groups, and available time to practice with VR were noted for successful implementation. Ultimately, both patients and healthcare providers highlighted the potential benefits and specific applications of VR, such as VR as a tool for exposure to triggering scenarios, practicing new behaviors and coping strategies, or using VR for observation and assessment of risk factors. All while emphasizing its deliberate and well-managed integration into therapeutic practice.

4.2 Comparison with prior work

The study's main findings underscore the positive first impressions of both patients and healthcare providers concerning the potential and application of virtual reality (VR) in treatment. Patients appreciated VR for offering them a valuable opportunity to visually simulate personal experiences, practice new behavior and coping skills, and reflect on this together with their healthcare provider, something not feasible in traditional treatment. This is in line with recent research, where it is previously discussed that the potential of VR offers a unique opportunity to bridge the gap between the safe and controlled treatment environment and the external, often unreachable real world (Botella et al., 2017; Tuente et al., 2020). The potential of VR is evident for both patients and healthcare providers. Consequently, when introducing VR, these stakeholders require neither an extensive persuasion nor an educational campaign to appreciate VR's utility within the context of treatment. This initial recognition of the technology's applicability creates a favorable starting point for implementation (Kouijzer et al., 2023).

Beyond these predominantly optimistic evaluations of VR's benefits and opportunities, patients and healthcare providers indicated a preference for more realistic detail in VR environments and characters. In contrast to the wish for more realistic details for VR to be effective, there are examples of VR applications in healthcare practice that use an abstract environment and are very effective in achieving their goals. As an illustration, Ijsfontein, a software developer specializing in behavior change and learning, designed an application where patients with depression engage in roleplay using highly abstract VR characters. Testing the application with varying degrees of realism revealed that the highly abstract version had the most significant effect on patients. This underscores that an effective

application does not need to solely rely on visually realistic detail (Cornet et al., 2019). This is in line with the findings of the current study. Patients expressed a strong sense of presence, the psychological experience of "being there" (Cummings and Bailenson, 2016), even when immersed in VR environments that lacked a satisfactory level of realism according to most participants. Patients felt alert and compelled to respond to the situations in VR. This showed that patients experienced a sense of belonging with a body in the virtual world, emphasizing the process of embodiment and the sense of presence (Ventura et al., 2018).

However, this feeling of alertness also contributed to an increase in patient's psychological distress during interactions with the virtual environment. For example, most patients experienced more distress when a virtual character imitated speech or motion. This heightened distress was also mentioned to be linked to the uncertainty of the unfamiliar situation patients found themselves in when experiencing VR for the first time. The overarching elevation in psychological distress experienced by all patients can be explained by VR's capacity to evoke powerful emotional and psychological reactions (González Moraga et al., 2022). Within a therapeutic setting, patients engage with scenarios that provoke their problematic behavior, potentially leading to an even more pronounced impact on psychological distress (González Moraga et al., 2022). Prior investigations have demonstrated that psychological distress and emotional reactions can intensify when personal triggers are incorporated (Botella et al., 2017). This rise in psychological distress and heightened alertness underscores VR's potential as an instrument for behavior change. However, it also highlights the need for a cautious approach when introducing VR due to its potential intrusiveness, especially when patients are immersed in triggering scenarios. A gradual, incremental approach is advised, commencing with neutral scenarios devoid of visuals or auditory triggers and progressively incorporating triggers and interactions within the virtual scenario.

When introducing VR in treatment, the importance of expectation management should be highlighted. It is essential for healthcare providers to explain and show what patients can expect when they enter a virtual environment. In addition, they could emphasize when introducing VR to patients that while the virtual environment may not look hyper-realistic, it can create a strong sense of presence and generate the associated emotional response. This approach can help manage expectations and reduce potential implementation barriers, which may be largely unfounded, increasing the intention to use VR and successful adoption of the technology (Tamilmani et al., 2021).

Additionally, the decision to incorporate VR into patient treatment should be deliberate and well-considered, involving discussions with both the patient and the healthcare team. Managing expectations is essential, ensuring that patients have a comprehensive understanding of the role VR will play in their treatment and realistic expectations when engaging in VR during treatment sessions. This process of shared-decision making fosters a collaborative environment in which patients and healthcare providers jointly determine the most suitable treatment approach that aligns with patients' preferences and clinical needs, promoting patient-centered care and enhancing implementation efficiency and treatment adherence (Rogers, 1995; Chong et al., 2013; Kouijzer et al., 2023).

Despite the overarching positive reviews regarding VR, several critical factors require consideration when introducing the technology in practice. It is crucial for healthcare providers and patients to acquaint themselves with the new technology and have access to implementation resources and activities that provide the necessary skills and confidence to employ VR within treatment. This emphasis on skill development and confidence-building aligns with the concept of "trialability" as proposed by Rogers (1995), which underscores the importance of allowing potential adopters to experiment with technology before fully committing to its integration in practice. In this context, facilitating opportunities for healthcare providers and patients to explore VR functionalities, understand its potential benefits, and engage in hands-on experiences can significantly contribute to the successful implementation of VR in healthcare settings.

This study showed the importance of thorough implementation. However, generally, the emphasis on systematic, multi-level implementation is lacking concerning VR's application in healthcare (Levac et al., 2015; Birckhead et al., 2019; Kouijzer et al., 2023). The current study provides valuable initial insights into important aspects of the introduction of VR technology into treatment practice. It offers points of attention and improvement when introducing a new form of treatment to its end-users.

Yet it also highlights the need for further comprehensive research to fully explore and understand these multifaceted implementation factors. For future directions, placing a heightened emphasis on systematic implementation would be valuable. Adhering to a well-structured implementation framework can empower researchers and practitioners to ensure a comprehensive and well-planned implementation process (Kouijzer et al., 2023). Implementation frameworks such as the Consolidated Framework for Implementation Research (CFIR) (Damschroder et al., 2009) or the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework (Greenhalgh and Abimbola, 2019) could provide guidance in systematically assessing and addressing implementation challenges and considerations in the context of integrating VR in healthcare. This study serves as a starting point, emphasizing the ongoing journey of exploration required to ensure the effective and sustainable adoption of VR within a forensic healthcare setting.

4.3 Strengths and limitations

The strengths of this study lie in its involvement of both patients and healthcare providers, who contributed their initial impressions regarding the introduction of VR into practice. The involvement of end-users and a focus on user experience is especially important when new technology is used (Ventura et al., 2018). Moreover, the deliberate selection of participants from diverse backgrounds and with varying intellectual abilities ensured a comprehensive representation of potential end-users of the VR application. The inclusion of patients with diverse cognitive abilities is especially important, considering that technological interventions must be personalized to the specific needs and capacities of end-users for successful integration into treatment. The study results capture a broad spectrum of perspectives and experiences, providing an overview of essential elements for the introduction of new technology into clinical practice. It facilitates a more informed and efficient integration of VR, enhancing the likelihood of a safe and successful adoption and long-term use of VR, thereby augmenting its value for patients, healthcare providers, and

treatment outcomes (van Gemert-Pijnen et al., 2018; Kip et al., 2019a,b).

While it is beneficial for patients to be exposed to VR before the interview, a limitation of this study focuses on the VR immersion in this study not being intended as part of the patient's treatment. The exposure to VR was primarily meant to gather first impressions as opposed to being an integral part of their treatment program, but it could distort the overall perception of the outcomes as these focus on integrating VR into treatment. In addition, the exposure to VR within this study was comparatively brief, and although patients experienced VR immersion, their exposure was restricted to neutral scenarios devoid of personal triggers. This aspect might have limited the complete potential impact of VR on participants. Therefore, it is important to consider VR's long-term effects on treatment and the exploration of its potential beyond the scenarios investigated in this study. Undertaking a more comprehensive investigation, involving longitudinal evaluations of outcome measures, could furnish a more complete understanding of the enduring benefits and potential limitations of VR within the context of forensic psychiatry.

An additional limitation to consider is that patients and healthcare providers involved in this study all volunteered to participate. A selection bias might have taken place, resulting in participants that generally held a favorable attitude toward the integration of VR in treatment. This could restrict the generalizability of the findings. The study sample might not be representative of the broader population, and therefore, the results may not capture the full range of perspectives and experiences related to VR use in forensic healthcare. Therefore, the results should be interpreted with caution and placed in a broader perspective. It might prove beneficial to track a diverse group of healthcare providers and their patients as they engage with VR in practice when the technology is successfully implemented. This approach would yield a broader understanding of implementation barriers and points of attention in the use of VR.

In addition to the potential of VR in forensic mental healthcare, as is discussed in this study, the limitations of applying the technology in mental healthcare should also be considered as well. Ethical considerations, for example, particularly in discussions surrounding the exposure of vulnerable populations should be taken into account. VR can be used to provoke emotions of anger and aggression. When eliciting physical and verbal anger, it can be questioned to what extent the patient may be stimulated in eliciting these emotions. The strength of VR is that reality can be realistically simulated, but the possibly provoked intense emotions and aggressive behavior need to be taken into account when integrating technology into mental healthcare (Kip et al., 2024). Finally, as shown above, the generalizability of findings from VR implementation studies poses a challenge, emphasizing the importance of diverse sample populations in research to draw broader conclusions about the efficacy and impact of VR implementation interventions in mental healthcare. These considerations highlight the nuanced approach required when integrating VR technology into clinical practice, needing careful navigation of ethical, technical, clinical, and research-related challenges.

5 Conclusion

The integration of VR into forensic mental healthcare holds great potential for behavior change. However, its immersive characteristics also increase the chance of amplifying psychological distress. This emphasizes the need for caution when using VR– especially when a vulnerable patient group is subjected to triggering scenarios. This study advocates for a gradual introduction of the technology and provides valuable insights into key elements for this introduction in clinical practice. Personalized treatment plans, developed collaboratively between patients and healthcare providers, and shared decision-making in setting up and integrating VR in treatment are crucial for navigating the introduction of VR effectively. Healthcare providers require adequate training and support to confidently use VR, in which attention should be paid to managing expectations for patients and providing them with adequate support throughout the introduction and integration of VR in treatment. While these key elements are well-known and should be considered as standard practice, they are not always applied when integrating a new form of technology in treatment. It is important to take these elements into account when introducing technology, particularly when the technology is used as a powerful tool to change behavior in vulnerable patient populations. This study highlights that even the initial step of integrating VR into practice – the introduction phase – demands careful planning and a personalized approach. This underscores the need for ongoing refinement and a systematic approach to the overall implementation of VR. These efforts are crucial to fully realize its potential in clinical practice.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the University of Twente (Behavioral, Management, and Social Sciences, number 210645). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MK: Writing – original draft. HK: Writing – review & editing, Supervision. SK: Writing – review & editing. YB: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this study was provided by Stichting Vrienden van Oldenkotte (grant number: P2021-17 Transfore UT-FoReTech 2 Verder Vooruit). They had no role in the study design, collection, analysis, or interpretation of the data, writing the manuscript, or decision to submit the paper for publication.

Acknowledgments

We would like to acknowledge Thien Vuong for helping with the data collection of the patient interviews.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1284983/full#supplementary-material

References

Adams, W. C. (2015). "Conducting semi-structured interviews" in *Handbook of practical program evaluation*. *Fourth* ed. Eds. Newcomer K.E., Hatry, H.P., and Wholey, J.S. (Hoboken, New Jersey, USA: Wiley Blackwell), 492–505.

Baniasadi, T., Ayyoubzadeh, S. M., and Mohammadzadeh, N. (2020). Challenges and practical considerations in applying virtual reality in medical education and treatment. *Oman Med. J.* 35, e125–e110. doi: 10.5001/omj.2020.43

Birckhead, B., Khalil, C., Liu, X., Conovitz, S., Rizzo, A., Danovitch, I., et al. (2019). Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. *JMIR Mental Health* 6:e11973. doi: 10.2196/11973

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Qual. Quant.* 36, 391–409. doi: 10.1023/A:1020909529486

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., and Baños, R. (2017). Recent Progress in virtual reality exposure therapy for phobias: a systematic review. *Curr. Psychiatry Rep.* 19:42. doi: 10.1007/s11920-017-0788-4

Chong, W. W., Aslani, P., and Chen, T. F. (2013). Shared decision-making and interprofessional collaboration in mental healthcare: a qualitative study exploring perceptions of barriers and facilitators. *J. Interprof. Care* 27, 373–379. doi: 10.3109/13561820.2013.785503

Clarke, V., and Braun, V. (2017). Thematic analysis. *J. Positive Psychol.* 12, 297–298. doi: 10.1080/17439760.2016.1262613

Cornet, L. J. M., Den Besten, A. L., and van Gelder, J. L. (2019). *Virtual reality en augmented reality in Justitiële context*. Enschede, The Netherlands: University of Twente.

Cummings, J. J., and Bailenson, J. N. (2016). How immersive is enough? A Meta-analysis of the effect of immersive technology on user presence. *Media Psychol.* 19, 272–309. doi: 10.1080/15213269.2015.1015740

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., and Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* 4:50. doi: 10.1186/1748-5908-4-50

Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford, United Kingdom: Oxford University press.

Fassaert, T., Segeren, M., Grimbergen, C., Tuinebreijer, W., and De Wit, M. (2016). Violent offenders as a target population for public mental health care. *J. Forensic Leg. Med.* 40, 54–57. doi: 10.1016/j.jflm.2016.03.003

Fromberger, P., Jordan, K., and Müller, J. L. (2014). Use of virtual reality in forensic psychiatry. A new paradigm? *Nervenarzt* 85, 298–303. doi: 10.1007/s00115-013-3904-7

García-Sancho, E., Salguero, J. M., and Fernández-Berrocal, P. (2014). Relationship between emotional intelligence and aggression: a systematic review. *Aggress. Violent Behav.* 19, 584–591. doi: 10.1016/j.avb.2014.07.007

Garofalo, C., Velotti, P., and Zavattini, G. C. (2018). Emotion regulation and aggression: the incremental contribution of alexithymia, impulsivity, and emotion dysregulation facets. *Psychol. Violence* 8, 470–483. doi: 10.1037/vio0000141

Garrett, B., Taverner, T., Gromala, D., Tao, G., Cordingley, E., and Sun, C. (2018). Virtual reality clinical research: promises and challenges. *JMIR Serious Games* 6:e10839. doi: 10.2196/10839

Given, S. K., and Lisa, M. (2008). *Convenience sample. The sage encyclopedia of qualitative research methods*. Thousand Oaks, CA: Sage.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Prob.* 12, 436–445. doi: 10.1525/sp.1965.12.4.03a00070

González Moraga, F. R., Klein Tuente, S., Perrin, S., Enebrink, P., Sygel, K., Veling, W., et al. (2022). New developments in virtual reality-assisted treatment of aggression in forensic settings: the case of VRAPT. *Front Virtual Real.* 2:2. doi: 10.3389/frvir.2021.675004

Greenhalgh, T., and Abimbola, S. (2019). The NASSS framework a synthesis of multiple theories of technology implementation. *Stud. Health Technol. Inform.* 263, 193–204. doi: 10.3233/SHTI190123

Kip, H., Kelders, S. M., Bouman, Y. H. A., and Van Gemert-Pijnen, L. J. E. W. C. (2019b). The importance of systematically reporting and reflecting on eHealth development: participatory development process of a virtual reality application for forensic mental health care. *J. Med. Internet Res.* 21:e12972. doi: 10.2196/12972

Kip, H., Kelders, S. M., Weerink, K., Kuiper, A., Brüninghoff, I., Bouman, Y. H. A., et al. (2019a). Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. *Front. Psychol.* 10:406. doi: 10.3389/fpsyg.2019.00406

Kip, H., Ligthart, S. L. T. J., and Meynen, G. (2024). Ethiek en recht bij toepassing van nieuwe technologie in de forensische psychiatrie. 91–96.

Kouijzer, M. M. T. E., Kip, H., Bouman, Y. H. A., and Kelders, S. M. (2023). Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. Implement. *Sci. Commun.* 4:67. doi: 10.1186/s43058-023-00442-2

Levac, D., Glegg, S. M., Camden, C., Rivard, L. M., and Missiuna, C. (2015). Best practice recommendations for the development, implementation, and evaluation of online knowledge translation resources in rehabilitation. *Phys. Ther.* 95, 648–662. doi: 10.2522/ptj.20130500

Mullen, P. E. (2000). Forensic mental health. *Br. J. Psychiatry* 176, 307–311. doi: 10.1192/bjp.176.4.307

Portigliatti Pomeri, A., La Salvia, A., Carletto, S., Oliva, F., and Ostacoli, L. (2021). EMDR in Cancer patients: a systematic review. *Front. Psychol.* 11:590204. doi: 10.3389/fpsyg.2020.590204

Renaud, P., Trottier, D., Rouleau, J. L., Goyette, M., Saumur, C., Boukhalfi, T., et al. (2014). Using immersive virtual reality and anatomically correct computer-generated characters in the forensic assessment of deviant sexual preferences. *Virtual Real* 18, 37–47. doi: 10.1007/s10055-013-0235-8

Roberton, T., Daffern, M., and Bucks, R. S. (2015). Beyond anger control: difficulty attending to emotions also predicts aggression in offenders. *Psychol. Violence* 5, 74–83. doi: 10.1037/a0037214

Rogers, E. M. (1995). Diffusion of innovations: modifications of a model for telecommunications. *Die diffusion von innovationen in der telekommunikation.*, 17, 25–38. doi: 10.1007/978-3-642-79868-9_2

Skip, R. A., Koenig, S. T., and Brett, T. T. (2018). Clinical virtual reality: Emerging opportunities for psychiatry. *Focus (Am Psychiatr Publ)* 16, 266–278. doi: 10.1176/appi.focus.20180011

Slater, M., Perez-Marcos, D., Ehrsson, H. H., and Sanchez-Vives, M. V. (2008). Towards a digital body: the virtual arm illusion. *Front. Hum. Neurosci.* 2:6. doi: 10.3389/neuro.09.006.2008

Slater, M., Spanlang, B., and Corominas, D. (2010). "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence" in *ACM SIGGRAPH 2010 papers, SIGGRAPH 2010* (New York, USA: Association for Computing Machinery, Inc).

Smeijers, D., Bulten, E. H., Verkes, R. J., and Koole, S. L. (2021). Testing the effects of a virtual reality game for aggressive impulse management: a preliminary randomized controlled trial among forensic psychiatric outpatients. *Brain Sci.* 11:1484. doi: 10.3390/brainsci11111484

Smith, V., Warty, R. R., Sursas, J. A., Payne, O., Nair, A., Krishnan, S., et al. (2020). The effectiveness of virtual reality in managing acute pain and anxiety for medical inpatients: systematic review. *J. Med. Internet Res.* 22:e17980. doi: 10.2196/17980

So, B. P. H., Lai, D. K. H., Cheung, D. S. K., Lam, W. K., Cheung, J. C. W., and Wong, D. W. C. (2022). Virtual reality-based immersive rehabilitation for cognitive-and behavioral-impairment-related eating disorders: a VREHAB framework scoping review. *Int. J. Environ. Res. Public Health* 19:5821. doi: 10.3390/ijerph19105821

Sygel, K., and Wallinius, M. (2021). Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. *Front. Psychiatry* 12:673089. doi: 10.3389/fpsyt.2021.673089

Tamilmani, K., Rana, N. P., Wamba, S. F., and Dwivedi, R. (2021). The extended unified theory of acceptance and use of technology (UTAUT2): a systematic literature review and theory evaluation. *Int. J. Inf. Manag.* 57:102269. doi: 10.1016/j.ijinfomgt.2020.102269

Tong, A., Sainsbury, P., and Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int. J. Qual. Health Care* 19, 349–357. doi: 10.1093/intqhc/mzm042

Tuente, S. K., Bogaerts, S., Bulten, E., Keulen-De Vos, M., Vos, M., Bokern, H., et al. (2020). Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *J. Clin. Med.* 9, 1, 1–18. doi: 10.3390/jcm9072258

van Gemert-Pijnen, L., Kelders, S., Kip, H., and Sanderman, R. (2018). *eHealth research, theory and development*. Oxford, New York: Routledge.

Ventura, S, Cebollla, A, Herrero, R, and Banos, R. M. Qualitative research of an innovative virtual reality embodiment system: the machine to be another. Proceedings of the 12th international conference on disability virtual reality and associated technologies; (2018).

Appendix

Tables A1, A2.

TABLE A1 Semi-structured interview scheme for interviews with patients.

Questions for reflection during the immersion in every scenario	
What did you think of it?	RQ1
What did it feel like?	RQ1 + 2
On a scale of 1 to 10, with 1 being completely relaxed and 10 experiencing extreme stress, how high was your level of distress?	RQ2
Questions for evaluation post-immersion	
What is your first impression of VR?	RQ1
What did you like or positively appreciate about VR?	RQ1
Are there any things you did not like or would improve in VR?	RQ1
How realistic/authentic did it feel?	RQ1
What did you notice about yourself in VR? When did you notice this?	RQ1 + 2
What was your level of distress during the VR scenarios?	RQ2
What stood out to you while you were in VR?	RQ1
Would you be willing to use VR in treatment? Why (not)?	RQ3
What should we consider when introducing VR to patients?	RQ3

TABLE A2 Semi-structured interview scheme for interviews with healthcare providers.

What is your initial impression of using the VR dashboard?	RQ1
What did you find challenging about using the VR dashboard? • How would you evaluate the usability of the VR dashboard? • How would you evaluate the design of the VR dashboard? • To what extent do you understand the use of the VR dashboard?	RQ1
Do you need more information on the dashboard to build a VR scenario? If so, what information do you think is currently missing?	RQ1
What would you like to change or add to this dashboard to improve it?	RQ1
Are there any (content-related) environments, characters or triggers that are missing so far?	RQ1
To what extent would you like to use this VR application in treatment? • How would you apply this VR application in treatment? • Within which existing treatment protocols would you apply VR? • What would that look like in practice? Can you give examples?	RQ3
What are additional points of attention in terms of timing, introduction, and explanation that are needed to introduce VR to patients?	RQ3
Would you prefer any implementation activities or materials, such as training or protocols to support you during the introduction of VR in treatment? • What kind of implementation activities/materials would you prefer? • What would that look like in practice? Can you give examples?	RQ3

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 09 November 2023
DOI 10.3389/fpsyg.2023.1201485

OPEN ACCESS

EDITED BY
Radwa Khalil,
Jacobs University Bremen, Germany

REVIEWED BY
Fernando Renee González Moraga,
Regional Forensic Psychiatric Clinic Växjö, Sweden
Raquel Simões de Almeida,
School of Health, Polytechnic of Porto, Portugal

*CORRESPONDENCE
Lisa Klein Haneveld
✉ l.kleinhaneveld@transfore.nl

RECEIVED 20 April 2023
ACCEPTED 25 September 2023
PUBLISHED 09 November 2023

CITATION
Klein Haneveld L, Kip H, Bouman YHA, Weerdmeester J, Scholten H and Kelders SM (2023) Exploring the added value of virtual reality biofeedback game DEEP in forensic psychiatric inpatient care—A qualitative study. *Front. Psychol.* 14:1201485. doi: 10.3389/fpsyg.2023.1201485

Exploring the added value of virtual reality biofeedback game DEEP in forensic psychiatric inpatient care—A qualitative study

Lisa Klein Haneveld[1,2]*, Hanneke Kip[2], Yvonne H. A. Bouman[1], Joanneke Weerdmeester[3,4], Hanneke Scholten[4,5] and Saskia M. Kelders[2,6]

[1]Department of Research, Transfore, Deventer, Netherlands, [2]Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands, [3]HKU University of the Arts, Centre for Educational Research and Innovation, Utrecht, Netherlands, [4]Games for Emotional and Mental Health (GEMH) Lab, Enschede, Netherlands, [5]Communication Science, Department of Technology, Human and Institutional Behavior, University of Twente, Enschede, Netherlands, [6]Opentia Research Unit, North-West University, Vanderbijlpark, South Africa

Background: Low motivation and suboptimal cognitive skills are common among forensic psychiatric patients. By focusing on doing and experiencing, innovative technologies could offer an alternative to existing treatment for this patient group. One promising technology is DEEP, a VR biofeedback game that teaches diaphragmatic breathing, which has shown its potential in reducing stress in other populations. This exploratory study aimed at identifying if, how and for whom DEEP can be of added value in forensic mental healthcare.

Methods: This study used a qualitative approach. Six focus groups with 24 healthcare providers and 13 semi-structured interviews with forensic psychiatric inpatients were conducted in two Dutch forensic mental healthcare organizations. All healthcare providers and patients experienced DEEP before participating. The data were coded inductively, using the method of constant comparison.

Results: The data revealed six themes with accompanying (sub)codes, including (1) the possible advantages and (2) disadvantages of DEEP, (3) patient characteristics that could make DEEP more or (4) less suitable and beneficial, (5) ways DEEP could be used in current treatment, and (6) conditions that need to be met to successfully implement DEEP in forensic mental healthcare. The results showed that DEEP can offer novel ways to support forensic psychiatric patients in coping with negative emotions by practicing diaphragmatic breathing. Its appealing design might be suitable to motivate a broad range of forensic psychiatric patient groups. However, DEEP cannot be personalized, which might decrease engagement and uptake of DEEP long-term. Regarding its place in current care, DEEP could be structurally integrated in existing treatment programs or used *ad hoc* when the need arises. Finally, this study showed that both healthcare providers and patients would need practical support and information to use DEEP.

Conclusion: With its experience-based and gamified design, DEEP could be useful for forensic mental healthcare. It is recommended that patients and healthcare providers are included in the evaluation and implementation from

the start. Besides, a multilevel approach should be used for formulating implementation strategies. If implemented well, DEEP can offer new ways to provide forensic psychiatric patients with coping strategies to better control their anger.

KEYWORDS

DEEP, virtual reality, forensic mental healthcare, focus groups, interviews

1. Introduction

In forensic mental healthcare, patients receive mandatory treatment due to their delinquent or criminal behavior, which could partly be explained by one or more psychiatric illnesses (Bloem et al., 2011; Meynen, 2017; Sygel and Wallinius, 2021). Treatment of this heterogeneous patient group can be complex, due to differences in type of offences and diagnoses, generally low treatment motivation, and often suboptimal cognitive skills, such as reading, writing and self-reflection (Drieschner and Boomsma, 2008; Svensson et al., 2015; Kip et al., 2018). All these factors result in a unique setting, with a need for treatment that is tailored to individual needs. While treatment of forensic psychiatric patients resulted in a reduction of recidivism rates, there is room for optimizing treatment outcomes and further reducing these rates (Delfin et al., 2019; Probst et al., 2020; Edberg et al., 2022). One of the commonly used treatment frameworks in forensic mental healthcare is Cognitive Behavioral Therapy (CBT). While CBT has been very effective in treating depression and anxiety in general mental healthcare, for treatment of aggression in (forensic) psychiatric patients it has been less effective (Del Vecchio and O'Leary, 2004; Saini, 2009; Henwood et al., 2015). Additionally, an earlier study concluded that forensic psychiatric inpatients do benefit from CBT regarding psychopathology and coping, but only a small group showed reliable change over time (Timmerman and Emmelkamp, 2005). A possible explanation for these results is that many CBT-based treatment methods rely on the ability of patients to think and talk about their behavior, which might be challenging for forensic psychiatric patients due to limited cognitive skills (Kip et al., 2019c; Kip and Bouman, 2020, 2021). These shortcomings of existing treatments highlight the need for novel strategies and interventions that better suit the needs, skills, and interests of forensic psychiatric patients.

eHealth – which is the use of technology to support health, wellbeing, and healthcare - can be used to improve care for forensic psychiatric patients (Kip et al., 2020a; Stoll et al., 2020; Tossaint-Schoenmakers et al., 2021). Over the last 10 years, an increasing number of eHealth interventions has been introduced in forensic mental healthcare (Kip et al., 2019b; Torous et al., 2020; Kirschstein et al., 2023). Many technologies that still rely on written text, such as web-based interventions and do not seem to fully fit the forensic psychiatric patient population (Kip et al., 2020b). However, experience-based technologies such as virtual reality, serious games and wearables with biofeedback can offer new, innovative possibilities to increase emotion regulation and wellbeing (Ilioudi et al., 2023; Kothgassner et al., 2023), and involve forensic patients and teach them the skills that are needed to prevent recidivism (Kip et al., 2019c; Kip and Bouman, 2020, 2021). A Dutch study with 110 forensic psychiatric healthcare providers and patients demonstrated the enthusiasm for these technologies, particularly because of their immersive qualities and focus on "doing and real-time experiencing," instead of "thinking and talking" about behavior (Kip et al., 2019c). These findings indicate that an experience-based, gamified approach accounts for the low literacy- and treatment motivation levels of the forensic population. Despite the apparent potential, not much research has been conducted on the added value of these experience-based, gamified technologies in forensic mental healthcare.

Over the years, virtual reality (VR) has become a growing topic of interest within psychological therapy (Rizzo et al., 2018; Sygel and Wallinius, 2021; Kothgassner et al., 2023; Kouijzer et al., 2023). In VR, it is possible for the user to be physically present and interact with a virtual environment in an engaging way (Kim et al., 2017; Pillai and Mathew, 2019). Additionally, while more research on the effectiveness is necessary, VR seems to be a potentially suitable tool for treatment of aggression (Kip et al., 2019b; Klein Tuente et al., 2020; Smeijers et al., 2021). An example of such a VR-application is DEEP. This experience-based and gamified VR technology provides a unique combination of breathing techniques and biofeedback to teach its player diaphragmatic (or deep) breathing in an intuitive and engaging manner to reduce stress and anxiety (Weerdmeester, 2021). By applying deep breathing, the user "swims" through a fictional virtual underwater environment and explores colorful caves while following a, by artists designed, route (see **Figure 1**). The biofeedback in DEEP can provide users with real-time information about their breathing by using a waistband that measures the movement of their diaphragm and by showing them visualizations of how well they are inhaling and exhaling (e.g., corals that light in DEEP and visual breathing circles). From a practical point of view, DEEP takes the relatively little time and technical skills of healthcare providers (Weerdmeester, 2021). This might make DEEP it easier to adopt than for instance interactive VR with roleplaying functionalities. Additionally, DEEP is designed to be an appealing and gamified environment, which might positively affect treatment motivation (Kip et al., 2019b). Moreover, DEEP is designed to be a calm environment for users to relax and decrease stress, which could be a helpful way for psychiatric patients to increase their wellbeing when a physical calm room is not available (Ilioudi et al., 2023). Hence, DEEP might provide possibilities to add something entirely new to forensic treatment, because of its emphasis on deep breathing and providing real-time biofeedback to its user. This is often not possible in standardized treatment, where the patient must rely on their memory and ability

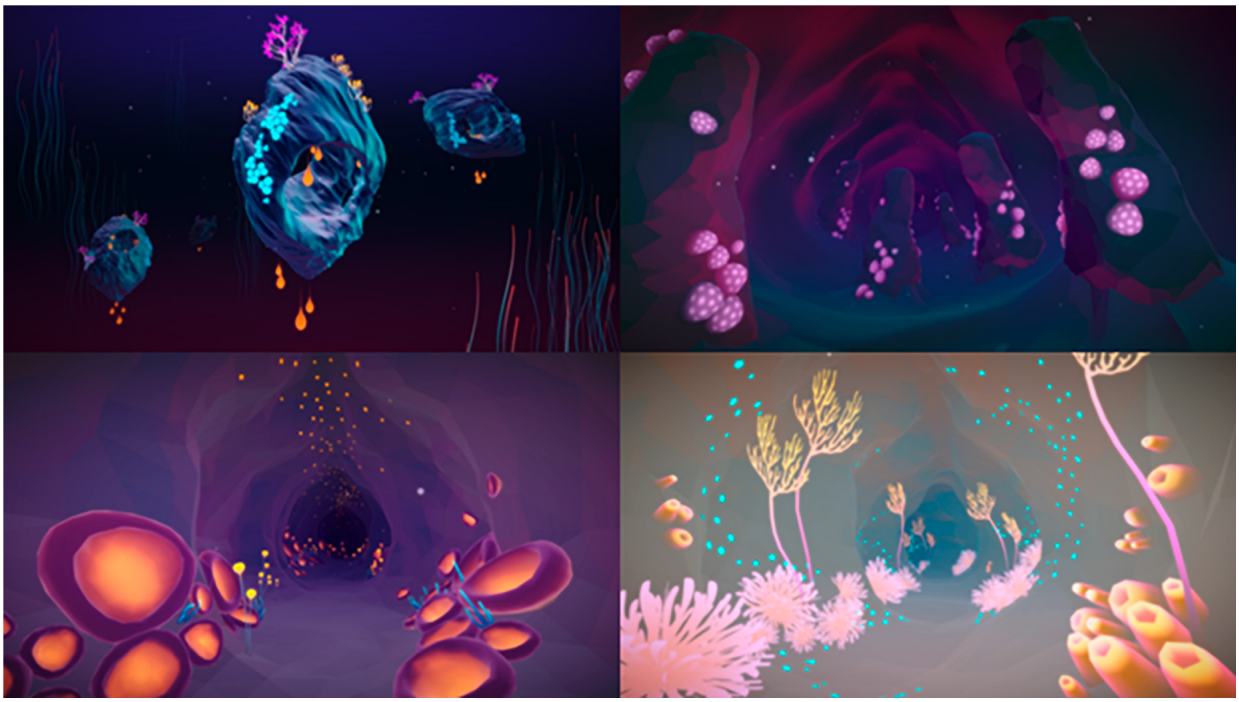


FIGURE 1
Visual examples of DEEP. Reproduced with permission from Development Team Explore DEEP (https://exploredeep.com).

to reflect on past experiences and behavior. While DEEP has not yet been studied in the forensic psychiatric population, it seems that an interventions like DEEP can be of added value for the forensic mental healthcare due to its use of biofeedback, deep breathing, and its immersive, gamified design.

Due to its working mechanisms, it DEEP might be a useful approach to prevent aggression. For example, studies have shown that biofeedback can be helpful in reducing aggression. To illustrate, interventions based on heart rate (variability) biofeedback can decrease anger in adolescents and improve emotion-regulation in offenders, which is an often-re-occurring treatment goal for forensic psychiatric patients (Savard, 2017; Gray et al., 2019). Moreover, diaphragmatic breathing has shown to be supportive in reducing aggressive behavior in various psychiatric patient groups, by modulating the heart rate and specific neural circuits that are involved in emotion regulation (Gillespie et al., 2012; Phillips et al., 2019). Regarding the effectiveness of DEEP: recent studies have also shown DEEP to be effective in reducing stress and anxiety in students and in showing less disruptive behavior in adolescents with behavioral problems (Van Rooij et al., 2016; Bossenbroek et al., 2020). Despite the promising results of biofeedback in general and DEEP in specific settings, it is unclear if and how DEEP can be used by forensic psychiatric patients and support them in their treatment outcomes (e.g., reduce stress and anger, or increase emotion-regulation). Likewise, the implementation of VR in forensic mental healthcare can be challenging, due to the necessity of healthcare providers to develop the right attitude and skill set to integrate VR in their treatment (Kip et al., 2020a; Tossaint-Schoenmakers et al., 2021; Kouijzer et al., 2023). Therefore, it is important to gain insight into how to successfully implement DEEP in forensic mental healthcare.

1.1. The current study

The objective of this study is to identify if and how DEEP can be of added value for forensic psychiatric inpatient settings according to healthcare providers and patients in forensic psychiatric inpatient care. The four accompanying research questions are to explore (1) expected advantages and disadvantages of DEEP in forensic mental healthcare, (2) for which types of forensic psychiatric inpatients DEEP could be most effective, (3) in what ways DEEP could be used in current treatment of inpatients, and (4) which factors are important for implementing DEEP within clinical forensic mental healthcare, according to healthcare providers and patients.

2. Materials and methods

2.1. Study design and setting

In the current study, focus groups with healthcare providers and semi-structured interviews with patients were conducted to answer the research questions. The Consolidated criteria for Reporting Qualitative research (COREQ) guidelines were used to report the focus groups and interviews (Tong et al., 2007). Before conducting the focus groups and interviews, participants verbally received information about DEEP as well as the scope of the study. Additionally, it was ensured that the forensic psychiatric patients understood that participating on this study was voluntarily and that it was not part of their treatment program. With support from the researcher (LK), who was already trained in using DEEP, all

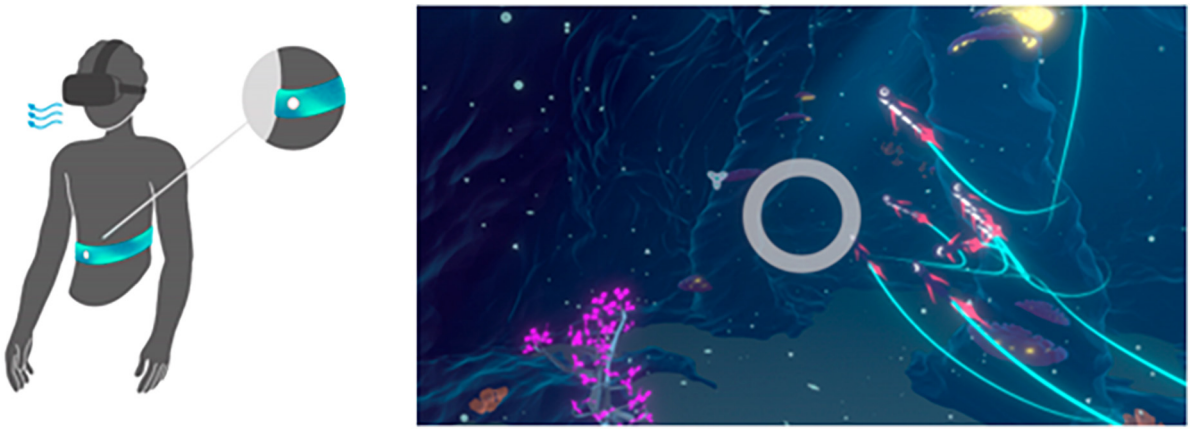


FIGURE 2
Waistbelt and wide circle to provide biofeedback to DEEP user. Reproduced with permission from Development Team Explore DEEP (https://exploredeep.com).

participants were able to test DEEP to experience the underlying mechanisms beforehand. This was ethically feasible, because during earlier studies in which participants used DEEP no severe physical risks were found, except for some motion sickness that can occur when someone is not used to VR (Bossenbroek et al., 2020; Weerdmeester et al., 2021). Ethical approval was given by the Ethics Committee of the University of Twente (Behavioral, Management and Social Sciences, nr. 21119).

This study has been conducted at inpatient clinics of two forensic mental healthcare organizations: Transfore and De Woenselse Poort. Both organizations offer forensic mental healthcare to in- and outpatients who have committed or are on the verge of committing a criminal offence, due to their psychiatric problems. Transfore has multiple locations in the east of The Netherlands and offers treatment to over 1,500 patients per year. De Woenselse Poort is based in the south of The Netherlands and provides treatment for around 550 patients per year. In total, six inpatient clinics (three per organization) participated.

2.2. DEEP

DEEP is a VR-game that uses biofeedback to teach relaxation skills to its user and is based on scientific knowledge about anxiety and stress regulation (Van Rooij et al., 2016; Weerdmeester, 2021). The player is placed in a surrealistic and immersive underwater world that they can move through by using their own diaphragmatic -or deep- breathing. During DEEP, the player wears a waistband that continuously measures their breathing. Within the game, a wide circle in front of the player mirrors the breathing of the player (see **Figure 2**). Additionally, the corals and plants in the world grow, shrink, and change in illumination with every inhale and exhale. Both techniques provide feedback on changes in physical signals, in this case, the way someone is breathing, which is called biofeedback (Weerdmeester, 2021). By deep breathing via their diaphragm, the player can move through the game, while shallow upper-chest breathing -which often happens in stressful situations- is not or barely measured by the waistband. Therefore, shallow breathing hinders the progress in the game, while diaphragmatic breathing results in progress. As a result, the player has continuous insight in their own progress and whether their breathing needs adjustment. By learning deep breathing in a gamified and engaging way, the user might become able to better cope with negative emotions and increase their wellbeing (Tellhed et al., 2019; Ahmadpour et al., 2020; Weerdmeester et al., 2021). In this study, the same version of DEEP was used as the one used in an earlier conducted randomized controlled trial (Weerdmeester et al., 2021). The version was built for the *Oculus Rift dk1* headset and *HTC vive*.

2.3. Participants

In this study both healthcare providers and patients were included since they will have an active role in the use of DEEP. Healthcare providers will have a role in introducing DEEP and supporting patients, while patients are the end-users of DEEP. All healthcare providers directly involved in any treatment, support or supervision of inpatients were eligible to participate and were included via convenience sampling. With help of managers and the use of information folders, the study was introduced to the inpatient healthcare providers. Patients were recruited via the healthcare providers who participated in the focus groups. They were asked to inform their patients about the study and distribute information flyers. Patients were excluded from participation if they were diagnosed with a current psychosis or if a therapist indicated that they were not able to participate for any relevant reason.

2.4. Materials and procedure

Six focus groups with healthcare providers and thirteen interviews with patients were conducted between November 2021 and April 2022 by one researcher (LK) and took place at the location of the forensic mental healthcare organization where the healthcare providers were working, and the patients received treatment.

2.4.1. Focus groups

Each focus group with healthcare providers took around 60 min (min = 43, max = 71). At the start of the focus groups, participants were informed about the design and mechanisms of DEEP as well as the scope of this research. After the introduction of the goal and structure of the focus group, all participants signed the informed consent. In the first part of the focus groups each participant was able to use DEEP for a couple of minutes to gain an impression of if and how the biofeedback in DEEP changed the way they were breathing. During this part, all caregivers tried out DEEP. However, one caregiver only used DEEP for a minute as they experienced motion sickness.

In the second part of the semi-structured focus group, a topic list was used to explore the first impressions of DEEP (see **Appendix A**) and elicit scenarios of when, how and for whom DEEP could be used in forensic treatment. This topic list was developed by the researchers and evaluated by the other project members. During the focus groups, the participants were able to express their first impression of DEEP and what they liked and disliked about using it, using questions as *"To what extent do you think that DEEP can be of added value for forensic treatment?"* They were also asked about the advantages and added value DEEP might have for the forensic psychiatric treatment, and specifically for what type of patients. An example of a question is: *"What type of patients would benefit most from using DEEP?"* Finally, the participants answered questions about possible barriers of using DEEP as well what they found important preconditions for implementing DEEP. For example: *"When implementing DEEP in your forensic clinic, what factors are important to take into account?"*

2.4.2. Interviews

The semi-structured interviews with patients took around 20 min (min = 15 and max = 27 min). The interviews used the same themes from the topic list that was used in the focus groups, but were kept short and concise, which fits the skills and attention span of most forensic psychiatric patients (Kip et al., 2019a, 2022). Before conducting the interview, the participating patients were given a brief introduction about DEEP and the interview. During this introduction, the researcher verbally explained the content provided in the information folder in short sentences, allowing patients of all cognitive levels to understand the scope of the study and interview. All participants were also given the opportunity to ask questions. After this introduction, the participants signed the informed consent, which consisted of short statements for them to read. After the informed consent was signed, the patients were able to become acquainted with DEEP for several minutes. They did so in the same way as the caregivers. Only one patient had trouble using DEEP and stated to feel scared. However, after a moment of standing still and getting used to the environment, they were able to move around and experience DEEP.

In the second part of the interview, a topic list was used (see **Appendix B**), which was evaluated beforehand by caregivers specialized in working with forensic psychiatric patients with lower cognitive abilities. The semi-structured approach allowed the researcher to ask probing questions, deviate from the order of questions, or tailor them depending on the patient's cognitive skills and attention span. Patients were asked questions about their first impression of DEEP and which advantages they thought it might have for them via questions such as: *"Which advantages do you think DEEP might have for you? Could it help you?."* They were also asked if they were willing to use DEEP more often and if so, when. An example question is: *"On which moments or periods would you like to use DEEP?"* Moreover, questions were asked about possible disadvantages and when patients are not willing to use DEEP. Finally, they were asked about considerations before implementing DEEP and what they would require before they would be able to use DEEP.

2.5. Data analysis

All focus groups and interviews were audio-recorded, transcribed verbatim and coded inductively. By using the method of constant comparison (Boeije, 2002), the raw data from the focus groups and the interviews were organized into categories after which codes were created to connect the fragments in each category. The coders (LK and HK) used Microsoft Excel to code the fragments as it is a simple and cost-effective way to thematically and inductively code qualitative data (Bree and Gallagher, 2016). First, one coder (LK) read the transcripts to become familiar with their content. Second, all fragments related to one of the research questions were selected and divided over each research question: (dis-)advantages of using DEEP (RQ 1), characteristics of patients that would or would not benefit of using DEEP (RQ 2), ways of using DEEP in treating patients (RQ 3) and conditions of implementing DEEP in forensic mental healthcare organizations (RQ 4). Based on these fragments, one coding scheme for every research question (theme) was created inductively: the main and subcodes, with accompanying definitions, were formulated using a bottom-up approach, using the content of the fragments. Several main codes and subcodes were identified and used to code the fragments. Throughout this process, the coding schemes and definitions were constantly adapted and updated by two researchers (LK and HK), using an iterative approach. After coding four focus groups and nine interviews, data saturation was reached and no new codes were identified. The first coder used the coding schemes to code all fragments, which then was sent to the second coder (HK) to code 20% of the fragments independently with an agreement rate of 78%, which is acceptable in the coding process (Miles and Huberman, 1994). The disagreement between the two coders was discussed until consensus was reached, and minor adaptations were made to the coding schemes accordingly.

3. Results

3.1. Demographics

For this study, 24 healthcare providers and 13 patients were included. Regarding the healthcare providers, 22 were included in 6 focus groups with an average of four participants per group (min = 2, max = 6). Four focus groups took place at Transfore (n = 14) and two at De Woenselse Poort (n = 10). Two of the 24 healthcare providers were interviewed individually, because they were not able to join the organized focus groups but were still interested in using DEEP and sharing their experiences. Of all

TABLE 1 Possible advantages of using VR-DEEP, according to healthcare providers (n = 24) and patients (n = 13).

Codes	Definition of code	Codes[a]	Healthcare providers[b]	Patients[c]
Advantages–content				
Relaxation and emotion regulation	Playing DEEP contributes to a feeling of relaxation and rest and helps the player in recognizing and dealing with arousal, stress or anger	34	4 (8)	10 (26)
Deep breathing	Playing DEEP contributes to the recognition and practice of diaphragmatic breathing	23	13 (16)	4 (7)
Meditation and Mindfulness	The principles of DEEP are reminding of principles of meditation and Mindfulness, e.g., the focus on the body and breathing combined with the surrounding and music	6	2 (2)	1 (4)
Advantages–design				
Appealing design	In DEEP the player has a fun and extraordinary experience by stepping into a surrealistic and relaxing environment	56	18 (25)	9 (31)
Accessible	DEEP requires little cognitive and reflective skills and can be used by everyone	14	11 (13)	1 (1)
User-friendly	DEEP can be set up easily, quickly and independently	3	2 (3)	

[a]The total number of times a code was mentioned in all focus groups and interviews. [b]The number of different healthcare providers that mentioned a code, and (#) the total number of times the code was found in all focus groups. [c]The number of patients that mentioned a code, and (#) the number of times the code was found in all interviews with patients.

participating healthcare providers, 7 identified as male and 14 as female. All healthcare providers worked at an inpatient clinic: one healthcare provider was a psychologist, one a coordinating therapist, two were team managers, two were forensic nurses in training and 18 healthcare providers were socio-therapists/forensic nurses. All interviewed patients received inpatient care, of which 11 identified as male and 2 as female. An overview of the included healthcare providers and patients per forensic inpatient clinic are provided in **Appendix C**.

3.2. Possible advantages of using DEEP

Healthcare providers and patients were asked to provide their view on DEEP and the possible advantages it might have for forensic inpatients. The participants provided possible advantages on the content as well as the design of DEEP. The identified codes are provided in **Table 1**.

Regarding the content of DEEP, both healthcare providers and patients indicated that playing DEEP might be able to contribute to more *relaxation and better emotion regulation*. Four healthcare providers and 10 patients thought that DEEP might predominantly help to relax and de-stress during tense situations or after conflict. Second, nine patients mentioned that they often feel stress, while seven of them added that using DEEP made them feel more relaxed.

Second, DEEP might help focus on *deep breathing*. During the focus groups, 13 healthcare providers acknowledged DEEP to be a useful intervention in learning patients to breathe slowly and deeply, which might help them to de-stress and cope with stressful or emotional situations in the future. Additionally, four patients stated that focusing on their breathing was a new but useful experience to find relaxation.

A third code that was identified by two healthcare providers is that DEEP might be able to introduce the player to principles of other types of interventions such as *mindfulness and meditation*, in a down-to-earth and playful manner, which might make it easier for patients to open themselves up to it. This view was acknowledged by one patient in stating the following: *"There are men here that like mindfulness and meditation and use it in their daily lives. But there are also men who find it rubbish. Maybe DEEP can convince them that getting some rest in your mind is really good and does not have to be boring."*

As for the design features of DEEP, 18 participants of the focus groups found DEEP to have an *appealing design*, which might help increase treatment motivation of forensic inpatients. Ten healthcare providers illustrated that this group often shows low treatment motivation and are not always willing to try new interventions. Because DEEP is developed as a game and is meant to provide a relaxing experience, patients might become more enthusiastic to try new interventions that are focused on experiencing, instead of reflecting. Participant 24 indicated the following regarding the suitability of the design of DEEP for forensic inpatients:" DEEP *does not ask you to do anything, except doing something you always do, namely breathing, and at the same time you learn to focus on all kinds of internal things unconsciously. While doing something fun. Yes, I do think that is especially fitting for our patient group."* According to nine patients, DEEP is a new, fun, and relaxing experience that they do not always get to have in their current treatment. Some added that DEEP enabled them to step into another world, where they can swim, steer, and breathe, which felt like an enjoyable experience.

Second, 11 healthcare providers mentioned that DEEP is *accessible* for everyone and does not require much cognitive skill, which is a great advantage for the forensic patient group. Participant 5 stated the following regarding the accessibility of DEEP: *"I think that it is a game with many features. But it functions also as a resting place, which asks relatively little of you as a player."* Finally, two healthcare providers appreciated that DEEP is *user-friendly* and that it can be started up easily and quickly, which enhances the chance that healthcare providers will structurally use an intervention with their patients.

TABLE 2 Possible disadvantages of using VR-DEEP, according to healthcare providers (n = 24) and patients (n = 13).

Codes	Definition of code	Codes[a]	Healthcare providers[b]	Patients[c]
Disadvantages of DEEP				
Simple design	DEEP has a simple and straight-forward design, which might affect immersion negatively and becomes boring	(24)	5 (8)	8 (16)
Suboptimal controls and biofeedback	The controls and biofeedback of DEEP is dated and does not always work properly, e.g., the waistbelt and the in-game visuals	(18)	8 (10)	5 (8)
Overwhelming design	DEEP requires the player to look around, steer, breath and process visual feedback, which might overwhelm or induce stress	(7)	3 (3)	2 (4)
No personalization	It is not possible to choose other worlds or levels in DEEP and there are no available triggers that can be personalized	(4)	2 (3)	1 (1)
Many wires	DEEP needs many wires to connect the VR-set, waistband and sensor to a laptop, obstructing the player to rotate properly	(3)	3 (3)	
VR motion sickness	DEEP can induce motion sickness or dizziness to its player	(2)	2 (2)	
No clear connection with daily life	In DEEP diaphragmatic breathing is trained in a surrealistic environment, with no direct connection to real life situations.	(2)	2 (2)	
Expensive	Purchasing DEEP and a VR-set is expensive	(1)	1 (1)	

[a]The total number of times a code was mentioned in all focus groups and interviews. [b]The number of different healthcare providers that mentioned a code, and (#) the total number of times the code was found in all focus groups. [c]The number of patients that mentioned a code, and (#) the number of times the code was found in all interviews with patients.

3.3. Possible disadvantages of using DEEP

Additionally, to the first category, healthcare providers and patients were asked to provide their insight on possible disadvantages of DEEP for forensic inpatients. The identified codes are provided in Table 2.

As a first disadvantage of DEEP, five healthcare providers mentioned that DEEP might not be able to engage the player long term, because of its *simple and straight-forward design*. During the focus groups, it was mentioned that the new generation inpatients is really into new gaming technology and might find DEEP too simple or too easy to keep emerging themselves after they used DEEP a couple of times. Eight patients shared this view and one of them illustrated it as following: *"In DEEP, you can only just swim around, without being able to communicate with the world around you. You cannot do anything or touch anything. I feel like that might become boring after a while."*

Second, eight healthcare providers indicated that the *controls and biofeedback felt suboptimal* or dated. Additionally, five patients mentioned that the waist belt did not always give them correct feedback on their breathing. This was again confirmed by some healthcare providers, of whom some also stated that the biofeedback that was given in-game (e.g., corals that light up with every breath), was sometimes hard to perceive and connect to what the player was doing. This is illustrated by participant 3: *"I am wondering if DEEP should give more visual stimuli, so that questions like "am I doing alright?" are answered. During DEEP I was able to focus on my breathing, but for our patient group it is good to clearly experience what happens if they are breathing "well" or "bad" If it stays subtle it might become too abstract for a part of our patient group."*

Third, three healthcare providers indicated that DEEP could be *overwhelming* for some of their patients, because it focuses on many features: breathing, steering, and looking around while processing the biofeedback. A couple of patients stated that they were overwhelmed, especially those who never experienced virtual reality before.

Fourth, two healthcare providers mentioned the *lack of personalizatio*n in DEEP, with no available triggers for different patients. They stated that for some patients it might be helpful to use more triggers in DEEP, so that they can train their deep breathing during a tense situation. For other patients it might help to have a less dark and surreal environment because they are easily overwhelmed.

In this study, an older version of DEEP was used that needed wires to connect a laptop to the VR-set and the waist belt. Hence, three healthcare providers felt that the *many wires* might obstruct someone to move properly while playing DEEP. Two healthcare providers were also wondering if playing DEEP could not induce *motion sickness*. Two other healthcare providers were questioning if the design and mechanics of DEEP *lack a clear connection to daily life*. Because DEEP uses a surreal environment it might become difficult for patients to apply the things they learn in their real-life situations. Finally, one healthcare provider mentioned that purchasing DEEP might be too *expensive*, when compared to other interventions, such as free apps that focus on meditation and deep breathing.

3.4. Patients that might benefit from using DEEP

The third main category of codes relates to patient groups that might benefit most from using DEEP. Only a couple of patients were able to indicate which patient groups might or might not benefit from DEEP. Most patients stated that they had a hard time predicting which patients would like DEEP or might find DEEP useful. They reported that they just focused on themselves and did not feel like interfering in other patients' treatment. The identified codes are provided in Table 3.

First, nine healthcare providers indicated that almost everyone might benefit from using DEEP, patient, or non-patient. However, seven of them added that it is essential that the *user has an open*

TABLE 3 Possible patient characteristics that might benefit from DEEP, according to healthcare providers (n = 24) and patients (n = 13).

Codes	Definition of code	Codes[a]	Healthcare providers[b]	Patients[c]
Patient characteristics that might benefit				
Everyone who has an open mind	Not only forensic patients, but everyone can benefit from DEEP, as long as one is open to the effects of breathing and relaxation exercises.	10	7 (9)	1 (1)
Emotion regulation issues	Patients who have difficulty controlling their emotions	8	7 (8)	
Body signals recognition issues	Patients who have difficulty recognizing and verbalizing their body signals	5	4 (5)	
Suboptimal cognitive skills	Patients with less optimal cognitive skills	4	4 (4)	
Personality issues	Patients with personality issues, such as borderline personality disorder or antisocial disorder	3	2 (3)	
Introversion	Patients who are introverted or are not comfortable talking about what they are experiencing	2	1 (1)	1 (1)

[a]The total number of times a code was mentioned in all focus groups and interviews. [b]The number of different healthcare providers that mentioned a code, and (#) the total number of times the code was found in all focus groups. [c]The number of patients that mentioned a code, and (#) the number of times the code was found in all interviews with patients.

mind toward the idea of deep breathing and using it as a coping mechanism in reducing negative emotions.

Second, seven healthcare providers mentioned that patients who have *difficulties in regulating their emotions* could benefit from DEEP, especially those with stress, anxiety, and reactive aggression. This is further explained by the following quote of participant 7: *"I think that patients who feel tense really quickly and therefore can react impulsively might benefit from DEEP. For example, (name patient) who tried DEEP, is quite impulsive in his behavior and reactions. And I saw how DEEP affected him and how he felt more relaxed. So, for patients like him DEEP could be a good fit."*

Third, four healthcare providers indicated that DEEP could help patients who have *issues with recognizing their bodily signals* in helping them to focus on their breathing and what they are experiencing while doing that. Participant 21 illustrated this as following: *"It seems to me that this is a good game for most of our patients. They sometimes have trouble pin-pointing their feelings. By using DEEP they can just experience their feelings, while focusing on their breathing."*

Fourth, four healthcare providers mentioned that patients with *suboptimal cognitive skills* could benefit from using DEEP. They stated that DEEP is easy to use, and its goal can be understood without relying on reading, writing, or reflecting skills of its user.

Fifth, two healthcare providers indicated that DEEP can be useful for people with issues *regarding their personality*, because they are not able to manipulate the situation, but must put trust in the intervention and their own breathing. Finally, one healthcare provider and one patient mentioned patients who are *introverted* as a group that might benefit from DEEP. This patient stated the following: *"In DEEP you can just be in your own world. You are able to artificially open yourself to the experience, without having to share it with others if you don't want to. That might provide a shelter for someone to go back to and de-stress."*

3.5. Patients that might benefit less from using DEEP

The fourth category relates to patient groups that might benefit less from using DEEP. An overview of this category and the identified codes are given in Table 4.

Sixteen healthcare providers provided no specific patient group that would not benefit from DEEP. Five healthcare providers were wondering whether patients with *psychotic vulnerability* could benefit from DEEP. They indicated that those patients might become overwhelmed or paranoid, which is illustrated by the following quote: "*I might not use DEEP with patients who are in a psychotic, paranoid state. If someone is totally shielded by the VR-headset he or she has to trust their surroundings. They might wonder what would happen if they surrendered to another reality, which is necessary while playing a VR-game."*

Additionally, one healthcare provider mentioned that patients with *epileptic sensitivity* might not benefit from DEEP, because of the flashing lights of the corals and plants in the game. Another healthcare provider mentioned *(game) addiction sensitivity* as something to keep in mind as some patients who are suffering from this seem to be more interested in finding new ways to distract themselves, than using the intervention to learn something new. One patient acknowledged this as well and stated the following: *"Of course it would be nice if you can just step into a room and put DEEP on your head. But with these kinds of people, you have to be careful that it does not become something like drugs, that it occupies their mind 24/7."*

Finally, one healthcare provider indicated that patients with *balance issues* might not benefit from DEEP or virtual reality in general, as they might become disoriented.

3.6. Possible ways DEEP could be used in current forensic care

Healthcare providers and patients were asked about possible ways DEEP could effectively be integrated in current forensic care. The identified main and subcodes are provided in Table 5.

All healthcare providers and most patients mentioned one way how DEEP could be used in current forensic care. First, nine healthcare providers and seven patients mentioned that DEEP could be introduced *ad hoc* by a healthcare provider or used on patients' own initiative. This way DEEP can be used as a flexible

TABLE 4 Possible patient characteristics that might less benefit from DEEP, according to healthcare providers (n = 24) and patients (n = 13).

Codes	Definition of code	Codes[a]	Healthcare providers[b]	Patients[c]
Patient characteristics that might benefit less				
Psychotic vulnerability	Patients who have psychotic sensitivity or are having a psychotic episode	7	5 (7)	
Epileptic sensitivity	Patients who are sensitive for epileptic episodes	3	1 (2)	1 (1)
(Game)addiction sensitivity	Patients who have (game)addiction sensitivity or who tend to obsessively use new things	2	1 (1)	1 (1)
Balance issues	Patients who have issues in their balance, which makes them sensitive for dizziness or motion sickness	1	1 (1)	

[a]The total number of times a code was mentioned in all focus groups and interviews. [b]The number of different healthcare providers that mentioned a code, and (#) the total number of times the code was found in all focus groups. [c]The number of patients that mentioned a code, and (#) the number of times the code was found in all interviews with patients.

TABLE 5 Possible ways DEEP could be used in current treatment, according to healthcare providers (n = 24) and patients (n = 13).

Codes	Definition of code	Codes[a]	Healthcare providers[b]	Patients[c]
Use of DEEP in practice				
Ad hoc on patient initiative	Using DEEP as a flexible tool that provides patients the chance to use DEEP on their own initiative when they feel they need it.	19	9 (10)	7 (9)
Complementing existing breathing exercises	Using DEEP as a complementary tool to make current breathing exercises more compelling and accessible.	12	6 (9)	2 (3)
Structural within treatment	Using DEEP as a structural tool at a fixed moment (in treatment), where patients can train their deep breathing and emotion regulation, without feeling tense.	11	5 (7)	3 (4)

[a]The total number of times a code was mentioned in all focus groups and interviews. [b]The number of different healthcare providers that mentioned a code, and (#) the total number of times the code was found in all focus groups. [c]The number of patients that mentioned a code, and (#) the number of times the code was found in all interviews with patients.

intervention that can provide direct support to patients when they experience anger, stress, or anxiety. When using this approach, it might be possible to decrease these types of negative emotions in a short period of time. A patient illustrates this in the following quote: *"I would like to use DEEP when I need it, for example that I make a reservation for DEEP for 1 h. Because if I use DEEP when I do not need it and can't use DEEP when I do, I might feel like crap the rest of the week."*

Second, six healthcare providers and two patients indicated that DEEP can *complement already existing breathing exercises* that are provided in current treatment. More clinics use breathing exercises in their treatment plan, however many patients are hesitant to participate, because they find it too abstract or boring. According to the healthcare providers, DEEP could help to introduce deep breathing to patients in an engaging and accessible way.

Finally, five healthcare providers mentioned that DEEP could be used structurally, which meant multiple times per week on fixed moments. They stated that when using DEEP structurally, the patient can train their deep breathing without feeling tense or emotional, while learning to use their deep breathing as a coping skill when stressful situations do occur. The following quote of participant 6 illustrates this: *"It would be nice to use DEEP structurally. Then you are able to start with arousal and stress regulation. And it is something to put in the treatment plan. It means that someone who might feel OK at the moment is able to practice with DEEP. When someone feels that their tension is high, you are able to rely on the things and skills you have built upon together."* Three patients agreed with the structural use of DEEP. They stated that using DEEP on fixed moments during the week or within therapy might give them a sense of support and structure.

3.7. Conditions for implementing DEEP within forensic mental healthcare

Finally, healthcare providers and patients were asked to provide possible conditions that should be considered when implementing DEEP within forensic mental healthcare. The identified codes are provided in Table 6.

First, 11 healthcare providers mentioned that it is important that *patients receive support* and instructions from their healthcare providers, especially when they use DEEP for the first time. Many healthcare providers stated that it is important to increase the confidence of patients while using DEEP, so that they feel it can help them in their treatment goals. Five patients indicated that DEEP can feel overwhelming: receiving some instructions before and support during DEEP might help them to stay calm. This is illustrated in the following quote by a patient: *"I would like some instructions and support. I feel like I might enjoy it even more, (. . .). But sometimes it is good to try something out and fail, this is how life works as well. But with some instructions I might have enjoyed it more. I also think I might have had a different kind of breathing throughout the whole game."*

Second, nine healthcare providers and four patients indicated the importance of using DEEP in a quiet *room with no distractions*. Some healthcare providers mentioned that it is important to use DEEP in private, because many patients are easily distracted or feel

TABLE 6 Possible conditions for implementing DEEP in forensic mental healthcare, according to healthcare providers (n = 24) and patients (n = 13).

Codes	Definition of code	Codes[a]	Healthcare providers[b]	Patients[c]
Implementing DEEP				
Support for patients	Healthcare providers must be able to support and instruct their patients, both before, during and after using DEEP, so patients can gain the confidence to discover DEEP	17	11 (12)	5 (5)
Room to play DEEP with no distractions	An available room must be provided that has a closing door and curtains, so patients can use DEEP undisturbed	16	9 (12)	4 (4)
Training for healthcare providers	A proper DEEP-training must be provided to all healthcare providers, so that they have the skillset to use DEEP with their patients (in their current treatment)	6	5 (6)	
Clear agreements on the usage of DEEP	Before implementing DEEP, healthcare providers and management need to make clear agreements on how, when and for whom DEEP will be used	5	4 (4)	1 (1)
Locks to prevent theft or vandalism	The VR-area should be secured in preventing theft or vandalism of the VR-set(s)	5	2 (2)	3 (3)
Good (spinning) desk chairs	To use DEEP optimally a good desk chair must be provided that has wheels and can rotate	4	3 (3)	1 (1)
Time to discover DEEP together	Healthcare providers need time and room in their agenda to discover together how DEEP can be used most effectively	1	1 (1)	
User support for healthcare providers	Healthcare providers need a helpdesk or support line where they can get information or help when using DEEP	1	1 (1)	

[a]The total number of times a code was mentioned in all focus groups and interviews. [b]The number of different healthcare providers that mentioned a code, and (#) the total number of times the code was found in all focus groups. [c]The number of patients that mentioned a code, and (#) the number of times the code was found in all interviews with patients.

uncomfortable by their peers. Participant 2 provided the following quote illustrating this topic: *"Well, I think that you should take privacy into account. Because when you use a VR-headset you are kind of in your own world. Many boys in our group were bullied and are hesitant to really emerge themselves, so we must provide room for them to do so."*

Third, five healthcare providers mentioned the need for *training for all healthcare providers*, so that everyone in the clinic has the confidence and the skills to use DEEP with their patients. Fourth, four healthcare providers indicated that before implementing DEEP, management and the healthcare providers should make *clear agreements* to ensure that everyone knows why and for which type of patient groups DEEP is implemented, and in what way it can be used in inpatient care. These agreements should be visible for all colleagues to ensure that no clinic is hesitant to start using DEEP. One patient gave the following statement regarding this topic: *"All these interventions are good but should also be concrete. Sometimes you can talk and talk and talk, and still, it takes forever before something is available and useful. And if something is available, it is put away, because someone has to stay with the patient, and no one has time for it. These things should be taken into account, because there are many great projects, but the whole system you put it in, is really bureaucratic."*

Fifth, two healthcare providers and three patients stated that it is necessary to make sure the *VR-area with DEEP can be secured safely*. Some of them shared their experience with new equipment that got stolen or vandalized. Sixth, three healthcare providers and one patient mentioned the need for a *good desk chair*, so that the user can move through DEEP without straining their neck. Seventh, one healthcare provider indicated the importance of providing *time in their agenda* to introduce DEEP to their patients and discover the intervention at their own pace. Finally, one healthcare provider shared the need for *user support* for when they need information or have questions regarding DEEP.

4. Discussion

4.1. Answering the research questions

The aim of this study was to gain insight in the possible added value of DEEP for treatment of forensic psychiatric inpatients. First, several possible advantages of using DEEP in forensic psychiatric inpatient clinics were identified. First, participants especially valued the promise DEEP has for engaging unmotivated patients, by using a gamified approach and an appealing design. Moreover, by its focus on the experience in a VR-environment with no use of written text or assignments, participants expected DEEP to resonate with patients with suboptimal cognitive skills. However, the results also showed some points of improvement of DEEP. Participants indicated that the hardware of DEEP felt in some ways suboptimal: the VR-controls and -headset showed some malfunction and multiple wire connections were necessary before using DEEP. Additionally, some design features of DEEP might not fit the skills of forensic patients, according to the participants. They found that DEEP uses a straight-forward design, but still requires multitasking from its user while playing. Participants wondered if this could lead to frustration in some forensic inpatients. Second, most healthcare providers concluded that DEEP might be useful for any type of patient if they are willing to have an open mind to the effects of diaphragmatic breathing on their wellbeing. Some patients' characteristics, like psychotic vulnerability, were mentioned as potentially unsuitable for using DEEP by several participants, due to its surrealistic and immersive design. Patients

were less able to provide insight into specific patient groups that might benefit from using DEEP because they were mostly preoccupied with their own treatment. Third, participants shared several views about how DEEP could be integrated into current forensic practice. On the one hand, DEEP could be used on structurally (e.g., on fixed days) aimed at the acquisition of skills, such as using deep breathing to cope with negative emotions. On the other hand, DEEP could be used *ad hoc* as a short-term relief of anxiety or anger or as an intervention for when patients are preparing for a challenging task during their treatment. Finally, many practical preconditions were given for implementing DEEP in forensic mental healthcare. For the implementation of DEEP to be successful, our study emphasized the needs of patients and healthcare providers, such as practical support, sharing of information, clear and concise instructions, and scheduled time to include DEEP in their daily practice.

4.2. The added value of DEEP in forensic mental healthcare

One of the main findings of this study is that healthcare providers and patients stated that DEEP has the potential to add something new and unique to the forensic mental healthcare. In an earlier study, DEEP was shown to reduce state-anxiety, providing a relaxed state that remained around 2 h (Bossenbroek et al., 2020), which was underlined by the expectations of our participants. Participants indicated that several working mechanisms of DEEP that were found to be of possible added value can also be found in other eHealth technologies, such as mindfulness and deep breathing. Consequently, apps might be a cheaper alternative for VR interventions such as DEEP. Research has indeed shown that mindfulness apps that focus on breathing exercises and meditation also show promising results in treating forensic psychiatric patients (Bostock et al., 2019; Walsh et al., 2019). Additionally, the study conducted by Weerdmeester et al. (2021) found regular breathing apps as effective as DEEP for undergraduates, while remaining free and easy to use. Our study showed that while apps seem to be fitting as well, DEEP seems to be more suitable for a forensic setting: both healthcare providers and patients stated that mindfulness apps often focused too much on spirituality and still require reflective and reading skills of its user. Other studies found this as well, stating that many apps require cognitive and digital skills and are based on inwardness, reflection, and introspection (Howells et al., 2010; Kip et al., 2019c; Kip and Bouman, 2020). Even though some breathing apps are less language-based and less focused on spirituality, especially unmotivated forensic psychiatric patients might need more than just an app to feel motivated and engaged to work on their deep breathing. This expectation is strengthened by the study of Weerdmeester et al. (2021), in which it was found that the engagement of DEEP users increased, while no change in engagement was found in users of a breathing app. By using an experience-based and gamified approach, DEEP might offer a unique and engaging way to introduce principles of mindfulness and diaphragmatic breathing to forensic inpatients. However, it is important to further evaluate DEEP to study whether DEEP can be effective for the forensic inpatient care, confirming the expectations of this study. Moreover, future research could explore if mindfulness apps could provide an effective, fitting, and cost-efficient alternative to DEEP in treatment of complex patient populations.

4.3. Usage of DEEP: structural vs. ad hoc

Some healthcare providers and patients would like to use DEEP structurally to train breathing skills, for example on fixed days. Others preferred to use DEEP *ad hoc*, on moments where patients felt they could benefit from DEEP. When used structurally, DEEP can be used to teach forensic patients deep breathing skills as a coping-strategy, allowing them to deploy these strategies when they feel anxious or angry. In this way, patients can acquire the deep breathing skills on good days when they have enough mental space to learn, ensuring that on bad moments, they can cope with their negative emotions. Regarding the *ad hoc* approach, DEEP is not used during regular treatment sessions, but can be used on the patients' initiative when they already feel elevated levels of negative emotions such as anger, stress or anxiety and need a way to escape and relax. This approach might increase the sense of self-management and ownership of patients, as they can decide when they would like to use DEEP and if they would like to use it alone or with their healthcare provider. While forensic psychiatric inpatients are often treated involuntarily, empowering them to take an active role in their care seems important for it to be effective (Senneseth et al., 2022). An increase of ownership and self-management might also motivate and engage patients to use DEEP as an intervention, making it important to adapt it to the needs of patients in helping them reach their treatment goals (O'Leary et al., 2015). In conclusion, DEEP is an intervention that can be used in diverse ways, depending on the needs and preferences of patients. Future research is needed to explore the efficacy and feasibility of both structural and *ad hoc* ways of using DEEP.

4.4. Implementing DEEP in forensic mental healthcare

This study showed the importance of a structural, thorough implementation of DEEP. This is in line with multiple studies that show that the implementation of eHealth is often complex and

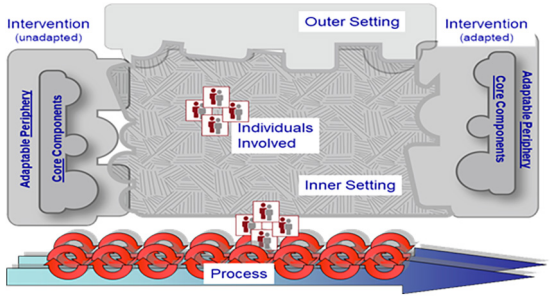


FIGURE 3
The Consolidated Framework for Implementation Research (CFIR). Reproduced with permission from the CFIR Research Team-Center for Clinical Management Research (https://cfirguide.org).

requires a thorough approach (Schreiweis et al., 2019; Kouijzer et al., 2023). Both healthcare providers and patients provided insight into the factors that are expected to be important for successful implementation of DEEP in forensic inpatient clinics. These factors belong to various levels, ranging from individual patients to the organization. The identified factors are in line with the domains of the *Consolidated Framework of Implementation Research*, which has been used often within healthcare settings (Damschroder et al., 2009, 2022). The CFIR is visually displayed in **Figure 3**. In this study, multiple advantages and barriers were mentioned about the design and working mechanisms of DEEP regarding the *intervention*, (e.g., the appealing but overwhelming design of DEEP) which might affect its long-term uptake. Our findings are in line with earlier studies that highlight the importance of overcoming a "one-size-fits-all approach" and tailoring the design of an intervention to its end-users (Polaschek, 2011; Wilson et al., 2017). Therefore, it seems to be important to further develop and adjust the hardware and software of DEEP to fit the needs and skills of forensic inpatients and their healthcare providers. This study also suggests the importance of including the *inner setting* and its *individuals* in the implementation process by not only explore the needs of the end-users (e.g., patients), but also those of managers and healthcare providers. This is important because they are the gateway keepers when introducing new technologies to their patients. Other studies point out the importance of involving a range of stakeholders in implementation as well, to account for all relevant perspectives (Marcu et al., 2011). Additionally, the results of this study show the importance of providing practical and content-related support to the healthcare providers and patients during the implementation *process*. This is supported by earlier research that emphasizes the importance holistic implementation approach, with attention to the organizations, patients, and healthcare providers (Kip et al., 2020b). Therefore, before starting the implementation process of DEEP, all needs of the organization, its employees and the patients should be considered to prevent practical barriers, such as the lack of a room for DEEP, to hinder the implementation. According to the CFIR framework, the role of the *outer setting* is essential for implementation, such as beneficial policies and (financial) incentives by government organizations and other external stakeholders to increase the uptake of innovative technologies (Kirk et al., 2016; Ross et al., 2016). The factors related to the outer setting were hardly identified in this study, underlining the need for future research to paint a more comprehensive picture of the wider context surrounding the organization (Christie et al., 2018). Finally, a recent study on the implementation of VR in healthcare highlights the importance of creating a systematic implementation plan, with concrete strategies and objectives linked to clear implementation outcomes (Kouijzer et al., 2023). In our study, participants clearly identified the importance of practical resources, such as room, time and support for caregivers to use DEEP with their patients. In future research, it remains necessary to further study DEEP in forensic psychiatric care to thoroughly and systematically investigate the process of implementing DEEP. Ideally, a comprehensive, multi-level implementation plan with concrete implementation strategies to address possible implementation obstacles has to be created. In conclusion, this study provides a starting point for implementation by identifying expected barriers and facilitators, and thus serves as a first step for future implementation research.

4.5. Ethical considerations

When conducting research in forensic psychiatric settings, attention must be paid to the informed consent of forensic psychiatric patients, as most of them receive mandatory treatment (Ligthart et al., 2022a,b; Meynen, 2022). Especially due to the obligatory nature of treatment, it is essential that patients are able to voluntarily provide a fully informed consent. To achieve this, patients who were interested in participating were given an information letter and were verbally informed about DEEP, its mechanisms and the purpose of this study. Specific attention was paid to providing information in a way that was suitable for their cognitive and literacy skills. Furthermore, engaging and immersive VR environments can become triggering when being used in forensic mental healthcare (Kip et al., 2019b). In DEEP, users do emerge themselves in a surrealistic, dreamlike environment. While in this study, no negative impact of such an environment was identified, it might cause triggering effects for some patients. This highlights the importance of continuously paying attention to the experiences of patients that use DEEP, especially to those who are sensitive to light, sounds and psychoses, the inclusion of caregivers and patients in the validation and further development of DEEP.

4.6. Strengths and limitations

The main strength of this study is the involvement of both healthcare providers and patients. Both perspectives are important, since both groups are stakeholders of new interventions, while healthcare providers remain often overlooked (Kip et al., 2019a,b). However, DEEP was not available for participants to use in the clinics. This study aimed to overcome this limitation by letting both healthcare providers and patients tryout DEEP before starting with the focus groups and interviews, showing the importance of using concrete products and user-experiences (Beerlage-de Jong et al., 2017; Kip et al., 2022). Even though many healthcare providers and patients stated that they found it useful to try DEEP before participating in the focus groups and interviews, some participants still felt some hesitation to provide definite answers on how and for whom DEEP is of most added value, due to them not being able to use DEEP in their current practice.

In this study, participants volunteered to be included, which can point to selection bias: there is a risk that these participants are already more interested in using new technological interventions, which might not thoroughly represent the forensic inpatient care. We have accounted for this by elaborately focusing on possible disadvantages of DEEP, paying attention to which parts of the intervention might not be suitable for forensic psychiatric inpatients, and focusing on barriers for implementation.

During the interviews it became clear that some patients had trouble with thinking about possible ways DEEP could be of use in their treatment, because of difficulties with abstract reasoning. Therefore, the researcher sometimes decided to shorten the interview by leaving out the more abstract questions about implementation, and by putting more focus on the patient's first impression of DEEP and on how and when they would prefer to

use it, making it more concrete. Adapting the methodology can be seen as good practice in this type of study, to prioritize the wellbeing of the end user (Kip et al., 2022; Schouten et al., 2022). However, this could mean that the perspective of some patients is not fully included. Additionally, due to privacy reasons we did not ask the participants about sociodemographic information such as age, psychiatric diagnoses, or reason for treatment. However, including this information might have resulted in more interesting insights into the way the participants experienced their first impression of DEEP and how they feel DEEP might help them in treatment.

Finally, an earlier version of the game was used in this study, which means that some aspects that may have led to frustration or sub-optimal user experiences have already been improved. Regarding the hardware of DEEP: DEEP is now available on the *Oculus Quest 2* headset and the waist belt and hand controllers are now wireless. However, it is only possible to cast DEEP to a laptop via Wi-Fi connection, making it more complex for healthcare providers to support patients during the game. DEEP is still under development, which means that the hardware and software keeps being expanded and improved. However, this study did underline the importance of keeping its user engaged, while remaining user-friendly and not too overwhelming. This needs to be addressed in future versions of the game to make it suitable for use in a forensic setting.

5. Conclusion

With its experience-based and gamified design, DEEP could be of added value for forensic mental healthcare, particularly by providing an appealing and engaging way to teach deep breathing to forensic psychiatric inpatients with low treatment motivation. According to patients and healthcare provider, DEEP can be used both structurally within treatment and *ad hoc* on patients' initiative. Our study showed that it is important to include patients and healthcare providers in the evaluation and implementation of DEEP from the start. To account for all relevant perspectives and to ensure successful integration in current care pathways, there is a need for a multi-level approach when implementing DEEP. This approach should consider the intervention, the wider forensic context, the ethical considerations and the individuals that are possible end-users of DEEP. These factors can be used to formulate clear implementation strategies, such as a practical DEEP-training and clear protocols for caregivers on when, how and for whom DEEP is to be used, as well as to further evaluate DEEP in forensic mental healthcare. If implemented well, DEEP can offer new, experience-based ways to provide forensic psychiatric patients with strategies to better cope with their negative emotions, such as stress and anger and thus prevent recidivism.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Twente (Behavioral, Management and Social Sciences, nr. 21119). The patients/participants provided their written informed consent to participate in this study.

Author contributions

HK and LK contributed to the design and planning of the study and analyzed the focus group and interview data. LK collected the focus group and interview data. All authors contributed to the interpretation and reporting of the results and approved the manuscript.

Funding

This study was supported by a grant from Kwaliteit Forensische Zorg to HK (KFZ 2020-146).

Acknowledgments

The authors would like to thank the healthcare professionals and patients of Stichting Transfore and De Woenselse Poort for participating in this study. The authors would especially like to thank Jente van Rijn for her support and assistance during this study.

Conflict of interest

JW has been affiliated with Explore DEEP Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahmadpour, N., Keep, M., Janssen, A., Rouf, A. S., and Marthick, M. (2020). Design strategies for virtual reality interventions for managing pain and anxiety in children and adolescents: Scoping review. *JMIR Serious Games* 8:e14565. doi: 10.2196/14565

Beerlage-de Jong, N., Wentzel, J., Hendrix, R., and van Gemert-Pijnen, L. (2017). The value of participatory development to support antimicrobial stewardship with a clinical decision support system. *Am. J. Infect. Control* 45, 365–371. doi: 10.1016/j.ajic.2016.12.001

Bloem, O., Nijman, H., and Bulten, B. (2011). "Psychopathologie onder gedetineerden. [Psychopathology in detainees]," in *Handboek forensische geestelijke gezondheidszorg. [Handbook forensic mental healthcare]*, eds H. Groen, M. Drost, and H. L. I. Nijman (Utrecht: De Tijdstroom), 153–162.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Qual. Quant.* 36, 391–409. doi: 10.1023/A:1020909529486

Bossenbroek, R., Wols, A., Weerdmeester, J., Lichtwarck-Aschoff, A., Granic, I., and Van Rooij, M. M. (2020). Efficacy of a virtual reality biofeedback game (DEEP) to reduce anxiety and disruptive classroom behaviour: Single-case study. *JMIR Ment. Health* 7:e16066. doi: 10.2196/16066

Bostock, S., Crosswell, A. D., Prather, A. A., and Steptoe, A. (2019). Mindfulness on-the-go: Effects of a Mindfulness meditation app on work stress and well-being. *J. Occup. Health Psychol.* 24, 127–138. doi: 10.1037/ocp0000118

Bree, R. T., and Gallagher, G. (2016). Using Microsoft Excel to code and thematically analyse qualitative data: A simple, cost-effective approach. *All Ireland J. High. Educ.* 8, 2811–2824.

Christie, H. L., Bartels, S. L., Boots, L. M. M., Tange, H. J., Verhey, F. J. J., and De Vugt, M. E. (2018). A systematic review on the implementation of eHealth interventions for informal healthcare providers of people with dementia. *Internet Interv.* 13, 51–59. doi: 10.1016/j.invent.2018.07.002

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., and Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implement. Sci.* 4, 1–15. doi: 10.1186/1748-5908-4-50

Damschroder, L. J., Reardon, C. M., Widerquist, M. A. O., and Lowery, J. (2022). The updated Consolidated Framework for Implementation Research based on user feedback. *Implement. Sci.* 17:75. doi: 10.1186/s13012-022-01245-0

Del Vecchio, T., and O'Leary, K. D. (2004). Effectiveness of anger treatments for specific anger problems: A meta-analytic review. *Clin. Psychol. Rev.* 24, 15–34. doi: 10.1016/j.cpr.2003.09.006

Delfin, C., Krona, H., Andiné, P., Ryding, E., Wallinius, M., and Hofvander, B. (2019). Prediction of recidivism in a long-term follow-up of forensic psychiatric patients: Incremental effects of neuroimaging data. *PLoS One* 14:e0217127. doi: 10.1371/journal.pone.0217127

Drieschner, K. H., and Boomsma, A. (2008). The treatment motivation scales for forensic outpatient treatment (tms-f) construction and psychometric evaluation. *Assessment* 15, 224–241. doi: 10.1177/1073191107311650

Edberg, H., Chen, Q., Andiné, P., Larsson, H., and Hirvikoski, T. (2022). Criminal recidivism in offenders with and without intellectual disability sentenced to forensic psychiatric care in Sweden—A 17-year follow-up study. *Front. Psychiatry* 13:1011984. doi: 10.3389/fpsyt.2022.1011984

Gillespie, S. M., Mitchell, I. J., Fisher, D., and Beech, A. R. (2012). Treating disturbed emotional regulation in sexual offenders: The potential applications of mindful self-regulation and controlled breathing techniques. *Aggress. Violent Behav.* 17, 333–343. doi: 10.1016/j.avb.2012.03.005

Gray, E., Beech, A., and Rose, J. (2019). Using biofeedback to improve emotion regulation in sexual offenders with intellectual disability: A feasibility study. *Int. J. Dev. Disabil.* 65, 195–204. doi: 10.1080/20473869.2018.1565003

Henwood, K. S., Chou, S., and Browne, K. D. (2015). A systematic review and meta-analysis on the effectiveness of CBT informed anger management. *Aggress. Violent Behav.* 25, 280–292. doi: 10.1016/j.avb.2015.09.011

Howells, K., Tennant, A., Day, A., and Elmer, R. (2010). Mindfulness in forensic mental health: Does it have a role? *Mindfulness* 1, 4–9. doi: 10.1007/s12671-009-0001-x

Ilioudi, M., Lindner, P., Ali, L., Wallström, S., Thunström, A. O., Ioannou, M., et al. (2023). Physical versus virtual reality–based calm rooms for psychiatric inpatients: Quasi-randomized trial. *J. Med. Internet Res.* 25:e42365. doi: 10.2196/42365

Kim, M., Jeon, C., and Kim, J. (2017). A study on immersion and presence of a portable hand haptic system for immersive virtual reality. *Sensors* 17:1141. doi: 10.3390/s17051141

Kip, H., and Bouman, Y. H. A. (2020). Van denken en praten naar doen en ervaren: Op zoek naar de toegevoegde waarde van technologie voor de forensische psychiatrie. [From thinking and talking to doing and experiencing: Searching for the added value of technology for forensic mental healthcare]. *Sancties* 4, 249–260.

Kip, H., and Bouman, Y. H. A. (2021). A perspective on the integration of eHealth in treatment of offenders: Combining technology and the risk-need-responsivity model. *Front. Psychiatry* 12:703043. doi: 10.3389/fpsyt.2021.703043

Kip, H., Bouman, Y. H. A., Kelders, S. M., and Van Gemert-Pijnen, L. J. E. W. C. (2018). eHealth in treatment of offenders in forensic mental health: A review of the current state. *Front. Psychiatry* 9:42. doi: 10.3389/fpsyt.2018.00042

Kip, H., Keizer, J., Da Silva, M. C., Beerlage-de Jong, N., Köle, N., and Kelders, S. M. (2022). Methods for human-centered eHealth development: Narrative scoping review. *J. Med. Internet Res.* 24:e31858. doi: 10.2196/31858

Kip, H., Kelders, S. M., Bouman, Y. H. A., and Van Gemert-Pijnen, L. J. E. W. C. (2019a). The importance of systematically reporting and reflecting on eHealth development: Participatory development process of a virtual reality application for forensic mental health care. *J. Med. Internet Res.* 21:e12972. doi: 10.2196/12972

Kip, H., Kelders, S. M., Weerink, K., Kuiper, A., Brüninghoff, I., Bouman, Y. H. A., et al. (2019b). Identifying the added value of virtual reality for treatment in forensic mental health: A scenario-based qualitative approach. *Front. Psychol.* 10:406. doi: 10.3389/fpsyg.2019.00406

Kip, H., Oberschmidt, K., and Bierbooms, J. J. (2020a). Ehealth technology in forensic mental healthcare: Recommendations for achieving benefits and overcoming barriers. *Int. J. Forensic Ment. Health* 20, 31–47. doi: 10.1080/14999013.2020.1808914

Kip, H., Oberschmidt, K., Bierbooms, J., Dijkslag, D., Kelders, S., and Roelofsen, B. (2019c). *Technologie in de forensische zorg-crossing borders. [Technology in forensic mental healthcare – crossing borders]*. Utrecht: Kwaliteit Forensische Zorg.

Kip, H., Sieverink, F., Van Gemert-Pijnen, L. J. E. W. C., Bouman, Y. H. A., and Kelders, S. M. (2020b). Integrating people, context, and technology in the implementation of a web-based intervention in forensic mental health care: Mixed-methods study. *J. Med. Internet Res.* 22:e16906. doi: 10.2196/16906

Kirk, M. A., Kelley, C., Yankey, N., Birken, S. A., Abadie, B., and Damschroder, L. (2016). A systematic review of the use of the consolidated framework for implementation research. *Implement. Sci.* 11:72. doi: 10.1186/s13012-016-0437-z

Kirschstein, M. A., Singh, J. P., Rossegger, A., Endrass, J., and Graf, M. (2023). International survey on the use of emerging technologies among forensic and correctional mental health professionals. *Crim. Justice Behav.* 50, 175–196. doi: 10.1177/00938548211042057

Klein Tuente, S., Bogaerts, S., Bulten, E., Keulen-de Vos, M., Vos, M., Bokern, H., et al. (2020). Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: A multicenter randomized controlled trial. *J. Clin. Med.* 9:2258. doi: 10.3390/jcm9072258

Kothgassner, O. D., Reichmann, A., and Bock, M. M. (2023). Virtual reality interventions for mental health. *Curr. Top. Behav. Neurosci.* 65, 371–387. doi: 10.1007/7854_2023_419

Kouijzer, M. M., Kip, H., Bouman, Y. H., and Kelders, S. M. (2023). Implementation of virtual reality in healthcare: A scoping review on the implementation process of virtual reality in various healthcare settings. *Implement. Sci. Commun.* 4, 1–29. doi: 10.1186/s43058-023-00442-2

Ligthart, S., Meynen, G., and Thomas, D. (2022a). "Persuasive technologies and the right to mental liberty: The 'smart' rehabilitation of criminal offenders," in *The Cambridge handbook of information technology, life sciences and human rights*, eds M. Ienca, O. Pollicino, L. Liguori, E. Stefanini, and R. Andorno (Cambridge: Cambridge University Press). doi: 10.1017/9781108775038.005

Ligthart, S., Meynen, G., Biller-Andorno, N., Kooijmans, T., and Kellmeyer, P. (2022b). Is virtually everything possible? The relevance of ethics and human rights for introducing extended reality in forensic psychiatry. *AJOB Neurosci.* 13, 144–157. doi: 10.1080/21507740.2021.1898489

Marcu, G., Bartram, J., and Gabrielli, S. (2011). "a framework for overcoming challenges in designing persuasive monitoring and feedback systems for mental illness," in *Proceedings of the 5th international conference on pervasive computing technologies for healthcare (PervasiveHealth) and workshops*, (Dublin), 1–8. doi: 10.4108/icst.pervasivehealth.2011.246097

Meynen, G. (2017). Walls and laws: Structural barriers to forensic psychiatric research. *Eur. Psychiatry* 44, 208–209. doi: 10.1016/j.eurpsy.2017.04.010

Meynen, G. (2022). Technologie in de forensische psychiatrie: Het belang van ethisch en juridisch onderzoek (Van redactiewege). *Sancties* 4:205.

Miles, M. B., and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications.

O'Leary, K., Vizer, L., Eschler, J., Ralston, J., and Pratt, W. (2015). Understanding patients' health and technology attitudes for tailoring self-management interventions. *AMIA Annu. Symp. Proc.* 2015, 991–1000.

Phillips, L. A., Fritz, J. N., Rettig, L. A., and Martin, K. E. (2019). Diaphragmatic breathing as treatment for escape-maintained aggression. *Behav. Interv.* 34, 219–230. doi: 10.1002/bin.1663

Pillai, A. S., and Mathew, P. S. (2019). "Impact of virtual reality in healthcare," in *Virtual and augmented reality in mental health treatment*, Vol. 17, ed. G. Guazzaroni (Hershey, PA: IGI Global). doi: 10.4018/978-1-5225-7168-1.ch002

Polaschek, D. L. L. (2011). Many sizes fit all: A preliminary framework for conceptualizing the development and provision of cognitive-behavioral rehabilitation programs for offenders. *Aggress. Violent Behav.* 16, 20–35. doi: 10.1016/j.avb.2010.10.002

Probst, T., Bezzel, A., Hochstadt, M., Pieh, C., and Mache, W. (2020). Criminal recidivism after forensic psychiatric treatment. A multicenter study on the role of pretreatment, treatment-related, and follow-up variables. *J. Forensic Sci.* 65, 1221–1224. doi: 10.1111/1556-4029.14281

Rizzo, A., Koenig, S. T., and Talbot, T. (2018). Clinical virtual reality: Emerging opportunities for psychiatry. *Focus* 16, 266–278. doi: 10.1176/appi.focus.20180011

Ross, J., Stevenson, F., Lau, R., and Murray, E. (2016). Factors that influence the implementation of e-health: A systematic review of systematic reviews (an update). *Implement. Sci.* 11:146. doi: 10.1186/s13012-016-0510-7

Saini, M. (2009). A meta-analysis of the psychological treatment of anger: Developing guidelines for evidence-based practice. *J. Am. Acad. Psychiatry Law Online* 37, 473–488.

Savard, J. S. (2017). *Reducing adolescent anger and aggression with biofeedback: A mixed-methods multiple case study*. Dissertations. Seattle, WA: Antioch University Seattle.

Schouten, S., Kip, H., Dekkers, T., Deenik, J., Beerlage - de Jong, N., Ludden, G. D. S., et al. (2022). Best-practices for co-design processes involving people with severe mental illness for eMental health interventions: A qualitative multi-method approach. *Design Health* 6, 316–344. doi: 10.1080/24735132.2022.2145814

Schreiweis, B., Pobiruchin, M., Strotbaum, V., Suleder, J., Wiesner, M., and Bergh, B. (2019). Barriers and Facilitators to the Implementation of eHealth Services: Systematic Literature Analysis. *J. Med. Internet Res.* 21:e14197. doi: 10.2196/14197

Senneseth, M., Pollak, C., Urheim, R., Logan, C., and Palmstierna, T. (2022). Personal recovery and its challenges in forensic mental health: Systematic review and thematic synthesis of the qualitative literature. *BJPsych Open* 8:E17. doi: 10.1192/bjo.2021.1068

Smeijers, D., Bulten, E. H., Verkes, R. J., and Koole, S. L. (2021). Testing the effects of a virtual reality game for aggressive impulse management: A preliminary randomized controlled trial among forensic psychiatric outpatients. *Brain Sci.* 11:1484. doi: 10.3390/brainsci11111484

Stoll, J., Müller, J. A., and Trachsel, M. (2020). Ethical issues in online psychotherapy: A narrative review. *Front. Psychiatry* 10:993. doi: 10.3389/fpsyt.2019.00993

Svensson, I., Fälth, L., and Persson, B. (2015). Reading level and the prevalence of a dyslexic profile among patients in a forensic psychiatric clinic. *J. Forensic Psychiatry Psychol.* 26, 532–550. doi: 10.1080/14789949.2015.1037329

Sygel, K., and Wallinius, M. (2021). Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: A review of current assessment and treatment methods for practitioners. *Front. Psychiatry* 12:673089. doi: 10.3389/fpsyt.2021.673089

Tellhed, U., Daukantaitë, D., Maddux, R. E., Svensson, T., and Melander, O. (2019). Yogic breathing and mindfulness as stress coping mediate positive health outcomes of yoga. *Mindfulness* 10, 2703–2715. doi: 10.1007/s12671-019-01225-4

Timmerman, I. G. H., and Emmelkamp, P. M. G. (2005). The effects of cognitive-behavioral treatment for forensic inpatients. *Int. J. Offender Ther. Comp. Criminol.* 49, 590–606. doi: 10.1177/0306624X05277661

Tong, A., Sainsbury, P., and Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *Int. J. Qual. Healthcare* 19, 349–357. doi: 10.1093/intqhc/mzm042

Torous, J., Myrick, K. J., Rauseo-Ricupero, N., and Firth, J. (2020). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Ment. Health* 7:e18848. doi: 10.2196/18848

Tossaint-Schoenmakers, R., Versluis, A., Chavannes, N., Talboom-Kamp, E., and Kasteleyn, M. (2021). The challenge of integrating ehealth into health care: Systematic literature review of the Donabedian model of structure, process, and outcome. *J. Med. Internet Res.* 23:e27180. doi: 10.2196/27180

Van Rooij, M., Lobel, A., Harris, O., Smit, N., and Granic, I. (2016). "DEEP: A biofeedback virtual reality game for children at-risk for anxiety," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, (New York, NY: Association for Computing Machinery), 1989–1997. doi: 10.1145/2851581.2892452

Walsh, K. M., Saab, B. J., and Farb, N. A. (2019). Effects of a mindfulness meditation app on subjective well-being: Active randomized controlled trial and experience sampling study. *JMIR Ment. Health* 6:e10844. doi: 10.2196/10844

Weerdmeester, J. W. (2021). *Take a DEEP Breath: Exploring the Potential of Game-based Biofeedback Interventions for Anxiety Regulation*. Dissertation. Nijmegen: Radboud Universiteit.

Weerdmeester, J. W., Van Rooij, M. M. J. W., Maciejewski, D. F., Engels, R. C. M. E., and Granic, I. (2021). A randomized controlled trial assessing the efficacy of a virtual reality biofeedback video game: Anxiety outcomes and appraisal processes. *Technol. Mind Behav.* 2, 1–16. doi: 10.1037/tmb0000028

Wilson, H. J., Palk, G., Sheehan, M. C., Wishart, D., and Watson, B. (2017). Steering clear of driving after drinking: A tailored e-health intervention for reducing repeat offending and modifying alcohol use in a high-risk cohort. *Int. J. Behav. Med.* 24, 694–702. doi: 10.1007/s12529-017-9664-1

10.3389/fpsyg.2023.1201485

Appendix

Appendix A Semi-structured topic list for focus groups with healthcare providers.

To what extent is DEEP useful for your patient population?	RQ1
What kind of benefits could it have?	RQ1
What do you see as the potential added value of DEEP?	RQ1
Why do you expect DEEP to work?	RQ1
What downsides or difficult issues do you see? Any barriers?	RQ1
For what types of patients would you want to use this, and when?	RQ2
Are there also patients for whom you would not want to use DEEP?	RQ2
If we want to use DEEP in the clinic, what should we take into account?	RQ2-4
○ Are there certain things related to patient characteristics? ○ What do you need? How would you like to be supported? ○ What should the organization do? ○ Are there certain things in DEEP itself that need to be clearer/better? ○ Do you have any suggestions for improvement?	
What else would you like to know about DEEP? What are you curious about?	
Are there any other questions, comments, or ideas that are important for us to consider?	

Appendix B Semi-structured interview scheme for interviews with patients.

Would you want to use DEEP yourself? Why or why not?	RQ1
What benefits do you think DEEP could have for you?	RQ1
How could it help you? And for other patients admitted here?	RQ1-2
What possible downsides of DEEP do you see? What would be difficult, hard or not fun?	RQ1
When would you prefer to use DEEP? At what moments or periods?	RQ3
And when would you really not want to use DEEP?	RQ3
Imagine that we would use DEEP at your clinic. What should we take into account?	RQ4
○ What kind of information would you like to have before using DEEP? ○ How would you like to be helped when using DEEP? ○ What do you think is important for your care providers? How would you like to involve them? ○ What do you expect from us as researchers? ○ If you look at DEEP now, are there certain things you would like to see differently or better? Do you have any tips to improve it further?	
Do you have any further comments, questions or ideas?	

Appendix C Included healthcare providers and patients per forensic organization (N = 24; N = 13).

	Healthcare providers	Patients
Transfore	14	4
FPK	3	2
FPA	3	2
Forence	8	0
De Woenselse Poort	10	9
Balans	9	1
Keer	0	2*
Verbinding	1	6
Total	**24**	**13**

*At this clinic no focus groups were organized. However, during the introduction of DEEP and interviews with patients of other clinics two patients who received treatment at Keer saw the VR-set and spontaneously asked to be included themselves.

frontiers | Frontiers in Psychology

TYPE Study Protocol
PUBLISHED 04 April 2024
DOI 10.3389/fpsyg.2024.1324644

OPEN ACCESS

EDITED BY
Lisa Wagels,
University Hospital RWTH Aachen, Germany

REVIEWED BY
Elizabeth A. Boyle,
University of the West of Scotland,
United Kingdom
Cyril Boonmann,
University Psychiatric Clinic Basel, Switzerland

*CORRESPONDENCE
Bas R. van Wolffelaar
bvanwolffelaar@dewaagnederland.nl

RECEIVED 19 October 2023
ACCEPTED 22 March 2024
PUBLISHED 04 April 2024

CITATION
van Wolffelaar BR, van Horn JE and Hoogsteder LM (2024) Treatment of aggression regulation problems with virtual reality: study protocol for a randomized controlled trial.
Front. Psychol. 15:1324644.
doi: 10.3389/fpsyg.2024.1324644

Treatment of aggression regulation problems with virtual reality: study protocol for a randomized controlled trial

Bas R. van Wolffelaar[1]*, Joan E. van Horn[1] and Larissa M. Hoogsteder[1,2]

[1]De Forensische Zorgspecialisten, de Waag, Utrecht, Netherlands, [2]Program Group: Forensic Child and Youth Care, University of Amsterdam, Amsterdam, Netherlands

Background: Aggressive conduct among delinquents presents a pervasive issue, bearing substantial implications for not only society at large but also for the victims and the individuals displaying the aggression. Traditional approaches to treating aggression regulation deficiencies generally employ Cognitive Behavioral Therapy (CBT) in conjunction with analog role-playing exercises. A body of research supports the efficacy of various therapeutic models for aggression regulation, including Responsive Aggression Regulation Therapy (Re-ART). Role-playing within a therapeutic context has been shown to contribute significantly to reductions in violent reoffending. However, the practical application of these skills in real-world settings remains challenging due to the inherent risk of aggressive outbreaks. Additionally, the conventional role-playing scenarios, often conducted in a therapy room, lack contextual realism and may induce role confusion between the patient and the therapist. Virtual Reality (VR) technology could offer a viable solution to these limitations by allowing for skill training in both behavioral and cognitive domains within a realistic yet safe and controlled setting. The technology also facilitates real-time awareness of emotional states and tension levels in the patient. This paper describes the study protocol of a randomized controlled trial in which Re-ART offered in a virtual environment (Re-ART VR) is compared to Re-ART offered as treatment as usual.

Methods and analysis: Adult forensic outpatients with aggression regulation problems are randomly assigned to either Re-ART VR or Re-ART. The *Controlling Skills*, *Influence of Thinking* and *Handling Conflicts* modules will be offered to both groups during 3–6 months. Pre- and post-intervention measurements are performed. The primary outcome measurement is the degree of aggression regulation, while secondary outcome measurements include impulsivity and cognitive biases. Additionally, patient motivation and therapist motivation are expected to act as moderating factors.

Discussion: To date, scarcely previous research has been done on the effectiveness of VR in treatment of aggression regulation problems in forensic outpatients. Forensic outpatients who do not benefit sufficiently from mainly CBT-based interventions may benefit more from experiential learning. The unique capabilities of VR in this regard have the potential to enhance the treatment effect.

Clinical trial registration: [https://clinicaltrials.gov/], identifier [NL78265.018.21].

KEYWORDS

virtual reality, VR, responsive aggression regulation therapy, Re-ART, aggression, forensic outpatients, forensic psychiatry, randomized controlled trial

1 Introduction

Aggressive behavior poses a significant challenge to society, impacting not only the individuals involved but also the wider community. Despite a decline in international crime rates among young adult offenders (Berghuis and De Waard, 2017; Fernández-Molina and Bartolomé Gutiérrez, 2020), the persistence of violent behaviors like threats, abuse, vandalism, and public order disturbances remains a major concern due to their profound consequences on both victims and society (e.g., Aarten et al., 2020). Violent behavior has significant financial implications, affecting various aspects such as victims' costs and criminal justice expenditures (Cohen and Piquero, 2015), and legal financial obligations for offenders (Pleggenkuhle, 2018). Moreover, violent offenders often exhibit high rates of recidivism, perpetuating the cycle of harm (Falk et al., 2014; Alper et al., 2018; Weijters et al., 2018; Stewart et al., 2019). Improving therapy for violent offenders is essential. Research indicates that integrating elements of Cognitive Behavioral Therapy (CBT) with Virtual Reality (VR) holds promise. For this reason, this study focuses on the impact of VR in treatment of aggression regulation problems.

The primary approach to treating aggression regulation problems involves CBT. Overall, research highlights the significant impact of interventions incorporating role-playing, cognitive skills, and homework assignments in treatment, leading to notable decreases in general and/or violent re-offending compared to interventions lacking these elements (Jolliffe and Farrington, 2007; Papalia et al., 2019; Hoogsteder et al., 2023a). Responsive Aggression Regulation Therapy (Re-ART) is an example of an aggression regulation intervention that combines CBT-elements with role-playing and experiential exercises such as chair techniques, imaginations, and mindfulness exercises (Hoogsteder and Bogaerts, 2018). Initially, Re-ART was developed for boys and girls aged 16–21 years in residential care with severe aggression regulation problems and a moderate to (very) high recidivism risk. Its effectiveness is now demonstrated in young adults in both in- and outpatient forensic treatment (Hoogsteder et al., 2014a,b; Schippers et al., 2020). More specifically, compared to control groups receiving treatment as usual, Re-ART showed to improve coping skills, treatment motivation, and diminishes impulsivity and cognitive distortions, leading to a significant reduction in both violent and general recidivism after 2 and 3 years (Hoogsteder et al., 2018, 2021). By integrating evidence-based elements, Re-ART aims to provide a promising approach to address and manage aggressive delinquent behavior. However, despite these positive outcomes, the effect of aggression regulation therapy remains moderate, indicating the need for further refinements in treatment strategies (Papalia et al., 2020; McIntosh et al., 2021).

A new technology that seems promising in increasing the treatment effects of role playing, among other things, is VR. Role-playing in VR presents distinct advantages, allowing controlled exposure to stimuli that might be too risky or harmful in real-life situations, thereby safeguarding others. Moreover, the use of avatars in VR role play eliminates role confusion that can occur especially when therapists engage in so-called analog role-play with their patients (Fromberger et al., 2014). VR is valued for its ability to facilitate implicit and experiential learning (Slater and Sanchez-Vives, 2016; Kip et al., 2018). Comparative studies, such as that by Park et al. (2011), indicate that VR role-playing outperforms analog role-playing in enhancing conversational skills, assertiveness, treatment interest and skill generalization. Additionally, VR requires active engagement from therapists and patients. This is expected to boost patient motivation and healthcare professional enthusiasm, potentially enhancing therapeutic relationships and treatment effectiveness (Kip et al., 2019a,b).

Previous research investigating the application of VR in treating mental disorders has demonstrated its effectiveness for specific phobias, as well as other clinical conditions such as obsessive-compulsive disorder, eating disorders, autism, schizophrenia, and post-traumatic stress disorder (Eichenberg and Wolters, 2012; Turner and Casey, 2014; Valmaggia et al., 2016; Freeman et al., 2017; Wiebe et al., 2022). While it is generally acknowledged that VR shows promise as a valuable addition to existing treatments, further research is required to fully understand its effectiveness. In the context of forensic treatment, the use of VR in forensic mental health care is expected to have added value (Kip et al., 2019a,b; Roggeman et al., 2021), but research is still in its early stages. Evidence suggests the effectiveness of applying VR in forensic setting (Jo et al., 2022; Johnston et al., 2023). Findings from an initial randomized controlled trial indicated that an intervention that integrated VR with serious gaming did not yield superior results in reducing anger and aggressive behavior compared to the control condition among forensic psychiatric outpatients with aggression regulation issues. Nevertheless, qualitative data revealed that participants reported gaining more insight into their own behavior and the behavior of others (Smeijers et al., 2021). As far as our knowledge extends, no other studies have yet been conducted on the effectiveness of VR in treating aggression regulation problems within an outpatient forensic setting. Thus, more research is needed to explore its potential benefits in this specific domain (Sygel and Wallinius, 2021).

In forensic inpatient studies involving VR, positive pre-post effects were observed on self-reported direct aggression and hostility, anger control and impulsiveness compared to a waiting list control group. However, these effects ceased to exist after a 3-month follow-up period (Klein Tuente et al., 2020). The authors put forth several explanations for these findings, with the closed setting being the most significant factor. The confined environment made it challenging to apply the skills learned in the VR setting to real-life situations effectively. Moreover, the patients were not actively encouraged to practice their newly acquired skills between sessions or given homework assignments. Furthermore, the research protocol lacked options for personalized interventions, as advocated in the Risk-Need-responsivity model (RNR), the most applied rehabilitation model in forensic treatments (Andrews and Bonta, 2017). These limitations

likely contributed to the diminished medium-term impact of the VR intervention. To the best of our knowledge, no studies have been conducted to compare inpatient and outpatient groups regarding the transfer of acquired skills into practical application. However, one could hypothesize that the effectiveness of VR within an outpatient setting may be higher due to the increased opportunity for patients to practice and apply the learned skills in their daily lives. This is especially true considering that homework assignments are a standard component of the Re-ART treatment.

The primary objective of this Randomized Controlled Trial (RCT) is to assess the potential benefits of incorporating VR into the existing Re-ART. More specifically, the study aims to examine the extent to which VR enhances the degree of aggression regulation among aggressive forensic outpatients (Re-ART VR) above and beyond the effects found in their counterparts receiving the regular Re-ART intervention. The effect of cognitive biases and impulsivity as determinants of aggressive behavior is examined as secondary outcome measures as well. Moreover, the study will explore motivation as a moderating variable, both among patients undergoing treatment and among the therapists delivering the intervention. The hypothesis posits that patients treated with Re-ART VR will experience greater enhancements in the degree of aggression regulation, a more substantial reduction in cognitive biases and impulsivity compared to those receiving regular Re-ART. Additionally, it is expected that higher levels of patient motivation and therapist motivation will contribute to more pronounced improvements.

2 Methods and analysis

2.1 Study design

An RCT with a pretest-posttest design will be carried out in which forensic outpatient with aggression regulation problems are being assigned to either Re-ART using VR (Re-ART VR) or regular Re-ART. Pre- (prior to the Controlling Skills module—T1) and post-intervention (at the end of the Handling Conflicts module—T2) measurements are performed. Figure 1 presents the flow diagram.

2.2 Setting

The study will be conducted at a large Dutch center for forensic outpatient treatment. Mainly CBT-based interventions are offered to juvenile and adult offenders who, due to their offense behavior, (are prone to) encounter police force or judicial authorities. Offense behaviors can vary from aggression in or outside the family, property offenses with or without violence, or sexual offenses. Several general exclusion criteria for treatment - clinically assessed at the registration and intake phase - are applicable. General exclusion criteria for treatment are acute psychosis and serious addiction problems that require supervised detoxification, due to the absence of the specific prerequisites at the treatment center for addressing these issues. Patients enter treatment on a voluntary or mandatory basis. In both scenarios, the patients participating in the study meet the eligibility criteria for Re-ART. Voluntary treatment indicates that the patient enters treatment on his own initiative, on referral of a general practitioner or another mental health care institute. Mandatory treatment means that treatment is imposed by a judge. In these cases, a probation officer fulfills the supervisory role.

2.3 Participants

The study includes adult (18 years or older) male patients who meet the inclusion criteria for Re-ART, which entails exhibiting aggressive behavior and having a moderate to high risk of recidivism. The study takes place at four out of the thirteen locations of the treatment center, as these four locations are equipped with VR sets. Therapists receive training in CleVR's VR-CBT software (see section 2.6.4.1), either from a CleVR trainer or from colleagues with VR experience, gaining extensive experience through their regular clinical practice. The main researcher provides instructional materials, including videos demonstrating VR exercises.

2.3.1 Sample size

To ensure sufficient statistical power for detecting significant effects in our repeated measures ANCOVA with within-between interaction, an *a priori* power analysis was conducted (G*Power; Faul et al., 2009). Using an expected effect size (f) of 0.25, a significance level (α) of 0.05, and a desired power (1-β) of 0.80, the calculation revealed that a total sample size of 34 participants is required for the study (17 in each condition). The analysis assumes two groups (Re-ART VR vs. regular Re-ART) with two measurement points (pre- and post-intervention), a correlation of 0.5 among repeated measures, and no sphericity correction ($\varepsilon = 1$).

2.3.2 Recruitment

Therapists will approach patients who meet the study inclusion criteria (see Figure 1). Patients who meet the Re-ART inclusion criteria (e.g., moderate to high recidivism risk, poor impulse control) are considered for study inclusion, except for those who receive concurrent interventions (e.g., trauma-focused therapy or treatment of substance abuse) during the study. This is done to prevent potential interference with the efficacy measurements in the current study. Following a comprehensive oral and written explanation of the study, patients who meet the study inclusion criteria (adult, male) will be requested to provide informed consent during discussion of the final treatment plan. Patients will be offered a visit to the VR-room to familiarize themselves with what they can anticipate in the VR-condition. They can also contact the first author or the independent expert for more information. After the treatment plan is discussed with the patient, interventions that are indicated can take place prior to the start of the study (e.g., Re-ART module Stress Reduction, trauma-focused therapy). Pretreatment measurement takes place just before start of the study (Controlling Skills module). Upon finishing the questionnaires and forwarding them to the main researcher, patients will be informed about the condition they are assigned to.

2.3.3 Randomization

After obtaining written consent from the patient, the first author randomizes included patients using the online randomization program available at https://www.graphpad.com/quickcalcs/randomize2. Each time the program is utilized, it generates a distinct

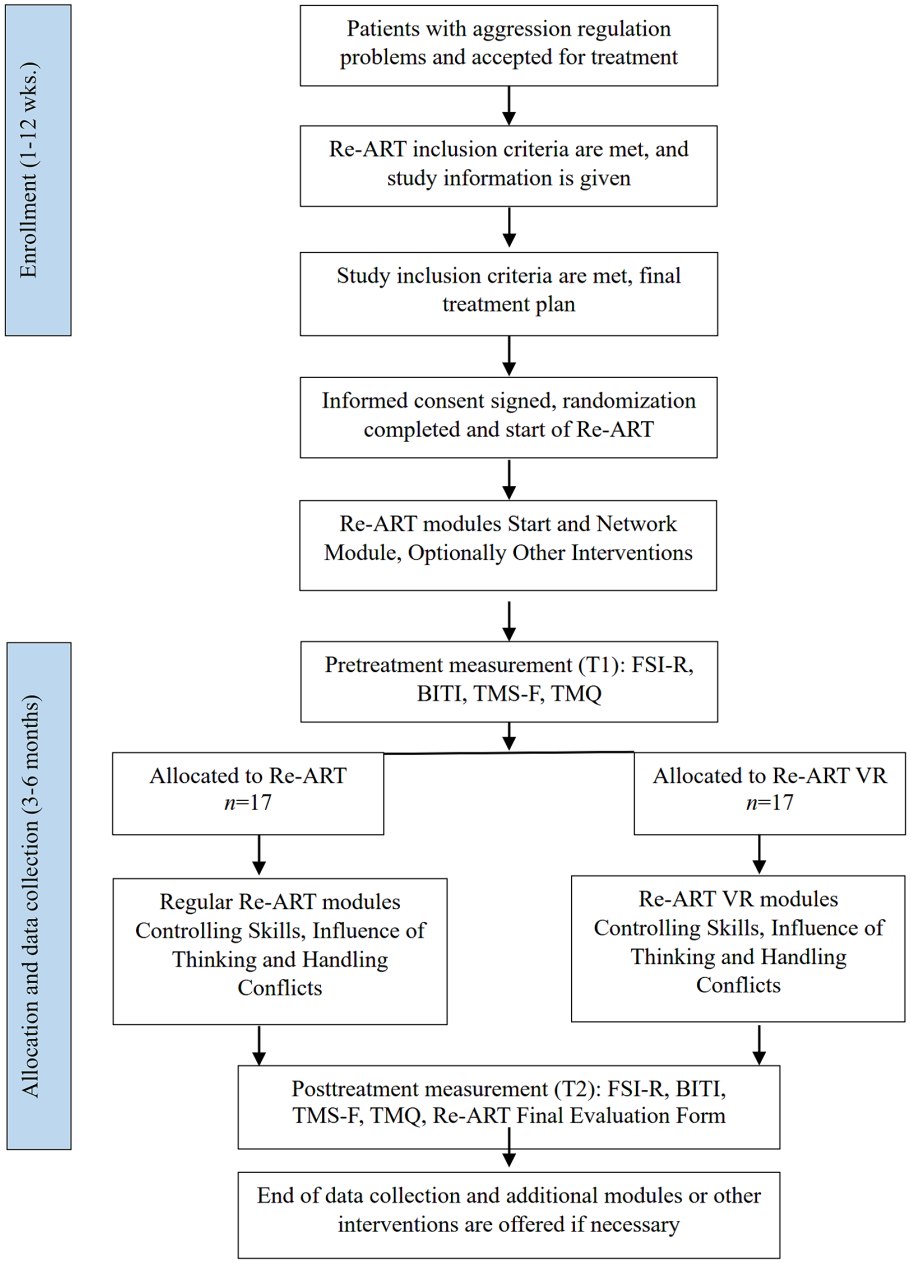


FIGURE 1
Flow diagram.

randomization list based on the date and time of usage. The randomization list will be accessible exclusively to the researchers.

2.4 Interventions

2.4.1 Responsive aggression regulation therapy

Re-ART is a protocolized intervention designed for (young) adults facing significant challenges with emotional and/or instrumental aggression regulation, and who exhibit a moderate to high recidivism risk. Incorporating the RNR principles (Andrews and Bonta, 2017), Re-ART demonstrates its flexibility by adapting the duration of the intervention (individual and group modules) according to the risk level, ranging from half a year to approximately 2 years (Risk principle). Additionally, it customizes treatment components based on the individual's specific criminogenic issues (Needs principle), and tailors the intervention to the patient's intelligence, learning style, pace, and preferred learning methods, among other factors (Responsivity principle).

The Re-ART treatment comprises standard modules, including Start, Network, Controlling Skills and Influence of Thinking. Optional modules, such as Stress Reduction, Impulse Control, Emotion Regulation, Observation and Interpretation, Self-Image and module for family systems, offer additional customization.

A comprehensive theoretical and program manual (Hoogsteder and Bogaerts, 2018) provide guidelines for each module, detailing the exercises and components to be included, along with the specific objectives that must be accomplished before progressing to the subsequent module. Moreover, the manuals outline the effective therapeutic techniques aimed at diminishing undesirable behaviors and fostering more desirable behaviors. These include motivational techniques to bolster patients' self-belief and motivation to change, the aggression chain, psychoeducation to raise awareness of the negative consequences of aggressive behavior, exercises to manage and reduce negative emotions and interventions to identify and change irrational thoughts. The intervention also encourages patients to consider alternate perspectives by adopting the viewpoint of others with more facilitative ways of thinking.

Patient engagement in Re-ART consists of individual sessions occurring at least once a week, lasting for a minimum of 1 h. Depending on the risk level, more sessions per week can take place. Group training can take place, but the focus predominantly lies on individual treatment.

The current study focuses on measuring the effect of the Controlling Skills, Influence of Thinking and Handling Conflicts modules, which are offered in that order and usually last about 3–6 months. The Controlling Skills module focuses on learning control methods, to help individuals take a time-out more easily or stay calm when a time-out is not feasible, and to apply these methods in practice. The Influence of Thinking module focuses on reducing distorting cognitions and applying helping thoughts during difficult situations. The Handling Conflicts module aims to teach various skills necessary to handle conflict constructively, such as communicating appropriately, dealing with authorities and dealing with criticism.

2.4.2 Responsive aggression regulation therapy with virtual reality

In the Re-ART VR condition, identical modules (Controlling Skills, Influence of Thinking and Handling Conflicts) with the same exercises and role plays are provided, utilizing VR. For example, situations that provoke anger are simulated, allowing patients to practice skills aimed at preventing the escalation of aggression (e.g., applying a helpful thought or responding assertively rather than aggressively). In addition to exercises and role plays, the modules encompass various components, including psycho-education, mapping triggers, identifying dysfunctional cognitions and exploring conflict styles. To ensure that in the experimental condition VR is a substantial part of the treatment, it was determined that a minimum of 60 or 66% (depending on the module) of the sessions per module should be offered with VR. To monitor the attainment of this percentage, therapists fill out the Program Integrity checklist (see section 2.6.4.2).

Following the completion of the study, additional modules, including Impulse Control, Observation and Interpretation, Emotion Regulation and the Self-image module, can be made available if deemed necessary. This provision becomes applicable after the post-test measurement is conducted and extends to other interventions as well, such as pharmacotherapy and treatment of substance abuse. Moreover, the therapist, in collaboration with the patient, then has the flexibility to opt for VR treatment, irrespective of the study condition in which the patient was during the study.

2.5 Criteria for discontinuing study participation

Participants who decide to discontinue participation in the study will receive a follow-up from their therapist to understand the reasons for their withdrawal. They will also have the option to be contacted directly by the main researcher. Feedback regarding the withdrawal will be sent via email to the main researcher to categorize responses for future discussion or reference. If any participant reports experiencing adverse effects or feels negatively impacted by their participation, they will be invited to a face-to-face meeting with the main researcher, therapist, and if necessary, a team leader. Additionally, participation in the study will be discontinued if, due to unforeseen circumstances during the study period, another intervention besides the three Re-ART modules must be provided.

2.5.1 Cyber sickness

Cyber sickness, also known as VR sickness or simulator sickness, is a common issue experienced by some individuals while using VR technology (Weech et al., 2019). Symptoms are like those of motion sickness, but they are less severe and less common (Kennedy et al., 1993). Cyber sickness occurs when there is a disconnect between the visual and vestibular systems, leading to sensory conflict and discomfort. Symptoms of cyber sickness include nausea or vomiting, dizziness, tired eyes, disorientation, dry mouth, sweating and problems with balance (LaViola, 2000; Davis et al., 2014). Some studies suggest that symptoms (at least partially) can be explained by overlap with anxiety symptoms (Fornells-Ambrojo et al., 2008). Therefore, it is expected that cyber sickness symptoms will decrease as anxiety symptoms decrease (Pot-Kolder et al., 2018). In case patients experience symptoms, such as dizziness or nausea, due to exposure to the VR environment, the therapist will help alleviate their discomfort. Several actions can be taken to prevent cyber sickness, such as building up the time a participant spends in the VR environment gradually and instruct the participant not to move his/her head too fast and in particular to walk straight (Rebenitsch, 2015). The main researcher provides a cyber sickness protocol, including instructions and strategies for therapists to prevent symptoms of cyber sickness. If cyber sickness symptoms persist and do not improve, the VR session will be halted, and a joint evaluation with the patient will determine the feasibility of continuing VR therapy. Should continuing VR therapy not be possible, the patient will be excluded from the study. Participants will be advised to seek a medical check-up with their general practitioner if any medical condition arises during the intervention period.

2.6 Materials

For an overview of the measurements (see Figure 1).

2.6.1 Primary outcome measurement

2.6.1.1 Degree of aggression regulation

The Forensic Symptoms Inventory-Revised Adults (FSI-R) (Van Horn, 2017) will be used to measure changes in the degree of aggression regulation. This self-report questionnaire consists of 32 items measuring eight domains of which a combined total score serves

as a measure for the degree of aggression regulation: Aggression (e.g., "I threatened others") and Anger (e.g., "I was annoyed"). Each subscale consists of four items, resulting in a combined scale of eight items that are rated on a 5-point scale ranging from 1 "(almost) never" to 5 "(almost) always."

Higher scores on the subscales are indicative of increased deficits in cognitive, behavioral, and/or emotional functioning. The Aggression subscale assesses the extent of physical and verbal aggression. Patients with a high score on this subscale struggle to effectively manage their feelings of anger and rage, resulting in physical and/or verbal violence toward objects and/or individuals. The Anger subscale evaluates elevated levels of both direct and underlying anger. Patients with high anger scores may face an increased risk of aggressive impulse breakthroughs.

Multi-Group Confirmatory Factor Analyses supported the measurement and structural invariance with respect to gender and age groups (18–23 years and ≥ 24 years) (Van Horn, 2018) as well as longitudinal measurement invariance (Ten Hag, 2016). Reliability coefficients were in the acceptable to good range ($\alpha \geq 0.68$) for all subscales (Ten Hag, 2016; Mac Intosh, 2021).

2.6.2 Secondary outcome measurement

2.6.2.1 Cognitive biases

To assess cognitive biases, the Brief Irrational Thoughts Inventory (BITI) will be used (Hoogsteder et al., 2014b). The BITI is a self-report questionnaire comprising 16 statements that capture three types of irrational thoughts: Aggression and Justification (9 items, e.g., "If someone touches me, I should hit him"), Sub-Assertiveness (4 items, e.g., "I think that people get angry" with me because I often say "No"), and Distrust (3 items, e.g., "Everyone is against me"). Respondents rate each item on a six-point Likert scale, ranging from 1 "totally disagree" to 6 "totally agree." The BITI's construct validity is supported by convergent, divergent and concurrent validity evidence. Furthermore, measurement invariance was established across gender and ethnic origin (native versus non-native Dutch respondents) through confirmatory factor analysis, confirming the robustness of the BITI as a valid measure (Hoogsteder et al., 2014b). The list appears to be useful for adults as well. The BITI was also found to be measurement invariant for the background characteristics of age, sex, intellectual disability and migration background (Schoute, 2022).

2.6.2.2 Impulsivity

From the FSI-R, the impulsivity subscale will be used to measure a general predisposition toward rapid, unplanned emotional (e.g., diminished ability to delay gratification) or behavioral (e.g., acting on the spur of the moment) reactions to internal or external stimuli without thinking about the consequences. The subscale contains four items (e.g., "I longed for more excitement"), each scored on a 5-point scale ranging from 1 "(almost) never" to 5 "(almost) always." For reliability and validity data of the FSI-R, see section 2.6.1.1.

FSI-R and BITI are part of the standard Routine Outcome procedure in the forensic outpatient treatment center. Routine assessments of the FSI-R and BITI, which are typically conducted every 3–4 months, will sometimes be omitted if they closely coincided with pre- or post-treatment measurements. This is done to lessen the burden on patients participating in the study.

2.6.3 Moderators

2.6.3.1 Patient motivation

Patient motivation is assessed using the Motivation to Engage in Treatment (MET) scale of the Treatment Motivational Scales for forensic outpatient treatment (TMS-F) developed by Drieschner and Boomsma (2008a). This self-report questionnaire comprises 85 statements designed to gauge the motivation of forensic patients for engaging in outpatient treatment. Beside the MET scale, the TMS-F encompasses six Internal Determinants (ID) scales and a social desirability scale. The MET scale serves as a measure for Patient Motivation, Drieschner and Boomsma (2008b) conclusion that in particular the MET scale predicts treatment engagement to a substantial degree. The MET scale contains 16 items (e.g., "I would stop therapy if I see no change in my life"), each scored on a 5-point scale ranging from 1 "totally disagree" to 5 "totally agree." Drieschner and Boomsma (2008a) found that the MET scale demonstrate adequate internal consistency, with the sum score of the items being reliably measured ($\alpha = 0.88$) for most purposes.

2.6.3.2 Therapist motivation

Therapist motivation is measured with a self-constructed questionnaire, the Therapist Motivation Questionnaire (TMQ). The TMQ consists of 13 items (e.g., "I experience positive emotions while giving treatment sessions to my patient") pertaining to various aspects, such as experiencing a positive relationship with the patient, having self-assurance in one's abilities, exerting effort to assist the patient with their issues and actively involving the patient in the treatment process. Answering categories vary on a 6-point scale from 1 "totally disagree" to 6 "totally agree." High scores reflect increased motivation.

2.6.4 Other materials and measurements

2.6.4.1 Virtual reality software

The VR-CBT software utilized in this study was collaboratively developed by CleVR[1] in conjunction with researchers from various academic institutions, mental health facilities, child and adolescent psychiatry and forensic psychiatry. CleVR's VR-CBT software has a CE mark of quality as a medical device and can be safely used as a medical aid within (mental) healthcare. The VR-CBT software enables the creation of personalized scenarios and interactive virtual (social) worlds.

Accompanied by a comprehensive manual (Social Worlds 4.1), the VR software empowers the therapist to govern the virtual environment through a laptop. During the VR session, the patient wears Oculus Rift VR glasses (type S, v1.2), fully encompassing their field of view and allowing complete immersion in the virtual environment. Concurrently, the therapist observes the same virtual setting on a computer screen, facilitating seamless communication with the patient via headphones connected to a microphone. This control encompasses adjusting characters' behavior within the virtual realm, including actions like approaching, arm movements and gaze, as well as regulating their emotional expressions, such as displaying anger. To further enhance the immersive experience, a dual-joystick

1 https://clevr.net/en/index.html

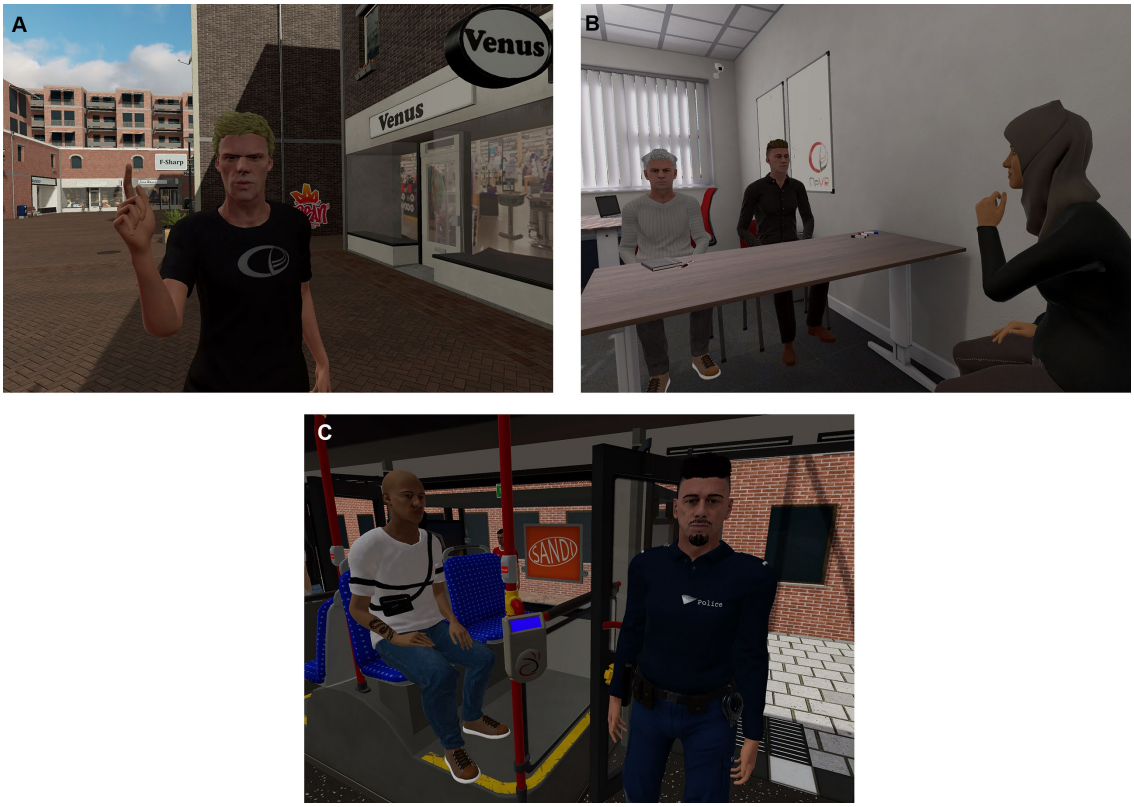


FIGURE 2
Virtual environments: shopping street **(A)**, office **(B)**, and bus **(C)**. Source: CleVR B.V., Delft, the Netherlands.

configuration allows the patient to navigate within the virtual environment while simultaneously providing real-time visual feedback of hand movements in the simulated world. Moreover, the therapist can manipulate the characters' voices by utilizing a voice shaper integrated with the MorphVox Pro program.[2]

The VR session begins as the patient put on the VR glasses and headphones, and if needed, holds the joysticks. At the outset, the patient is introduced to a "virtual waiting room," which serves as a neutral and low-stimulation environment, allowing them to adapt to the VR setting. Once the patient feels at ease, the therapist proceeds to load the targeted VR environment, and the role-playing or exercise commences. Importantly, throughout the VR session, the therapist retains the ability to "pause" the experience, accomplishing this by deactivating the voice modifier and directly communicating with the patient using their own voice within the virtual environment. Figure 2 shows some examples of virtual environments.

2.6.4.2 Program integrity

To check whether the study protocol in both conditions has been carried out according to the design, the therapist keeps a Program Integrity checklist in which the therapist records several aspects for each module, including the number of (analog or VR) role-play exercises conducted per session, the total sessions with and without role-play and whether the patient engaged in the role-plays in both a relatively relaxed state and a tense state, among other factors. In addition, the therapist is asked to report all details, including whether and why deviations from the protocol occurred and whether other therapy sessions were conducted that were not prescribed in the module.

2.6.4.3 Feedback on study protocol

To gather feedback on the study protocol, the Re-ART final evaluation form is employed. This form is a standard part of the Re-ART treatment, with versions for both patients and therapists. For the purposes of this study, modifications have been made to the forms. Questions related to feasibility and usability were added, while questions not relevant to the study were removed. Instead of administering the forms at the end of the Re-ART treatment, they will be given at the conclusion of the study.

2.7 Data management

Data relevant to the study will be stored in a distinct research file, accessible only to the researchers involved in the study. To ensure security, SPSS files will be stored in a protected Citrix environment. During data entry into SPSS, researchers will use unique codes to differentiate subjects. These codes can be cross-referenced with a separate password protected Excel spreadsheet to identify individual

2 https://screamingbee.com/Docs/MorphVOXPro/MorphDocIntro

subjects. The linkage between subject codes and patient files will be exclusively accessible to the researcher team, maintaining confidentiality and data integrity.

2.7.1 Statistical analyses

The data analysis will be conducted using IBM SPSS v29. The analytical plan aims to examine the effectiveness of Re-ART with VR (Re-ART VR) compared to standard Re-ART in a forensic outpatient setting. Intention to Treat (ITT) analyses as well as Per Protocol (PP) analyses will be conducted. The ITT analyses will include data from all randomized patients, regardless of drop-out, whereas PP analyses will include only those patients who completed the treatment as originally allocated.

To evaluate treatment effects on primary and secondary outcomes, a Repeated Measures Analysis of Covariance (RM-ANCOVA) will be performed. This approach allows to account for both baseline scores and time, along with other potential covariates, thus providing a more precise estimation of the intervention effect.

2.7.1.1 Preliminary analyses

Before conducting the main analyses, the data will be inspected for missing values (e.g., MCAR—Missing Completely At Random), outliers, and the assumptions of RM-ANCOVA, including sphericity, homogeneity of variances, and linearity. Missing data will be addressed using multiple imputations (Expectation Maximization—EM), and outliers will be examined and treated to ensure they do not bias the results.

To verify the randomization results, various background variables such as age, migration background, and primary diagnosis will be compared between patients in Re-ART VR and regular Re-ART using chi-square tests for dichotomous measures and independent *t*-tests for continuous measures. These variables will be included as covariates in the RM-ANCOVA if significant differences are found.

2.7.1.2 Primary outcome analysis: degree of aggression regulation

An RM-ANCOVA will be conducted with the FSI-R scores at both pre- and post-intervention as the within-subjects factor. The treatment condition (Re-ART VR vs. regular Re-ART) will serve as the between-subjects factor. Covariates like age, motivation level, and risk of recidivism will be included to control for initial differences.

The primary focus will be on the interaction term between treatment condition and time to examine whether the rate of improvement in aggression regulation differs between the two conditions.

2.7.1.3 Secondary outcome analysis: cognitive biases and impulsivity

For secondary outcomes, separate RM-ANCOVA's will be conducted for the BITI and the Impulsivity subscale of the FSI-R, using pre- and post-intervention scores as the within-subjects factors and treatment condition as the between-subjects factor. Additional covariates will be included as controls.

2.7.1.4 Moderators: patient and therapist motivation

To examine whether patient and therapist motivation moderate treatment outcomes, interaction terms between treatment condition and the respective motivation scales (MET scale of the TMS-F for patients and TMQ for therapists) will be included in the RM-ANCOVA models.

2.7.1.5 Sensitivity analyses

To assess the robustness of the results, sensitivity analyses will be performed to determine how the outcomes change when different covariates are included or excluded.

2.7.1.6 Correction for multiple comparisons

Given the multiple analyses conducted in this study, a correction for multiple comparisons will be applied to control the Type I error rate. We will use the Bonferroni correction method, which will adjust the alpha level accordingly.

3 Discussion

The main goal of this randomized controlled study is to evaluate the potential advantages of integrating VR into the existing Re-ART. VR is a relatively new technology in forensic outpatient treatment and preliminary evidence on its effectiveness has been provided only in forensic clinical samples (Klein Tuente et al., 2020). Based on the possibilities offered by VR (including being able to safely practice social skills in a role-playing game) and the opportunities to practice the learned skill in their own environment, it is plausible to assume that offering VR to forensic outpatients in addition to regular therapy will increase the treatment effect. To this end, an RCT-study has been set up to provide evidence for the added value of VR in enhancing aggression regulation skills in aggressive forensic outpatients (Re-ART VR) beyond the effects observed in those receiving the standard Re-ART intervention. Additionally, the research will explore motivation as a moderating factor, both among patients undergoing treatment and the therapists providing the intervention. It is expected that patients in the Re-ART VR-group will demonstrate a more significant improvement during the second assessment compared to patients in the regular condition, especially if both patients and therapists exhibit higher motivation levels. Including motivation as a moderator is a valuable addition because research shows that the lack of motivation in patients increases the chance of dropout, which in forensic patients eventually increases the chance of re-offending (Norcross et al., 2011).

3.1 Study strengths

The randomized controlled trial (RCT) as described in the study protocol offers several strengths, chief among them the robustness inherent to its design. The RCT method is a gold standard in intervention research, providing a high level of internal validity. Another strength is the incorporation of VR as a novel modality for delivering Re-ART. Given the immersive capabilities of VR, this study could pave the way for more engaging and effective treatment regimes in forensic outpatients, a population that typically presents complex treatment needs. The use of VR also allows for a controlled, safe environment where patients can practice skills that would otherwise be risky to rehearse in real-life settings. Additionally, the study employs multiple outcome measures, including assessments of the degree of aggression regulation, cognitive biases and impulsivity,

thereby providing a comprehensive evaluation of treatment effectiveness. The inclusion of both patient and therapist motivation as moderating variables is another crucial strength, recognizing the role that these factors play in treatment success.

3.2 Potential limitations

While the study is designed with rigor, it is not without limitations. One limitation could be the sample size. Although power analyses suggest that 34 subjects will be adequate, this number may not be large enough to detect small but clinically significant differences between the two treatment modalities. This is particularly the case because the study focuses on a component of treatment rather than the entire intervention, and the issues faced by the target group are intractable. Another limitation is the potential for a novelty effect with the VR condition. Patients may initially be more motivated or engaged simply because the VR treatment is new and different, and this could potentially confound the results (Klein Tuente et al., 2020). Additionally, the study's focus on adult male participants may limit the generalizability of the findings to broader forensic populations, such as females or adolescents. Furthermore, the limited duration of the intervention (3–6 months) and absence of follow-up measurement raises questions about the long-term sustainability of any observed treatment gains.

3.3 Challenges

A key challenge is ensuring treatment fidelity across both the standard Re-ART and Re-ART VR conditions. Maintaining the integrity of the interventions and controlling for therapist variables could prove difficult. While not expected, cyber sickness could be an additional challenge that could affect the patient's ability to complete the VR sessions (Weech et al., 2019). This could hinder the patient's ability to complete VR sessions, necessitating withdrawal from the study. Additionally, technical issues and the need for continual software updates represent logistical challenges. To mitigate this, the research team maintains an ongoing collaboration with the VR-CBT software provider, CleVR, to ensure the utilization of up-to-date and optimally functioning VR sets and modules. Another challenge stems from the relatively early stages of VR adoption in clinical forensic practice. Implementation hurdles include therapists' limited confidence and self-efficacy in using VR, along with practical resource constraints like time and technical support (Kouijzer et al., 2023). Therapists with limited VR experience in this study may hesitate to involve patients and may not fully realize its potential.

3.4 Ethical considerations

The study has been approved by the Medical Research Ethics Committee (MREC) of the Amsterdam University Medical Center (UMC), confirming its alignment with ethical standards. All participants will be given a detailed explanation about the study and are required to give written informed consent. Care will be taken to ensure that the VR environment does not trigger excessive stress or other adverse reactions in participants. In the case of adverse effects, the protocol outlines steps for discontinuing the intervention and providing appropriate care. Data confidentiality and participant anonymity will be maintained throughout the study.

3.5 Future directions

Given the novel nature of this study, the findings will have implications for future research and clinical practice in forensic outpatients. Should the VR condition prove to be more effective than the standard Re-ART, more robust studies must be conducted to corroborate these findings. Given the heterogeneity of forensic patients and the transdiagnostic focus of this study, future research could investigate the utility of VR in different subgroups to gain a better understanding of how and for whom VR can be valuable. The inclusion of secondary outcome measures and moderators in the study might provide initial insights into this. Furthermore, future studies should examine the use of VR in other forensic populations, such as sexual offenders, intimate partner violence offenders, and property offenders. Moreover, as VR technology continues to evolve, research will need to keep pace, examining the impact of new VR capabilities on treatment outcomes.

4 Conclusion

The RCT as described in the study protocol aims to provide valuable data on the effectiveness of incorporating VR into aggression regulation treatment for forensic outpatients. If successful, this study could represent a significant step forward in the use of technology-enhanced interventions in forensic outpatients. Given the high societal costs of aggressive and violent behavior, any improvement in treatment effectiveness has the potential for significant societal impact.

Ethics statement

The studies involving humans were approved by the Medical Research Ethics Committees (MREC) of the Amsterdam University Medical Center (UMC). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

BW: Conceptualization, Investigation, Methodology, Visualization, Writing – original draft. JH: Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing. LH: Conceptualization, Investigation, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

LH is the developer and owner of Re-ART and a Re-ART trainer and consultant.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aarten, P., Buiskool, B.-J., Hudepohl, M., Van Lakerveld, J., and Matthys, J. (2020). *Agressie en geweld in het veiligheidsveld: Literatuurstudie naar agressie en geweld jegens beroepsgroepen werkend onder beleidsverantwoordelijkheid van het Ministerie van Justitie en Veiligheid [Aggression and violence in the security field: literature review of aggression and violence toward professionals working under policy responsibility of the Ministry of Justice and security]*. Ministerie van Justitie & Veiligheid, Wetenschappelijk Onderzoek- en Documentatie Centrum (WODC), Afdeling Externe Betrekkingen (EWB), The Hague

Alper, M., Durose, M. R., and Markman, J. (2018). *2018 update on prisoner recidivism: a 9-year follow-up period (2005–2014)*, Report no. NCJ 250975, U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC.

Andrews, D. A., and Bonta, J. (2017). *The psychology of criminal conduct. 6th ed*. New York, NY: Routledge, Taylor & Francis Group.

Berghuis, A. C., and De Waard, J. (2017). Verdampende jeugdcriminaliteit: Verklaringen van de internationale daling [Evaporating juvenile crime: explaining the international decline]. *Justitiële Verkenningen* 43, 10–27. doi: 10.5553/JV/016758502017043001002

Cohen, M., and Piquero, A. (2015). Benefits and costs of a targeted intervention program for youthful offenders: the YouthBuild USA offender project. *J. Benef. Cost Anal.* 6, 603–627. doi: 10.1017/bca.2015.50

Davis, S., Nesbitt, K., and Nalivaiko, E. (2014). A systematic review of cybersickness. Proceedings of the 2014 Conference on Interactive Entertainment, Newcastle NSW, Australia, December 2-3,, New York, NY: Association for Computing Machinery, 1–9.

Drieschner, K. H., and Boomsma, A. (2008a). The treatment motivation scales for forensic outpatient treatment (TMS-F): construction and psychometric evaluation. *Assessment* 15, 224–241. doi: 10.1177/1073191107311650

Drieschner, K. H., and Boomsma, A. (2008b). Validation of the treatment motivation scales for forensic outpatient treatment (TMS-F). *Assessment* 15, 242–255. doi: 10.1177/1073191107311651

Eichenberg, C., and Wolters, C. (2012). "Virtual realities in the treatment of mental disorders: a review of the current state of research" in *Virtual reality in psychological, medical and pedagogical applications*. ed. C. Eichberg (London: InTechOpen), 35–64.

Falk, O., Wallinius, M., Lundström, S., Frisell, T., Anckarsäter, H., and Kerekes, N. (2014). The 1% of the population accountable for 63% of all violent crime convictions. *Soc. Psychiatry Psychiatr. Epidemiol.* 49, 559–571. doi: 10.1007/s00127-013-0783-y

Faul, F., Erdfelder, E., Buchner, A., and Lang, A. G. (2009). Statistical power analyses using G*power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Fernández-Molina, E., and Bartolomé Gutiérrez, R. (2020). Juvenile crime drop: what is happening with youth in Spain and why? *Eur. J. Criminol.* 17, 306–331. doi: 10.1177/1477370818792383

Fornells-Ambrojo, M., Barker, C., Swapp, D., Slater, M., Antley, A., and Freeman, D. (2008). Virtual reality and persecutory delusions: safety and feasibility. *Schizophr. Res.* 104, 228–236. doi: 10.1016/j.schres.2008.05.013

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 2393–2400. doi: 10.1017/S003329171700040X

Fromberger, P., Jordan, K., and Müller, J. (2014). Use of virtual reality in forensic psychiatry. A new paradigm? *Nervenarzt* 85, 298–303. doi: 10.1007/s00115-013-3904-7

Hoogsteder, L. M., and Bogaerts, S. (2018). *Beschrijving Agressie Regulatie op Maat Ambulant voor Volwassenen [Description Re-ART outpatient for adults]*. Nederlands Jeugdinstituut, Movisie en Trimbos instituut, Utrecht. Available at: https://www.justitieleinterventies.nl

Hoogsteder, L. M., Kuijpers, N., Stams, G. J. J. M., Van Horn, J. E., Hendriks, J., and Wissink, I. B. (2014a). Study on the effectiveness of responsive aggression regulation therapy (re-ART). *Int. J. Forensic Ment. Health* 13, 25–35. doi: 10.1080/14999013.2014.893711

Hoogsteder, L. M., Oomen, P., Sweers, N., and Hendriks, J. (2023b). A study on the effects of responsive outpatient aggression regulation therapy for juveniles. *Int. J. Forensic Ment. Health* 22, 175–186. doi: 10.1080/14999013.2022.2129524

Hoogsteder, L. M., Ruijssenaars, K., Kesteloo, S., Oomen, P., Dust, S., and Stams, G. J. J. M. A. (2023a). *Multilevel Meta-analysis: the effects of cognitive behavioral therapy in combination with experiential interventions for children and adolescents with externalizing problem behavior*.

Hoogsteder, L. M., Schippers, E. E., Sweers, N., and Stams, G. J. J. M. (2021). A quasi-experimental pilot study on the effects of responsive aggression regulation therapy (re-ART) outpatient for young adults. *J. Forensic Sci.* 66, 971–981. doi: 10.1111/1556-4029.14648

Hoogsteder, L. M., Stams, G. J. J. M., Schippers, E. E., and Bonnes, D. (2018). Responsive aggression regulation therapy (re-ART): an evaluation study in a Dutch juvenile justice institution in terms of recidivism. *Int. J. Offender Ther. Comp. Criminol.* 62, 4403–4424. doi: 10.1177/0306624X18761267

Hoogsteder, L. M., Wissink, I. B., Stams, G. J. J. M., Van Horn, J. E., and Hendriks, J. (2014b). A validation study of the brief irrational thoughts inventory. *J. Ration. Emot. Cogn. Behav. Ther.* 32, 216–232. doi: 10.1007/s10942-014-0190-7

Jo, H. J., Jung, Y. H., Hong, Y. J., Shin, Y. B., Baek, K. D., Kim, E., et al. (2022). The applicability of virtual reality-based training for controlling anger in aggressive individuals. *Cyberpsychol. Behav. Soc. Netw.* 25, 278–286. doi: 10.1089/cyber.2021.0193

Johnston, T., Seinfeld, S., Gonzalez-Liencres, C., Barnes, N., Slater, M., and Sanchez-Vives, M. V. (2023). Virtual reality for the rehabilitation and prevention of intimate partner violence - from brain to behavior: a narrative review. *Front. Psychol.* 13:788608. doi: 10.3389/fpsyg.2022.788608

Jolliffe, D., and Farrington, D. P. (2007). *A systematic review of the national and international evidence on the effectiveness of interventions with violent offenders, Ministry of Justice research series report no. 16/07*, Ministry of Justice, London.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* 3, 203–220. doi: 10.1207/s15327108ijap0303_3

Kip, H., Bouman, Y. H. A., Kelders, S. M., and Van Gemert-Pijnen, L. J. E. W. (2018). eHealth in treatment of offenders in forensic mental health: a review of the current state. *Front. Psychiatry* 9:42. doi: 10.3389/fpsyt.2018.00042

Kip, H., Kelders, S. M., Weerink, K., Kuiper, A., Brüninghoff, I., Bouman, Y. H. A., et al. (2019a). Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. *Front. Psychol.* 10:406. doi: 10.3389/fpsyg.2019.00406

Kip, H., Oberschmidt, K., Bierbooms, J., Dijkslag, D., Kelders, S. M., and Roelofsen, B. (2019b). *Technologie in de forensische zorg: crossing borders [Technology in forensic care: crossing borders]*. Kwaliteit Forensische Zorg, Utrecht.

Klein Tuente, S., Bogaerts, S., Bulten, E., Keulen-de Vos, M., Vos, M., Bokern, H., et al. (2020). Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *J. Clin. Med.* 9:072258. doi: 10.3390/jcm9072258

Kouijzer, M. M. T., Kip, H., Bouman, Y. H. A., and Kelders, S. M. (2023). Implementation of virtual reality in healthcare: a scoping review on the implementation process of virtual reality in various healthcare settings. *Impl. Sci. Commun.* 4:67. doi: 10.1186/s43058-023-00442-2

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bull.* 32, 47–56. doi: 10.1145/333329.333344

Mac Intosh, L. (2021). *Forensische klachtenlijst: Vergelijking tussen mannen en vrouwen in forensische behandeling en de algemene populatie [Forensic symptoms inventory: a comparison between men and women in forensic treatment and the general population]*. [unpublished masters's thesis] Amsterdam: Universiteit van Amsterdam.

McIntosh, L. G., Janes, S., O'Rourke, S., and Thomson, L. D. G. (2021). Effectiveness of psychological and psychosocial interventions for forensic mental health inpatients: a meta-analysis. *Aggress. Violent Behav.* 58:101551. doi: 10.1016/j.avb.2021.101551

Norcross, J. C., Krebs, P. M., and Prochaska, J. O. (2011). Stages of change. *J. Psychol.* 67, 143–154. doi: 10.1002/jclp.20758

Papalia, N., Spivak, B., Daffern, M., and Ogloff, J. R. P. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clin. Psychol. Sci. Pract.* 26:e12282. doi: 10.1111/cpsp.12282

Papalia, N., Spivak, B., Daffern, M., and Ogloff, J. R. P. (2020). Are psychological treatments for adults with histories of violent offending associated with change in

dynamic risk factors? A meta-analysis of intermediate treatment outcomes. *Crim. Justice Behav.* 47, 1585–1608. doi: 10.1177/0093854820956377

Park, K. M., Ku, J., Choi, S. H., Jang, H. J., Park, J. Y., Kim, S. I., et al. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. *Psychiatry Res.* 189, 166–172. doi: 10.1016/j.psychres.2011.04.003

Pleggenkuhle, B. (2018). The financial cost of a criminal conviction: context and consequences. *Crim. Justice Behav.* 45, 121–145. doi: 10.1177/0093854817734278

Pot-Kolder, R., Veling, W., Counotte, J., and van der Gaag, M. (2018). Anxiety partially mediates cybersickness symptoms in immersive virtual reality environments. *Cyberpsychol. Behav. Soc. Netw.* 21, 187–193. doi: 10.1089/cyber.2017.0082

Rebenitsch, L. (2015). Managing cybersickness in virtual reality. *XRDS* 22, 46–51. doi: 10.1145/2810054

Roggeman, S., Van Wallendael, K. L. P., Van West, D., and Jeandarme, I. (2021). Mogelijke toepassingen van virtual reality voor de forensische psychiatrie [Possible applications of virtual reality for forensic psychiatry]. *Tijdschr. Psychiatr.* 63, 810–815.

Schippers, E. E., Hoogsteder, L. M., and Stams, G. J. J. M. (2020). Responsive aggression regulation therapy (re-ART) improves executive functioning in adolescents and young adults with severe aggression problems: a pilot study. *J. Forensic Sci.* 65, 2058–2064. doi: 10.1111/1556-4029.14539

Schoute, A. (2022). *Het meten van irrationele gedachten: Validatie van de Verkorte Lijst Irrationele Gedachten (V-LIG) [Measuring irrational thoughts: validation of the brief irrational thoughts inventory (BITI)]*. [unpublished masters's thesis]. Amsterdam: Universiteit van Amsterdam.

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3:74. doi: 10.3389/frobt.2016.00074

Smeijers, D., Bulten, E. H., Verkes, R. J., and Koole, S. L. (2021). Testing the effects of a virtual reality game for aggressive impulse management: a preliminary randomized controlled trial among forensic psychiatric outpatients. *Brain Sci.* 11:1484. doi: 10.3390/brainsci11111484

Stewart, L. A., Wilton, G., Baglole, S., and Miller, R. (2019). *A comprehensive study of recidivism rates among Canadian federal offenders, report no. R-426*, Correctional Service of Canada, Ottawa.

Sygel, K., and Wallinius, M. (2021). Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. *Front. Psych.* 12:673089. doi: 10.3389/fpsyt.2021.673089

Ten Hag, M. C. (2016). *Longitudinale meetinvariantie van de Forensische Klachtenlijst 18+ [Longitudinal measurement invariance of the forensic symptoms inventory 18+]*. [unpublished masters's thesis] Amsterdam: Universiteit van Amsterdam.

Turner, W. A., and Casey, L. M. (2014). Outcomes associated with virtual reality in psychological interventions: where are we now? *Clin. Psychol. Rev.* 34, 634–644. doi: 10.1016/j.cpr.2014.10.003

Valmaggia, L. R., Latif, L., Kempton, M. J., and Rus-Calafell, M. (2016). Virtual reality in the psychological treatment for mental health problems: a systematic review of recent evidence. *Psychiatry Res.* 236, 189–195. doi: 10.1016/j.psychres.2015.11.042

Van Horn, J. E. (2017). *Forensische Klachtenlijst – Revised (FKL-R) volwassenen: handleiding [Forensic symptoms inventory—revised (FSI-R) adults: manual]*. De Waag, Utrecht.

Van Horn, J. E. (2018). The forensic symptoms inventory—revised (FSI-R adults): measurement and structural invariance across gender and age groups. *Psychol. Rep.* 121, 566–594. doi: 10.1177/0033294117732343

Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and Cybersickness in virtual reality are negatively related: a review. *Front. Psychol.* 10:158. doi: 10.3389/fpsyg.2019.00158

Weijters, G., Verweij, S., Tollenaar, N., and Hill, J. (2018). *Recidive onder justitiabelen in Nederland: Verslag over de periode 2006–2018 [recidivism among litigants in the Netherlands: report for the period 2006–2018], cahier report no. 2019–10*. Ministerie van Justitie en Veiligheid, Wetenschappelijk Onderzoek- en Documentatie centrum (WODC), The Hague

Wiebe, A., Kannen, K., Selaskowski, B., Mehren, A., Thöne, A. K., Pramme, L., et al. (2022). Virtual reality in the diagnosis and therapy for mental disorders: a systematic review. *Clin. Psychol. Rev.* 98:102213. doi: 10.1016/j.cpr.2022.102213

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 07 December 2023
DOI 10.3389/fpsyg.2023.1242243

OPEN ACCESS

EDITED BY
Paolo Taurisano,
University of Bari Aldo Moro, Italy

REVIEWED BY
Valeria Sebri,
European Institute of Oncology (IEO), Italy
Anson Kai Chun Chau,
The Chinese University of Hong Kong, Hong Kong SAR, China
Klaudia Proniewska,
Jagiellonian University Medical College, Poland

*CORRESPONDENCE
Chris N. W. Geraets
✉ c.n.w.geraets@umcg.nl

RECEIVED 22 June 2023
ACCEPTED 30 October 2023
PUBLISHED 07 December 2023

CITATION
Hedström R, Wallinius M, Sygel K and Geraets CNW (2023) Virtual reality-assisted assessment of paranoid ideation in forensic psychiatric inpatients: A mixed-methods pilot study.
Front. Psychol. 14:1242243.
doi: 10.3389/fpsyg.2023.1242243

Virtual reality-assisted assessment of paranoid ideation in forensic psychiatric inpatients: A mixed-methods pilot study

Richard Hedström[1,2], Märta Wallinius[1,2,3], Kristina Sygel[1,3,4] and Chris N. W. Geraets[1,2,5]*

[1]Research and Development Unit, Regional Forensic Psychiatric Clinic, Växjö, Sweden, [2]Evidence-Based Forensic Psychiatry, Department of Clinical Sciences Lund, Psychiatry, Lund University, Lund, Sweden, [3]Department of Psychiatry and Neurochemistry, Center for Ethics, Law and Mental Health, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden, [4]Department of Forensic Psychiatry, National Board of Forensic Medicine, Stockholm, Sweden, [5]Department of Psychiatry, University Medical Center Groningen, Groningen, Netherlands

Background: Reliable and valid assessment of paranoia is important in forensic psychiatry for providing adequate care. VR technology may add to current assessment procedures, as it enables observation within realistic (social) situations resembling the complexity of everyday life. VR constitutes a promising tool within forensics, due to the restricted nature of forensic psychiatric hospitals and ethical challenges arising from observing potentially dangerous behaviors in real life.

Objective: To investigate the feasibility of VR assessment for paranoid ideation in forensic psychiatric inpatients qualitatively by assessing the experiences of patients and a clinician, and to explore how the VR measures relate to established clinical measures.

Methods: One clinician (experienced psychiatrist) and 10 forensic psychiatric inpatients with a history or suspicion of paranoid ideation were included. Patients participated in two immersive VR scenarios (bus and supermarket) during which paranoia was assessed by the clinician. Qualitative interviews were performed with patients and the clinician performing the assessment to investigate experiences and feasibility. Further, measures of paranoia, social anxiety, and positive symptoms were obtained.

Results: Nine out of 10 participants with varying levels of paranoid ideation completed the assessment. Manifest inductive content analyses of the interviews revealed general experiences, advantages such as enabling observing participants from a different perspective, and challenges of the VR assessment, such as a lack of objectivity and the laboriousness of the assessment for the clinician. Although more paranoia was experienced during the supermarket scenario, correlates with classical measures were only significant for the bus scenario.

Discussion: The VR assessment was appreciated by most patients and the clinician. Based on our results short, standardized VR assessment scenarios are feasible, however, they do not appear reliable or objective for assessing paranoia. The clinical usefulness is most likely as a collaborative tool and add-on measure to existing methods.

KEYWORDS

virtual reality, assessment, diagnostics, paranoia, psychiatry, mental disorders, forensic psychiatry

1 Introduction

Paranoid ideation is one of the most common symptoms of a psychotic disorder (Freeman, 2007). Such thoughts can be mild or severe and manifest as persecutory delusions, which are characterized by the belief that harm is occurring, or will occur, and that someone intends to inflict harm (Freeman, 2007). Up to 90% of patients with a psychotic disorder experience paranoid ideation to some degree (Moutoussis et al., 2007). In forensic psychiatry, where patients with a combination of severe mental disorders and criminal conduct are treated, paranoid ideation is prevalent and has been proposed as an underlying, and sometimes even causal, risk factor for violent offenses (Coid et al., 2016; Darrell-Berry et al., 2016).

Reliable and valid assessment of paranoia is important for providing adequate healthcare for patients with psychotic disorders and may in some cases also be relevant to the formulation of violence-preventive strategies for patients. Current assessments in forensic psychiatry mainly consist of clinical interviews, self-reports, and staff observations (Aboraya et al., 2005). A downside of self-reports and interviews is that they rely on memory, insight, and motivation of patients. Furthermore, the secluded environment strongly differs from life outside the clinic, which can affect the reliability and ecological validity of observations. Novel technologies such as virtual reality (VR) could potentially provide new possibilities for psychiatric assessments (Freeman et al., 2017; van Bennekom et al., 2017; Geraets et al., 2022).

Immersive VR enables patients to interact with computer-generated virtual worlds, usually by wearing a head-mounted display. VR simulations have been shown to trigger emotional, psychological, and physical reactions similar to real-life reactions (Martens et al., 2019). To effectively elicit such emotions and responses, VR simulations have to induce a sense of "presence." For a VR user to experience presence requires experiencing a sense of both "place illusion" and "plausibility illusion"; that is believability of both the virtual environment itself and the unfolding scenario (Slater, 2009; Skarbez et al., 2017; Geraets et al., 2021).

Several systematic reviews have investigated the existing evidence on VR assessment for paranoia (Freeman et al., 2017; van Bennekom et al., 2017; Rus-Calafell et al., 2018). An advantage of VR for the assessment of paranoia lies in its controllability. In a VR environment, we can manipulate, and thus know, whether virtual characters (avatars) show friendly, neutral or hostile behavior. In contrast, when using self-report of daily life situations, it is unknown whether the self-reported paranoia or hostility reflects a persecutory delusion or whether it reflects an actual, imminent social threat. As such, in a VR situation with exclusively neutral cues, a patient's self-reported paranoid ideation and level of perceived hostility are easier to evaluate. Among the reviewed VR paradigms are scenarios involving riding the underground with several avatars, and exploring public environments such as cafés, a library, and a supermarket, while informing on thoughts of participants about the virtual scenarios directly afterward (Rus-Calafell et al., 2018). The reviews conclude that almost all reviewed VR scenarios could elicit and measure self-reported paranoia in clinical and non-clinical populations to some extent (Freeman et al., 2017; van Bennekom et al., 2017; Rus-Calafell et al., 2018). However, not all studies agree that VR can reliably differentiate between clinical and non-clinical groups. In accordance with this, many of the current studies have been performed as proof-of-concept studies for VR environments or to investigate mechanisms. A lack of knowledge still exists concerning the clinical use and potential for VR-assisted assessment of paranoid ideations.

Virtual reality technology may add value to current forensic psychiatric assessment procedures, as it enables observation within realistic (social) situations resembling the complexity of everyday life, where cognition and behavior can be monitored in real time (Riva, 1997; Freeman et al., 2017; Pan and Hamilton, 2018). Further, VR has the advantage that you can (repeatedly) expose people to (social) situations that are controlled, safe, and can be accessed within the isolated high-security environment of the clinic. VR has been used safely in both patients with paranoia and forensic psychiatric patients (e.g., Pot-Kolder et al., 2018; Klein Tuente et al., 2020). Thus, the use of VR in assessments constitutes a promising tool especially within forensic psychiatry, considering the restricted nature of forensic psychiatric hospitals and the ethical challenges arising from observing potentially dangerous behaviors in real-life settings (Freeman et al., 2017; van Bennekom et al., 2017).

In the current pilot study, we investigated a novel VR-assisted assessment for paranoid ideation in forensic psychiatric inpatients. This was done both qualitatively and quantitatively by (1) evaluating how the VR-assisted assessment was experienced by patients and the clinician, (2) investigating how clinicians' observations from VR scenarios can add to paranoia assessments, and (3) describing how the VR measures relate to established clinical measures of paranoia and anxiety.

2 Materials and methods

2.1 Design and participants

In total, 10 forensic psychiatric inpatients from a high-security forensic psychiatric clinic in Sweden were included in this single-group, mixed-methods pilot study. Inclusion criteria were: (1) aged 18 or older, (2) currently receiving forensic psychiatric treatment, and (3) a history of, or indications of current, paranoid ideation. Exclusion criteria were: (1) insufficient command of the Swedish language, (2) inability to provide informed consent due to current psychiatric status, (3) presence of an organic brain disease, e.g., dementia or epilepsy, and (4) posing severe security risks (e.g., violence) hindering safe participation. A clinician was recruited, from the consultant psychiatrists at the clinic, to perform the VR-assisted assessments.

Both quantitative and qualitative data were gathered from the patients through interviews, self-reports, observations, and medical record reviews. Qualitative (interview) data was obtained from the clinician conducting the VR-assisted assessment.

2.2 Ethics

This study was approved by the Swedish Ethics Review Authority (2021-06353-01) and was conducted according to the principles of the Declaration of Helsinki. Treating psychiatrists only referred patients to the study if they were assessed as able to provide informed consent and did not pose current and severe security risks. The patients were informed, verbally and in writing, about the study by a research assistant. Specific care was taken to explain that participation or non-participation would not in any way affect the patients' inpatient care and that they could cease participation at any time without providing a reason. If patients were willing to participate, written, informed consent was obtained.

2.3 Procedures

Patients were referred to the study by their treating psychiatrist, according to inclusion and exclusion criteria. After eligible participants had been informed of the study, those who wanted to participate signed informed consent. Subsequently, background data were collected, and the participants completed the VR scenarios and measurements. The clinician could discontinue a VR scenario when there was a risk of harm (e.g., due to falling). Participation took between 1.5 and 4 h and could be completed in 1 or 2 days (in case it was too intensive to finish all the measures on the same day). Four participants completed all measures within the same day. Five patients completed the clinical trait measures (see the measurement section) the next day, and one patient finished it 2 days later. During the VR scenarios and VR-specific measures, both a clinician and a research assistant were present. After completion of participation, participants received a gift voucher of 99 SEK (approx. $9) for the kiosk at the clinic.

2.4 VR scenarios

In this study, the Social Worlds© VR software created by CleVR (Delft, The Netherlands) was used. Participants used an Oculus Rift S to view the virtual environments, wore headphones, and moved around using Oculus controllers, see Figure 1. The clinician guided the participant through the VR scenarios by using a microphone. First, a 2-min VR practice scenario was performed, and then a supermarket and bus scenario each lasting 5 min. These two virtual environments were chosen as these are neutral environments and relatable to most people, even for those who have been incarcerated or involuntarily confined to a hospital for a long period of time.

Practice scenario: The participant was immersed in a virtual supermarket and was free to explore the supermarket to become accustomed to the VR experience and equipment, and to report signs of cybersickness. The virtual environment was programmed to include 12 freely moving customer avatars, which meant a moderate level of crowding where participants passed avatars throughout the supermarket, but never were surrounded by them. Avatars would shortly turn their attention to participants only if they came close. No other interaction with avatars was possible and the avatars were set to continuously show a neutral facial expression toward the participant.

Scenario 1: Directly after the practice scenario, the participant was instructed through the headphones to search the virtual supermarket for milk and to inform the clinician once he or she had found it. Next, the participant was instructed to find the cashier and to queue behind the avatars until the scenario ended after 5 min. Due to the random movements of the 12 avatars (acting neutrally) participants faced different numbers of avatars while queuing, and some participants experienced avatars "cutting the line" from behind. During this scenario only ambient sounds were present and avatars, including the cashier, would not respond to any potential comments by participants.

Scenario 2: The participant was seated on a bus which left the bus station and drove through a city neighborhood. On the bus were 11 other avatars, all programmed to act neutrally. The participant could not move around the environment but could look around the bus in 360 degrees as it moved. A female and a male avatar were seated opposite the participant. During the bus ride, the participant overheard the female avatar (enacted by the clinician using the microphone with voice distortion) having a (scripted) heated telephone conversation about her being late ("Mhm. I've told you I will be late! Mhm. No!" etc.). The male avatar would occasionally look at the participant, pick up his phone or look across at the female avatar at moments when she spoke louder. After ending the call, the female avatar turned to the participant and asked for directions in a calm and neutral manner. The scenario ended after the participant answered.

2.5 Quantitative measurements

2.5.1 Sociodemographic characteristics

Age, gender, diagnoses, and IQ scores were collected from patients' medical records. IQ scores were measured with the Wechsler Adult Intelligence Scale Version R, III, or IV. Reliability

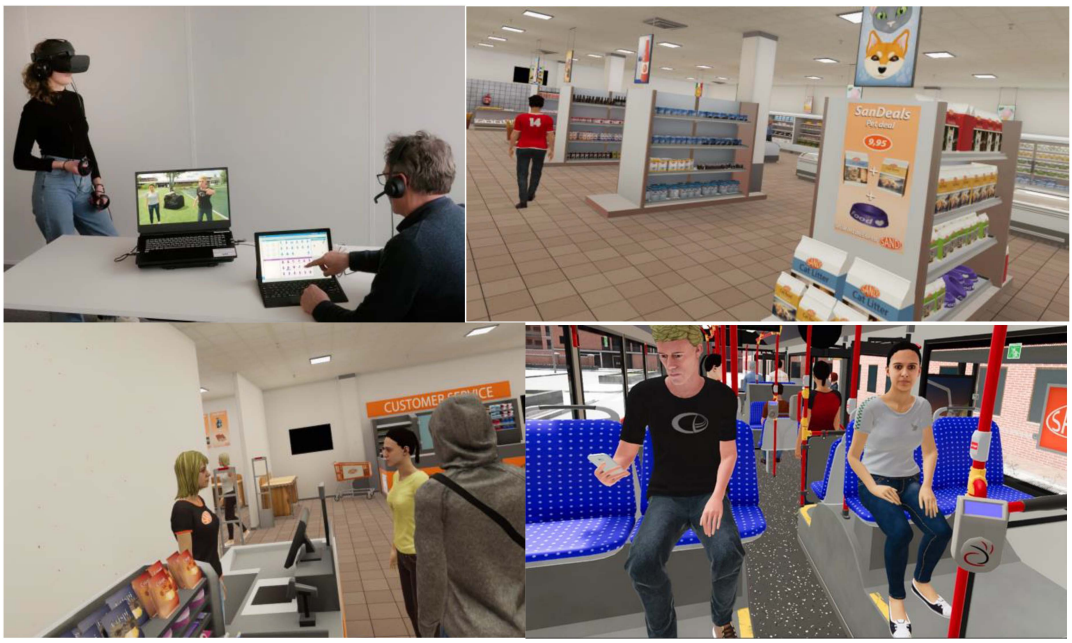


FIGURE 1
The VR setup and the VR environments as seen by the patients in Scenarios 1 and 2. Reproduced with permission from CleVR.

for the WAIS-R has only been estimated for its subscales, which were deemed moderate to good (ranging from 0.65 to 0.88) (Wechsler, 1996). Full-scale reliability for the later versions were deemed excellent: 0.90 for WAIS-III (Wechsler and Nyman, 2003) and 0.96 for WAIS-IV (Wechsler and Nyman, 2010).

2.5.2 Clinical trait measures

Positive psychotic symptoms were assessed using the Positive and Negative Syndrome Scale interview (PANSS). The Positive symptom subscale is made up of seven items measuring the presence and severity of positive symptoms on a 7-point Likert scale. This subscale has good test-retest reliability of $\alpha = 0.81$ and interrater-reliability of 0.73 (Kay et al., 1989).

Social anxiety was measured with the Social Interaction Anxiety Scale (SIAS). The SIAS consists of 20 items assessing the tendency to fear and avoid social situations on a scale from 0 (not at all) to 4 (extreme), resulting in a total score ranging between 0-80. Internal and test-retest reliability for the original English version is considered very good, while showing discriminant validity toward non-clinical samples, depression, and other anxiety disorders (Mattick and Clarke, 1998).

Paranoid thoughts were assessed with the Revised Green Paranoid Thoughts Scale (R-GPTS). The R-GPTS has two subscales: Ideas of Reference (8 items) and Ideas of Persecution (10 items). Items are rated on a 5-point Likert scale, ranging from 0 (not at all) to 4 (totally), the total score ranges from 0 to 72. The instrument is considered to have excellent reliability, especially at elevated levels of paranoia (Freeman et al., 2021).

2.5.3 VR measures

State anxiety levels were measured before and directly after each VR scenario on a verbal analog scale (VAS) by rating current anxiety on a scale from 0 (not at all) to 100 (extremely anxious).

VAS perceived hostility was assessed after each VR scenario by asking: "How hostile were the people in the VR environment toward you, on a scale from 0 (not hostile) to 100 (extremely hostile)."

State paranoia in each VR scenario was assessed with the 20-item State Social Paranoia Scale (SSPS). Ten items assess state paranoia, i.e., negative intention about the virtual characters (e.g., "Someone was hostile toward me") and 10 items describe positive or neutral interpretations of the virtual characters. Items were rated on a 5-point Likert scale, resulting in scores ranging from 10 to 50 on state paranoia. The original English instrument has demonstrated excellent internal reliability and adequate test-retest reliability. Combined with clear divergent and convergent validity it is considered to have good psychometric properties (Freeman et al., 2007).

Presence in VR was assessed after completing both scenarios using the 14-item Igroup Presence Questionnaire (IPQ). Items were scored on a 7-point Likert scale and analyzed according to previously established factors of general presence, spatial presence, involvement, and realness. The instrument has demonstrated good psychometric properties (Schubert et al., 2001).

2.6 Qualitative data

During the scenarios, the participant's behavior in VR (physical and verbal expressions) was observed by the clinician. Behavior was rated using a structured observation protocol with open questions assessing (1) social physical behavior (e.g., "Does the participant look at avatars, does he avoid avatars, etc."), (2) emotional expressions ("What emotions does the patient show during the VR session? How?"), (3) verbal expressions ("Does the patient say anything? What?"), and (4) other observations considered relevant

for the assessment of symptoms of paranoia, see Supplementary Table 1 for the observation protocol. The clinician's observations of the participant's physical and verbal expressions in VR were summarized and manually divided into categories.

The clinician conducted a semi-structured interview with the participant regarding his/her experiences in VR directly after the scenarios, to gain more insights into symptom-related experiences in VR. The interviews were audio recorded and transcribed using the NVivo software.

To assess the acceptability and feasibility of the VR simulation, the researcher conducted semi-structured interviews (see Supplementary Table 2) with both the participant and the clinician on their experiences with the VR assessment. All interviews were audio recorded and transcribed using the NVivo software.

2.7 Data analyses

Descriptive statistics were calculated for all quantitative data by presenting the mean and standard deviation or n, and the median and range. Explorative, non-parametric Spearman's rho correlation analyses were performed to assess relations between VR state paranoia measures (SSPS and VAS scores) and the clinical trait measures (PANSS, R-GPTS, and SIAS). Significance was accepted at 0.05 due to the explorative design of the study. For visualization of VR state and trait measures spider graphs were made per participants, by transforming the scores of each measure to percentages.

Analysis of the qualitative data was conducted through manifest inductive content analysis using NVivo, in accordance with the process described by Vears and Gillam (2022). The interview scripts were coded by authors RH and MW and then summarized into content categories and subcategories using an iterative process.

3 Results

Between March 2022 and May 2022 23 inpatients from the clinic's high-security units, and 53 inpatients from the clinic's medium security units were assessed for inclusion with their treating psychiatrist (see Figure 2). Exclusion occurred mainly due to not having a history of paranoid ideation or current symptoms, a cognitively impaired psychiatric state, lacking Swedish language skills and/or severe risks of violence. The risk of violence was most pronounced for high-security candidates, but was also a factor when excluding several patients in medium security. Out of the original 76 patients, 23 medium security inpatients fulfilled the criteria and were approached for the study. In total, 11 patients signed informed consent. One patient subsequently withdrew consent, resulting in a final sample of 10 participants: a 43% inclusion rate. Table 1 shows the demographic and clinical characteristics of the participants: 9 out of 10 had a current diagnosis of a psychotic disorder. Participants represented diverse psychiatric treatment histories, with a wide range of psychotic experiences and length of outpatient and inpatient care.

The sample showed a diverse presentation of current positive psychotic symptoms as measured with the PANSS. Among the 10 participants, the most noteworthy symptoms were suspiciousness/persecution (n = 1 extreme case, n = 2 severe cases) and delusions (n = 2 severe cases, n = 1 moderately severe case). Absent or minimal positive psychotic symptoms were found among three participants. Regarding paranoid ideation as measured in R-GPTS, one participant showed very severe ideas of social references, while four participants showed severe or very severe ideas of persecution, respectively (see Table 1 and Figure 3).

3.1 Experiences of VR-assisted assessment by the patients and the clinician

Nine participants completed both VR scenarios. One participant only completed the bus scenario, as the clinician discontinued the supermarket scenario (after ± 3 min) after observing balance issues. Two other participants reported minor symptoms of cybersickness and showed varying degrees of balance issues during the supermarket scenario, but could continue. One participant asked to be seated during the supermarket scenario, due to fear of standing while wearing the VR-glasses. No serious adverse events were reported.

Participants experienced moderate presence in VR on all four subscales of the IPQ (range 0–6): general M = 5.1 (SD = 1.0) spatial presence M = 4.0 (SD = 0.9), involvement M = 3.4, (SD = 1.1), and realness M = 2.9 (SD = 1.1). Table 2 presents the manifest inductive content analyses of the interviews with the patients and the clinician regarding their perceptions of the VR-assisted paranoia assessment. In total, three content categories with specific subcategories were identified: (1) advantages of VR in assessment, (2) experiences with VR assessment, and (3) challenges administering VR scenarios.

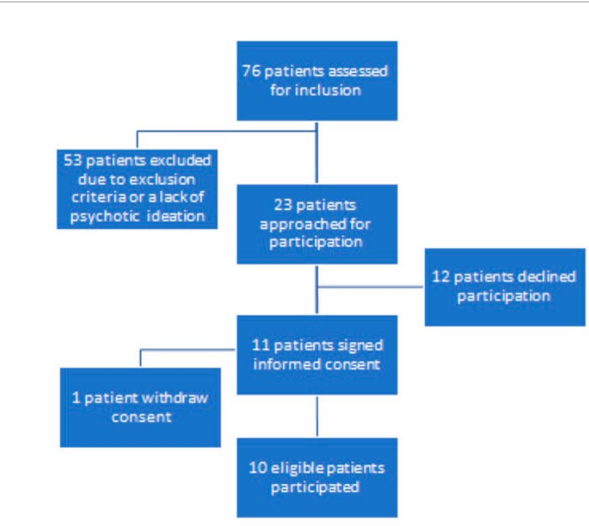


FIGURE 2
Visualization of the selection and recruitment process.

10.3389/fpsyg.2023.1242243

TABLE 1 Participant characteristics and clinical measures (n = 10).

	M (SD) or N	Median	Range
General characteristics			
Age in years	35.8 (9.6)	33.0	29–62
Female	2		
Male	8		
IQ estimate	9		
Within 85–115 range	7		
Within 85–95 range	2		
Number of diagnoses	2.1 (0.7)	2.0	1–4
Minimal one comorbidity	9		
Years since first reported psychotic symptoms	8.3 (3.3)	9.5	3–12
Years of forensic inpatient care	6.3 (10.4)	2.2	0.3–35
ICD-10 diagnoses			
Schizophrenia and Other Psychotic Disorders	9		
Substance-Abuse Disorders	7		
Autism and Other Pervasive Developmental Disorders	1		
Attention-Deficit Disorders	1		
Personality Disorders	2		
Clinical measures			
Positive symptoms (PANSS)	13.0 (5.8)	10	7–23
Delusions	2.7 (2.2)	1.5	1–6
Suspiciousness	3.4 (2.2)	2.5	1–7
Conceptual Disorganization	1.8 (1.0)	1.5	1–4
Hallucinatory Behavior	1.4 (1.3)	1	1–5
Excitement	1.1 (0.3)	1	1–2
Grandiosity	1.1 (0.3)	1	1–2
Hostility	1.5 (0.7)	1	1–3
Paranoid thoughts total (R-GPTS)	21.6 (20.8)	21	0–61
Ideas of social reference (R-GPTS A)	8.8 (9.8)	6	0–31
Average (0–9)	7		
Elevated (10–15)	1		
Moderately severe (16–20)	1		
Severe (21–24)	0		
Very severe (>24)	1		
Ideas of persecution (R-GPTS B)	12.8 (15.8)	5	0–46
Average (0–4)	4		
Elevated (5–10)	2		
Moderately severe (11–17)	0		
Severe (18–27)	3		
Very severe (>27)	1		
Social interaction anxiety (SIAS)	27.9 (22.0)	28	2–75

3.2 Clinician observations during VR

Using the structured observation protocol (**Supplementary Table 1**), the clinician mainly noted participants' direction of gaze and movement patterns, see **Table 3**. Interpretations of interactions were made in four cases, and in three cases participants gave verbal information on how they interpreted the avatars' behaviors or actions and recorded observations of emotions with positive,

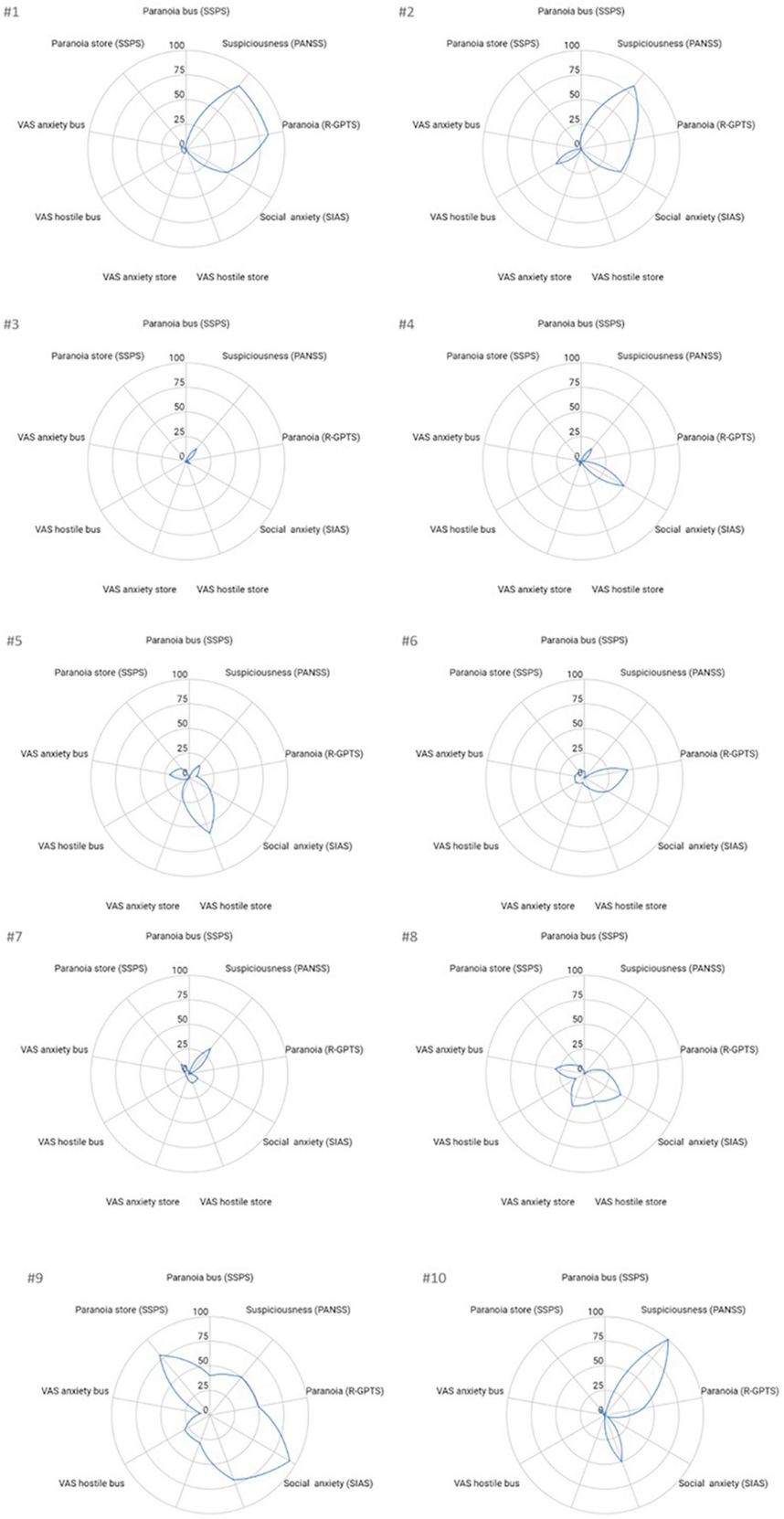


FIGURE 3
Visualization of trait and VR state measures per participant. The standardized data of participants #1 to #10 is provided for paranoia and anxiety related traits (PANSS, R-GPTS, and SIAS) and state (VAS scores and SPSS) measures. To enable comparison each score was transformed into percentages by dividing the score by the maximum score possible of that measure and multiplying by 100.

neutral, or negative valence. Furthermore, several participants narrated their experiences during the session. In two cases, the clinician noted interpretations of the emotional valence of the participants' communication with the avatars. Furthermore, the clinician noted that two patients did not fully experience the bus scenario as anticipated, which was observed as they did not listen to and/or answer the female avatar as expected; one answered the female character as she was on the phone, and both were unsure whether the female avatar's final question was directed at them.

TABLE 2 Manifest inductive content analysis with illustrative quotes from interviews with participants and clinician.

Explanation	Quote
Content category 1. Advantages of VR in assessment	
1.1. Appreciation of the VR-mediated assessment format	
VR scenarios were experienced as easy to participate in, even by patients previously expressing fear of using VR, with several preferring the VR task to standard clinical interviews/questionnaires. The clinician saw VR scenarios as a potential add-on to gold-standard interviews, as the VR context makes it possible to interact with/observe patients in a role other than as a clinical interviewer.	"It was better like this (. . .). Because when I talk to people, then they will not understand me. (. . .) But then, when they do understand, they get so scared that I have to change hospitals!" *(#10)* "Yes, some of them actually showed greater capacities or (. . .) an interest to collaborate than I expected. (. . .) There were some other sides to them, at least. I'm thinking of (participant). Now, he showed completely different sides to me! Well, maybe as a doctor I haven't seen these sides, because. . . At the ward and in the psychiatrist role (. . .) you have certain limitations in your contacts with patients. But there were other things in VR." (C)
1.2. Information gained by VR assessment	
Participants interpreted avatars' behaviors as both negative (e.g., "rude," "irritated," and "unsympathetic") and positive. Many patients experienced one or several avatars as hostile or condescending, based on eye contact or other cues (e.g., avatars approaching too closely, avatars being too quiet/too loud). The clinician emphasized how VR scenarios made it possible to observe patients from a third-person perspective, which was particularly helpful with patients who also communicated openly about their social interpretations in the VR context. This provided an increased understanding of paranoid interpretations and social skills, especially for patients with limited social experiences due to long periods of inpatient care and/or interpersonal difficulties.	"- It wasn't possible to. . . It's hard to know, actually. Because if he would have been standing there I would have asked him." (#5) - If you had the option of talking to him (. . .), what would you have said? (C) - I would probably have asked why. . . Why he's not standing in the queue? So, why is he standing there, staring? But that could provoke him too! Then maybe it's better not to say anything? (. . .) So, you better not start something unnecessary neither. (. . .) But you better keep in the back of your head that it's good to stay vigilant, yes. . ." (#5) "We do see that (patients) have difficulties, but maybe it's a little (. . .) clearer in VR? Because sometimes it is hard for them to tell us, or they do not want to discuss their symptoms (. . .). It can be of help to assess them in these situations. (. . .) Yes, there were actually at least two for whom it seemed that there was a lot, an awful lot, going on below the surface. Even if we suspect that all is not entirely well. But, the description that patients give about their conditions, or what we observe in the wards is not enough, and that's where a few other things show specifically during VR. Now I am thinking about a few patients regarding psychotic symptoms and one of them regarding communication overall, and how this person feels in social situations." (C)
Content category 2. Experiences with VR assessment	
2.1. Attitudes to VR technology	
Many participants expressed interest and curiosity for the VR technology, either because of its novelty, associations with video games or, in one case, because of the clinic's existing VR treatments. However, a few participants described simultaneous worries and fears about VR technology, e.g., fears of provoking or worsening psychotic symptoms, and a general uneasiness with the technology. The clinician expressed concern for one participant's strong fear of VR, as well as the possibility of handling patients' possible fears adequately within the 1-1.5-h timeframe of the experiment.	"It was really amazing. You know, this 360-degree vision you get, seeing everything around you instead of watching things on a screen!" *(#4)* "It felt like you almost needed to follow up, because (VR) provoked so many (. . .) worries and thoughts. . . You have to be prepared for. . . Really, some thoughts and symptoms can come to the foreground and actually need to be addressed later on, after the session." (C)
2.2. Perceived realness of VR	
Many participants described, or acted in accordance with, feelings of being immersed and present in the VR experience. These feelings were linked to the perception that avatars were actively looking at or interacting with them. However, some described VR as "different" and "virtual," or as "feeling unreal," linking it to awareness of the outside world, physical sensations from the VR equipment, limitations set by the VR scenarios (e.g., being unable to interact with objects and avatars or to move freely), and the design and behaviors of avatars which made several participants perceive them as odd, incomprehensible, non-human or "programmed." The clinician experienced the scenarios as unrealistic because of the scripted nature and wished for more interactions and freedom to navigate and try problem-solving in the VR setting.	"Yeah, no, but it's more that I can't actually imagine personalities and feelings and a consciousness in those characters. So that made it really difficult to answer these questions. So, if you mean. . . It is sort of a question of definitions for me. Noticed me? Well, then I guess, sure, she was talking to me so she did notice me, even if it didn't happen." *(# 4)* "And of course you can stand in a queue when nothing is happening, but then that will be something different. . . Either there is no cashier, in which case you have to resolve that? For example, what if you are in a real situation in a store, then you'll have to think, like: What do I do now? Should I leave and. . . Or something along those lines. Should I go look for the cashier? Should I wait until someone else does it? So there you have lots of scenarios." (C)

(Continued)

TABLE 2 (Continued)

Explanation	Quote
2.3. Patient engagement	
During follow-up interviews, all participants at times gave neutral or brief answers, with little or no self-reported suspiciousness about avatars, also when answering open-ended questions. The majority described most avatars as ordinary and unremarkable, or as lacking discernable emotions. The clinician observed varying levels of participant engagement between and during the scenarios and the follow-up interviews, with length of inpatient treatment as a possible confounder.	"- But what did you think of the avatars, those characters? (C) - Nothing special. (#3) - What were they doing? (C) - They just sat there, talking on the phone. (#3) - What kind of people were they? (C) - They were regular people? (#3) - How did they make you feel? (C) - Nothing special (#3) - What were they thinking about you? (C) - They didn't show" (#3) "Some of (the participants), maybe (. . .) the ones who haven't been inpatients for such a long time, maybe it's not as exciting for them to do something they haven't. . . Could it be that they are not as motivated as the ones for whom nothing happens, and now something does?" (C)
Category 3. Challenges administering VR scenarios	
3.1. The role of randomness and chance in VR scenarios	
The clinician noted that participants' decisions on how to explore and navigate the scenarios influenced their experiences, e.g., where, when, and how they saw or encountered avatars. Further, avatars also moved randomly, resulting in variance in avatars noticing, looking at, and approaching participants. This affected the participants' experiences, and made them non-comparable between participants.	"That guy behind me was awfully close. Although, I went. . . (. . .) It's hard to say. I don't think they. Honestly, I reacted to them and got kind of alert and careful back there." *(#5)* "Yes, I do understand that this is specific software, but in practice there will be enormous differences for the patients. Some take an endless amount of time to find products in the store and then they barely queue, while others who moved faster or just found their way by chance, had to wait for a real long time in this queue, which doesn't really exist (as avatars do not respect the queue, they do not take out groceries). And then, this is obviously not a natural situation." (C)
3.2. Patients' social experiences in relation to VR scenarios	
Participants described being with new people (in VR) as an unfamiliar experience, which for some made a strong impression. Simply being talked to or being in a social situation with avatars of the opposite sex was, for some, an unusual experience. The clinician described how participants' varying social experiences, as well as length of inpatient care, seemed to affect their social skills and comfort in social interactions, specifically at the supermarket or public transport environments. Accordingly, some participants described the VR as "enjoyable," while others brought up lifelong difficulties and uneasiness in stores and public transport.	"It was a little bit scary right there when you entered the store and then, all of a sudden, there were people everywhere and. . . That experience was some kind of smaller shock, you could say. Especially when you have been isolated from a lot of people yourself, that. . . But that passed pretty fast. My first thought was just that it must have been a long time since I found myself in a store." *(# 4)* "The comments from the patients. (. . .) That someone appreciated (. . .) just to have another person look at them (. . .). I would not have guessed that it was so important for this patient to be paid attention to in that way. And then it was interesting to observe how paranoid someone gets just by queueing in the grocery store, for example. So, it really felt like that person was terribly afraid back there. . . And another interesting piece of information was seeing how some of our patients. . . How long ago it was that they were among people outside the clinic (. . .) and how stressful those situations can be, like riding the bus or visiting a grocery store, even though it's VR." (C)
3.3. Misunderstandings in scenarios and interviews	
In follow-up interviews, some participants struggled noticeably with understanding interview questions - several expressed difficulties answering or asked for rephrasing. Also, some participants gave tangential answers, possibly related to psychotic thought processes. The clinician noticed that some participants misunderstood or were confused by scenarios. This was partly attributed to a lack of clear tasks in the experimental design, partly to distortions in voice transformation and partly to misunderstandings due to psychotic symptoms. Especially brief answers by participants made their understanding hard to evaluate.	"(The avatars) were just busy with their phones, right? There was someone who asked a question, but I didn't know if that was directed to me or to someone else." *(#6)* "These details (. . .) It really should be. . . I don't know, practiced or seen to somehow beforehand, in order to give the best results. So that it feels natural and not that. . . That the participants do not understand who is talking to them. Is it me or the avatars?" (C)
3.4. Usability and fit of VR equipment	
Some participants experienced problems with the size of the VR glasses, or discomfort using the glasses. Several participants either expressed or were observed by the clinician as having difficulties navigating in the VR context, with concerns regarding standing up or bumping into objects/avatars. The clinician described feeling stressed and discouraged when using the equipment and occasionally needing to restart the software, and recommended thorough software training and practice for clinicians.	"Yes, the weight of the VR glasses also kept me reminded that all this is not real. I would want them to weigh less, that's my opinion. When you move your head, then, then that actually felt a bit heavy." *(#6)* "In my opinion, there are technical things that need to be massively improved and planned for, before we start using this at a bigger scale. Because now the technical side was actually what influenced the most and didn't always make it possible to (. . .) work without issues. That's what popped up all the time, you know?" (C)

(Continued)

TABLE 2 (Continued)

Explanation	Quote
3.5. Simultaneous use of VR equipment and clinical observations	
Administering the protocol required simultaneously running the VR scenarios, observing and documenting behaviors. The clinician experienced this as too complex, with too many simultaneous tasks, and recommended future administrations to be conducted in teams of two clinicians.	"The tricky part was. . . Well, there were several aspects. (. . .) It was planned from the beginning that I was supposed to run the program and observe the patient at the same time. That means I have to watch the screen and observe the patient at the same time, which in my opinion has been a challenging task. Because on the one hand I have to observe how they interact with the avatars (. . .) on the screen, where they direct their attention and where they're going and so on. (. . .) But at the same time, I can also observe the patient's mimics and speech and verbal reactions. And so that they don't get dizzy (. . .). For me this was absolutely impossible!" (C)

TABLE 3 Summary of clinician's observations according to type of behavior/reaction.

Types of observation and number of participants in which the observation occurred	Illustrative examples: Supermarket scenario//Bus scenario
Topic 1. Social behaviors	
• Direction of gaze, e.g., at environment, at avatars by gender (n = 10) • Patterns of movement, e.g., hesitant, active, walk into avatar (n = 5) • Patterns of interaction, e.g., cautious, inquisitive, hesitant, none (n = 4)	• Hesitant toward avatars in the queue//looks more at male (#5) • Almost walks into some//looks at both (#2) • Active, inquisitive, walks into avatars//looking at both, especially male (#6) • //barely shifts gaze from avatars (#9)
Topic 2. Emotional pressure	
• Emotions with positive valence, relaxation, curiosity, interest (n = 3) • Emotions with neutral valence, e.g., "neutral," "no emotions" (n = 7) • Emotions with negative valence, i.e., anxiety, impatience (n = 5) • Description of participants actions, e.g., "snorts," pacing, looking around" (n = 5)	• No emotions//interest, looking around (#3) • Nothing apparent//no apparent emotions (#7) • Restless, difficulty standing still//impatient, restless (#9) • Pacing//no emotions, impatient, shaking leg (#1)
Topic 3. Verbal statements	
• Seeking further verbal instruction, e.g., asks how to move, if avatar is speaking (n = 4) • Narrating and/or explain their experience, e.g., repeats instructions, nothing is happening, expressing opinion on program, mentions hostile gaze of avatars, explains hostility score, afraid to stand up (n = 9) • Misunderstanding scenario (n = 2) • Observing emotional valence of communication with avatars, e.g., friendliness, politeness, derision (n = 2)	• Asks how to move//misunderstands then communicates with avatar (#7) • Repeats instructions//asks if female is taking to him (#5) • Reporting and wondering about next step//no (#4) • Comments on cutting in but shows patience//friendly and polite toward speaker (#10)
Topic 4. Other	
• Additional information, e.g., speed, gait, and misunderstandings of the task (n = 4)	• Solves task quickly//(#9)

3.3 Associations between VR assessments and clinical measures of paranoia and anxiety

Means, standard deviations, medians, and ranges of the VR assessment measures are presented in Table 4. Both VR paranoia measures, the SSPS state paranoia, and the single-item VAS hostility measure, correlated highly for both the supermarket (r = 0.90, $p < 0.001$) and the bus scenario (r = 0.94, $p < 0.0001$). On average, slightly more anxiety and paranoia were elicited by the supermarket scenario than the bus scenario. This is also reflected in the visualizations of the standardized scores of the state VR measures and clinical trait measures for each participant in Figure 3, where a strong heterogeneity between the profiles of patients can be noted. Some participants only showed elevated scores on established trait measures (#1, #2, #10) and not on VR measures, while some showed almost no paranoia or anxiety symptoms on any of the measures (#3, #7). Others (#8, #9) demonstrated a more integrated picture with elevated scores on both trait and VR measures. One participant (#5) reported a high level of experienced hostility in the supermarket scenario, but low scores on all other measures.

3.3.1 Supermarket scenario

Although the supermarket scenario triggered slightly more paranoid ideas, state paranoia (SPSS) and VAS hostility scores did not correlate with any of the clinical trait measures significantly. Also, for VAS anxiety scores, no clinical trait measures correlated. This indicates that people higher in trait anxiety and paranoia were not more prone to feeling anxious or paranoid during the VR exposures.

3.3.2 Bus scenario

State paranoia (SSPS) in the VR bus scenario correlated strongly with social interaction anxiety (r = 0.64, p = 0.05), ideas of social reference (r = 0.77, $p < 0.01$), ideas of persecution (r = 0.66, p = 0.04) and the paranoid thoughts total score (r = 0.81, $p < 0.01$), but not with PANSS delusions or suspiciousness. Similarly, VR VAS

TABLE 4 VR measures.

	M (SD) or N	Median	Range
VR-specific measures			
State paranoia supermarket (SSPS)	15.2 (9.6)	13	10–42
State paranoia bus (SSPS)	12.7 (5.0)	10.5	10–26
VAS anxiety pre	10.0 (18.5)	3	1–60
VAS anxiety supermarket	10.9 (12.7)	5.3	1–35
VAS anxiety bus	8.7 (9.5)	5	1–30
VAS hostility supermarket	23.3 (27.2)	10	1–70
VAS hostility bus	9.3 (11.5)	4.5	1–30

hostility during the bus scenario correlated strongly with ideas of social reference ($r = 0.69$, $p = 0.03$), and the total score on paranoid thoughts ($r = 0.67$, $p = 0.04$), but not with social interaction anxiety or PANSS delusions and suspicious. Thus in contrast to the supermarket, significant associations were found, however, when interpreting it should be noted that people scored rather low on state paranoia and anxiety measures for this scenario.

4 Discussion

The present study is, to our knowledge, the first to examine the feasibility and clinical relevance of VR-assisted assessment of paranoid ideations in a clinical forensic psychiatric setting. The ten participants showed a wide range in severity of psychotic symptoms. Overall, many were positive and curious about the assessment, even though hesitations and fears of the VR technology and its possible effects emerged. From the clinician's view, VR enabled observations of patients from a third-person perspective, and to initiate conversations on paranoia that otherwise would be difficult to create preconditions for. Several challenges were identified, such as difficulties in the practical use of VR while simultaneously performing clinical assessments and a lack of objectivity due to amongst others, variance in the avatar's automated behavior.

Self-reports of paranoid ideations in VR were partially related to trait paranoid ideation and social anxiety, but a lack of associations to the clinically assessed PANSS scales, was demonstrated.

4.1 VR-assisted assessment as perceived by patients and clinician

Many participants described VR scenarios as a novel and interesting experience when compared to standard clinical interviews. Several participants who initially worried about the VR technology and its possible impact on their wellbeing, still preferred VR over standard interviews. Interestingly, worries about VR, slight discomfort from equipment or side effects (cybersickness) did not make patients drop out of the experiment. The only discontinued participant was initiated by the clinician, as there was a risk of falling due to balance issues in VR which were clearly noticeable to the clinician but described by the participant as minor cybersickness. Although postural instability has been reported in other VR research (Tian et al., 2022), several studies with forensic and paranoid patients have not reported high rates of cybersickness or balance problems (Rus-Calafell et al., 2018; Klein Tuente et al., 2020). The balance issues could be related to the participants' unfamiliarity with VR, or the single-session format in which three scenarios were performed in a short period of time. Adaptations such as placing someone in a seated position may solve this issue.

Regarding two of the most central features of VR – experiences of presence and immersion – participants had diverse experiences. Some described experiencing strong presence and immersion, while some described circumstances affecting presence and immersion negatively, e.g., the feeling of wearing the equipment, restrictions in possible actions with the VR environment, and avatars' "unnatural" expressions and movements. Thus, several participants described experiencing breaks in presence that could be classed as breaks in either place illusion, plausibility illusion, or co-presence (Slater, 2009). Conversely, others reported a continued sense of presence and strong emotions, even when for example bumping into avatars or not being answered when talking to avatars in the supermarket scenario. More investigations of factors affecting presence and immersion are needed.

The clinician appreciated the VR assessment as an alternative means of communication with patients, underlining how VR facilitated observations and conversations with the participants who were most suspicious or hostile to forensic psychiatry and clinical interviews. This is in line with previous studies, indicating that VR interventions potentially increase the degree of personal disclosure in other psychiatric settings (Pan and Hamilton, 2018). We therefore, humbly, suggest that VR may constitute a way to establish constructive two-way communication with patients who, due to hostility, severe paranoia, or previous negative experiences from standard interventions in compulsory (forensic) psychiatric care have been less responsive to previous interventions. Further clinical advantages with VR were described by the clinician as a means for roleplay and observations for patients with limited exposure to social situations, either because of long-term inpatient care or more generalized social difficulties.

The scenarios created for the current study were short and designed to be both neutral and standardized, while letting participants approach avatars and explore the surroundings. This design still contained variations between administrations, and follow-up interviews revealed that the scenarios themselves also carried different meanings for different participants. For example, avatars moving randomly and participants exploring freely in the supermarket scenario meant that participants could bump into avatars and avatars rush past participants, causing different experiences and social situations. Thus, the ability to interact with VR environments, while potentially contributing to maintaining the plausibility illusion of a VR scenario, also creates more variation in the assessment scenario, thereby limiting standardization and objectivity (Pan and Hamilton, 2018). Furthermore, interviews

revealed that the scenarios related differently to patients' specific psychotic symptoms, cognitions, and social learning histories. A wide range of attitudes and reactions were reported to the supposedly neutral scenario environments, ranging from excitement about a virtual environment outside the clinic to VR environments triggering anxieties. These differences must be taken into consideration when evaluating our results since these experiences could be one explanation for why the three most paranoid participants did not report paranoid ideations during VR.

A prominent challenge for the clinician was the limited feasibility of simultaneously conducting clinical assessments while managing the VR technology. We acknowledge that this could be mitigated through automated VR scenarios, or through having an assistant performing the VR scenarios, allowing for the clinician to completely focus on observations of patients' behaviors. However, when protocols are less strict and not focused on objective assessment, this might also release strain on the clinician, as has been observed in treatment studies where conducting interactive scenarios and providing feedback (thus observing) with similar software was feasible for clinicians (Nijman et al., 2022). Also, providing more thorough training for the clinician (than was done for the current study) with the VR hardware and software seems to be an important feature if such assessments are to be implemented in clinical practice.

4.2 Paranoia in VR and associations with standard assessments

Paranoia in the VR scenarios was reported by some participants and was partly observed by the clinician. The clinician's observations showed a range of reactions to the scenarios (e.g., curiosity, hesitancy, lack of interest), but no strong signs of paranoid ideations and behavior among participants. For the majority of those demonstrating paranoia during the VR scenarios, the supermarket scenario provoked more paranoia than the bus scenario on both the single-item VAS scale and the state paranoia questionnaire. Inspection of the graphs demonstrating the overall paranoid symptoms presented by each participant showed that those presenting high scores at PANSS and R-GPTS scales, in general, did not experience anxiety or paranoia to a large degree during the VR scenarios. When associations were investigated in correlation analyses, no significant associations between VR-related measures of state paranoia and the clinically assessed PANSS scales were demonstrated, such relations were only found for self-reports of social anxiety and paranoid ideations (in the bus scenario). However, it also should be noted that VR state paranoia measures do not necessarily need to match the interview and questionnaire-based scores, as they concern different timeframes. I.e., the VR measure concerns the past 5 min, whereas the R-GPTS concerns the past month, and the PANSS the past week. In previous experimental studies, participants in VR scenarios have expressed ideas of persecution, social evaluative concerns, risk of physical harm and emphasized social cues in both patients and non-clinical people with high trait paranoia (e.g., being approached too closely, being the subject of the avatars' attention) (Freeman et al., 2008; Riches et al., 2020).

Given the obvious limitations of a very small sample, our results must be seen as preliminary. However, it provides information for future research that VR-elicited paranoia should be investigated in relation to clinically assessed paranoia, with no assumptions on the inherent overlap. Also, differences between clinical and non-clinical populations should be investigated, as a part of a validation process and investigating the utility as a potential "objective" assessment form, as small and non-significant differences have been reported previously when assessing paranoia during a VR underground train ride (Fornells-Ambrojo et al., 2015; Veling et al., 2016), though the majority of research did find significant differences between such groups (Rus-Calafell et al., 2018).

To be noted, two participants self-reported paranoid ideations in R-GPTS without receiving an elevated score of paranoia in the clinical PANSS interview. This could have been due to current real stressors and conflicts in the forensic psychiatric environment for these specific participants (i.e., actual threats), which was captured as paranoia by the R-GPTS. Fornells-Ambrojo et al. (2015) have previously highlighted the role of environment and adverse life-events in creating and maintaining paranoid ideations, recommending VR assessment to consider participants' social background and different neighborhoods. For our sample, these aspects were arguably important, and we recommend continued VR research in forensic settings to take the social environments of patients into consideration.

4.3 Therapeutic misconceptions

Expectations reflecting therapeutic misconceptions were evident during recruitment, administration, and follow-up interviews. Therapeutic misconception denotes the misunderstanding and conflation of research goals, protocols, and procedures with clinical treatment effects (Thong et al., 2016). While many participants seemed to understand the study, three out of ten expressed both hopes and fears of VR as an "objective" measure of their mental health, even after thorough information on the study aims and procedures was provided by the research assistant. These participants either disagreed with their current diagnosis and wished for it to be reexamined through this study, or rather worried about showing potential early signs of new psychotic episodes that would be detected through VR. Referring psychiatrists expressed similar expectations about paranoid ideations being discernable through VR scenarios, despite the study repeatedly being presented as a pilot project without the possibility of diagnosing psychotic symptoms, and that no individual findings would be reported back to the referring psychiatrist. These misconceptions of the diagnostic capabilities of VR were comparable to therapeutic misconceptions encountered in psychiatric treatment studies, e.g., potential benefits from the study for the participant's care, and misconceptions concerning the purpose of the study

(Appelbaum et al., 2012). Our study design contained several elements previously found to increase the risk of therapeutic misconception, e.g., researchers having simultaneous clinical assignments and conducting the study in forensic psychiatric settings (Pedersen et al., 2021). The current sample is also characterized by established risk factors for therapeutic misconception in psychosis patients: residential living, poor independence in activities of daily functioning, cognitive deficits, and positive psychotic symptoms (Thong et al., 2016). The expectations and fears that were voiced in our study underline the importance of clear communication adjusted to the participants' responsiveness, to decrease the risk of therapeutic misconception during research.

4.4 Limitations

The present study has several limitations. Firstly, 57% of the approached patients were not willing to participate, and we have no information on how these patients may have differed from those who chose to participate. Thus, the actual representativeness of the sample is unknown. However, this participation rate is comparable to other studies in forensic psychiatry (Pedersen et al., 2021). Second, participants' IQ scores were collected through medical records, but measured at different points in time during their illness and cannot be assumed to be reliable measures of current intellectual functioning. Because of recruitment criteria, our sample may have fewer cognitive impairments when compared to patients in similar units and with similar lengths of stay, limiting the generalizability. However, the nature of this study, being an explorative feasibility study, does not entail strict considerations regarding representativeness and generalizability.

Further, the order of the two scenarios was not randomized and therefore we cannot rule out that this might have influenced the assessment and account for differences between assessments from the supermarket and bus scenarios. Also, we do not know whether the (minor) symptoms of cybersickness may have influenced results, a larger sample is needed to investigate this. Finally, the validity of self-report measures on paranoia poses additional limitations in forensic settings, where some individuals may face actual, physical danger in the forensic hospital environment (e.g., due to threats from fellow inpatients).

5 Conclusion

In this study, VR-assisted assessment of paranoid ideations proved overall acceptable to forensic psychiatric patients with different presentations of psychotic symptoms, social experiences, familiarity with VR technology and attitudes to their diagnosis. The VR format was appreciated by a subgroup of patients and the clinician, although the VR assessment (both the clinical model and the practical use of VR technology while conducting clinical assessments) should be revised to enhance practicality. Standardized VR assessment scenarios seem feasible to perform, however, the current research shows that they do not appear reliable as a stand-alone, objective assessment of paranoid ideations in forensic psychiatric patients.

Even though our VR assessment model did not identify clearly defined paranoid symptoms, there seems to be value in introducing VR-assisted assessment to forensic psychiatric practices. Moving forward, VR-assisted assessment could be examined as a collaborative and personalized tool, which could be especially relevant for patients who are hard to engage through standard methods. Accordingly, we recommend future research on personalized VR assessment scenarios for forensic psychiatric patients. Such scenarios could, for instance, examine more therapeutic-oriented goals instead of aiming for objective symptom measurements. Finally, the forensic psychiatric setting added additional challenges concerning expectations of VR technology, patients' preoccupations with diagnoses in inpatient care, as well as the very real security concerns patients may face, all of which should be examined in future studies.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study involving humans was approved by the Swedish Ethical Review Authority Box 2110 750 02 Uppsala. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

CG and MW designed the study. RH and CG wrote the first draft of the manuscript and carried out the quantitative analysis of the results. RH and MW analyzed the qualitative interview data. KS analyzed observation protocols from the VR scenarios. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

This study was supported by the Regional Forensic Psychiatric Clinic in Växjö, Södra sjukvårdsregionen, Region Kronoberg, and the Swedish Research Council for Health, Working Life and Welfare (2018–01409).

Acknowledgments

A special thank you to the participating patients, Natalia Platonova and Ola Ståhlberg at the National Board of Forensic Medicine, Sweden, and treating psychiatrists and staff at the forensic psychiatric clinic where data collection was performed.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1242243/full#supplementary-material

References

Aboraya, A., France, C., Young, J., Curci, K., and Lepage, J. (2005). The validity of psychiatric diagnosis revisited: The clinician's guide to improve the validity of psychiatric diagnosis. *Psychiatry Edgmont. Pa Townsh.* 2, 48–55.

Appelbaum, P. S., Anatchkova, M., Albert, K., Dunn, L. B., and Lidz, C. W. (2012). Therapeutic misconception in research subjects: Development and validation of a measure. *Clin. Trials* 9, 748–761. doi: 10.1177/1740774512456455

Coid, J. W., Ullrich, S., Bebbington, P., Fazel, S., and Keers, R. (2016). Paranoid ideation and violence: Meta-analysis of individual subject data of 7 population surveys. *Schizophr. Bull.* 42, 907–915. doi: 10.1093/schbul/sbw006

Darrell-Berry, H., Berry, K., and Bucci, S. (2016). The relationship between paranoia and aggression in psychosis: A systematic review. *Schizophr. Res.* 172, 169–176. doi: 10.1016/j.schres.2016.02.009

Fornells-Ambrojo, M., Freeman, D., Slater, M., Swapp, D., Antley, A., and Barker, C. (2015). How do people with persecutory delusions evaluate threat in a controlled social environment? A qualitative study using virtual reality. *Behav. Cogn. Psychother.* 43, 89–107. doi: 10.1017/S1352465813000830

Freeman, D. (2007). Suspicious minds: The psychology of persecutory delusions. *Clin. Psychol. Rev.* 27, 425–457. doi: 10.1016/j.cpr.2006.10.004

Freeman, D., Loe, B. S., Kingdon, D., Startup, H., Molodynski, A., Rosebrock, L., et al. (2021). The revised Green et al., Paranoid Thoughts Scale (R-GPTS): Psychometric properties, severity ranges, and clinical cut-offs. *Psychol. Med.* 51, 244–253. doi: 10.1017/S0033291719003155

Freeman, D., Pugh, K., Antley, A., Slater, M., Bebbington, P., Gittins, M., et al. (2008). Virtual reality study of paranoid thinking in the general population. *Br. J. Psychiatry* 192, 258–263. doi: 10.1192/bjp.bp.107.044677

Freeman, D., Pugh, K., Green, C., Valmaggia, L., Dunn, G., and Garety, P. (2007). A measure of state persecutory ideation for experimental studies. *J. Nerv. Ment. Dis.* 195, 781–784. doi: 10.1097/NMD.0b013e318145a0a9

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 2393–2400. doi: 10.1017/S003329171700040X

Geraets, C. N. W., Stouwe, E. C. D., van der Pot-Kolder, R., and Veling, W. (2021). Advances in immersive virtual reality interventions for mental disorders: A new reality? *Curr. Opin. Psychol.* 41, 40–45. doi: 10.1016/j.copsyc.2021.02.004

Geraets, C. N. W., Wallinius, M., and Sygel, K. (2022). Use of virtual reality in psychiatric diagnostic assessments: A systematic review. *Front. Psychiatry* 13:828410. doi: 10.3389/fpsyt.2022.828410

Kay, S. R., Opler, L. A., and Lindenmayer, J.-P. (1989). The positive and negative syndrome scale (PANSS): Rationale and standardisation. *Br. J. Psychiatry* 155, 59–65.

Klein Tuente, S., Bogaerts, S., Bulten, E., Keulen-de Vos, M., Vos, M., Bokern, H., et al. (2020). Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: A multicenter randomized controlled trial. *J. Clin. Med.* 9:2258. doi: 10.3390/jcm9072258

Martens, M. A. G., Antley, A., Freeman, D., Slater, M., Harrison, P. J., and Tunbridge, E. M. (2019). It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *J. Psychopharmacol.* 33, 1264–1273. doi: 10.1177/0269881119860156

Mattick, R. P., and Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety11Editor's note: This article was written before the development of some contemporary measures of social phobia, such as the Social Phobia and Anxiety Inventory (Turner et al., 1989). We have invited this article for publication because of the growing interest in the scales described therein. S.T. *Behav. Res. Ther.* 36, 455–470. doi: 10.1016/S0005-7967(97)10031-6

Moutoussis, M., Williams, J., Dayan, P., and Bentall, R. P. (2007). Persecutory delusions and the conditioned avoidance paradigm: Towards an integration of the psychology and biology of paranoia. *Cognit. Neuropsychiatry* 12, 495–510. doi: 10.1080/13546800701566686

Nijman, S. A., Pijnenborg, G. H. M., Vermeer, R. R., Zandee, C. E. R., Zandstra, D. C., van der Vorm, D., et al. (2022). Dynamic interactive social cognition training in virtual reality (DiSCoVR) versus virtual reality relaxation (VRelax) for people with a psychotic disorder: A single-blind multicenter randomized controlled trial. *Schizophr. Bull.* 19:272. doi: 10.1093/schbul/sbac166

Pan, X., and Hamilton, A. F. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *Br. J. Psychol. Lond. Engl.* 1953, 395–417. doi: 10.1111/bjop.12290

Pedersen, S. H., Bergman, H., Berlin, J., and Hartvigsson, T. (2021). Perspectives on recruitment and representativeness in forensic psychiatric research. *Front. Psychiatry* 12:647450. doi: 10.3389/fpsyt.2021.647450

Pot-Kolder, R. M. C. A., Geraets, C. N. W., Veling, W., van Beilen, M., Staring, A. B. P., Gijsman, H. J., et al. (2018). Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. *Lancet Psychiatry* 5, 217–226. doi: 10.1016/S2215-0366(18)30053-1

Riches, S., Bird, L., Chan, N., Garety, P., Rus-Calafell, M., and Valmaggia, L. (2020). Subjective experience of paranoid ideation in a virtual reality social environment: A mixed methods cross-sectional study. *Clin. Psychol. Psychother.* 27, 337–345. doi: 10.1002/cpp.2431

Riva, G. (1997). Virtual reality as assessment tool in psychology. *Stud. Health Technol. Inform.* 44, 71–79. doi: 10.3233/978-1-60750-888-5-71

Rus-Calafell, M., Garety, P., Sason, E., Craig, T. J. K., and Valmaggia, L. R. (2018). Virtual reality in the assessment and treatment of psychosis: a systematic review of its utility, acceptability and effectiveness. *Psychol. Med.* 48, 362–391. doi: 10.1017/S0033291717001945

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence Teleoperators Virtual Environ.* 10, 266–281. doi: 10.1162/105474601300343603

Skarbez, R., Brooks, F. Jr., and Whitton, M. (2017). A Survey of Presence and Related Concepts. *ACM Comput Surv* 50 96:39. doi: 10.1145/3134301

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Thong, I. S. K., Foo, M. Y., Sum, M. Y., Capps, B., Lee, T. S., Ho, C., et al. (2016). Therapeutic misconception in psychiatry research: A systematic review. *Clin. Psychopharmacol. Neurosci.* 14, 17–25. doi: 10.9758/cpn.2016.14.1.17

Tian, N., Lopes, P., and Boulic, R. (2022). A review of cybersickness in head-mounted displays: raising attention to individual susceptibility. *Virtual Real.* 26, 1409–1441. doi: 10.1007/s10055-022-00638-2

van Bennekom, M. J., de Koning, P. P., and Denys, D. (2017). Virtual reality objectifies the diagnosis of psychiatric disorders: A literature review. *Front. Psychiatry* 8:163. doi: 10.3389/fpsyt.2017.00163

Vears, D. F., and Gillam, L. (2022). Inductive content analysis: a guide for beginning qualitative researchers. *Focus on Health Professional Education: A Multi-Professional J.* 23, 111–127. doi: 10.11157/fohpe.v23i1.544

Veling, W., Counotte, J., Pot-Kolder, R., Van Os, J., and Van Der Gaag, M. (2016). Childhood trauma, psychosis liability and social stress reactivity: A virtual reality study. *Psychol. Med.* 46, 3339–3348. doi: 10.1017/S0033291716002208

Wechsler, D. (1996). *Manual for the Wechsler Adult Intelligence Scale–Revised Swedish Edition*. Stockholm: Psykologiförlaget.

Wechsler, D., and Nyman, H. (2003). *WAIS III: Wechsler Adult Intelligence Scale*, 3rd Edn. Stockholm: Psykologiförlaget.

Wechsler, D., and Nyman, H. (2010). *WAIS IV: Wechsler Adult Intelligence Scale*, 4th Edn. Bloomington, MN: NCS Pearson Inc.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 18 September 2023
DOI 10.3389/fpsyg.2023.1231619

OPEN ACCESS

EDITED BY
Märta Wallinius,
Lund University, Sweden

REVIEWED BY
Fernando Renee González Moraga,
Regional Forensic Psychiatric Clinic Växjö, Sweden
Danai Korre,
University of Edinburgh, United Kingdom

*CORRESPONDENCE
Cylia Hendriks
✉ cylia.hendriks@inforsa.nl

RECEIVED 30 May 2023
ACCEPTED 28 August 2023
PUBLISHED 18 September 2023

CITATION
Hendriks C, Jansen JM, Smit M, Smulders LM, Popma A and Van Der Pol T (2023) VReedom: training for authorized leave of absence through virtual reality – a feasibility study. *Front. Psychol.* 14:1231619. doi: 10.3389/fpsyg.2023.1231619

VReedom: training for authorized leave of absence through virtual reality – a feasibility study

Cylia Hendriks[1,2]*, Jochem Milan Jansen[1,3], Merel Smit[1], Lisanne M. Smulders[1,2], Arne Popma[2] and Thimo Van Der Pol[1,2,4]

[1]Inforsa, Forensic Mental Healthcare, Amsterdam, Netherlands, [2]Department of Child and Adolescent Psychiatry and Psychology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, Netherlands, [3]Faculty of Law, Institute for Criminal Law and Criminology, Leiden University, Leiden, Netherlands, [4]Department of Research and Quality of Care, Arkin, Amsterdam, Netherlands

This study assessed the feasibility, implementation process and outcomes of the VReedom training; a virtual reality (VR)-based intervention designed to prepare forensic psychiatric patients for their first authorized leave. Clinical forensic mental healthcare organization Inforsa, operating at security level 3, introduced the VReedom training for forensic patients eligible for their first authorized leave, between March 1st and November 13th, 2022. Employing a retrospective observational cohort study design with patient dossier data as the primary source, the study also used participant observation, weekly evaluative questionnaires and focus group discussions as data sources. Five objectives were utilized to evaluate the feasibility: recruitment capacity and resulting sample characteristics, data collection and evaluation procedures, acceptability and suitability of the training and protocol, training management and implementation, and preliminary participant results. Despite the lack of a control group, findings align with literature suggesting VR's potential for enhancing treatment motivation and reducing stress in preparation for first authorized leave. Of 13 patients approached, 10 participated without dropouts, and no incidents occurred during training. Emotion elicitation was successful, supporting VR Exposure therapy's efficacy. Findings align with literature, emphasizing VR's value in forensic psychiatry. Establishing favorable implementation conditions was crucial, with positive reception from treatment providers. Also, the need for personalization and additional locations was identified, and the training seemed most suitable for patients with a tbs-measure. Future research with control groups is recommended to further validate the effectiveness of the VReedom training intervention, and further protocol development is necessary to make it suitable for a broader population. Current findings contribute to the refinement and expansion of evidence-based practices in the field of VR-assisted training and treatment in forensic psychiatry.

KEYWORDS

authorized leave, virtual reality, VR, feasibility, forensic psychiatry, TBS

1. Introduction

Patients in designated correctional mental health facilities display a high prevalence of schizophrenia spectrum disorders (SCZ), borderline personality disorders (BPD), antisocial disorders (ASPD), substance use disorders (SUD), and/or mood disorders (Campagnolo et al., 2019). Intensive mandatory treatment is received during the patient's residence in a designated

correctional mental health facility, and after a positive evaluation on eligibility for rehabilitation, these patients are prepared for their first authorized leave of absence (Watson and Choo, 2020). Authorized leave is a crucial step in the treatment of clinical forensic mental health patients, as it allows them to practice learned skills and helps them to adjust to conditions outside the forensic clinic (Porporino, 2010).

Authorized leave, however, can be stressful for patients who have spent time within the walls of a forensic mental health facility, as leaving the clinic and reintegrating into society can lead to feelings of anxiety, low self-esteem and (self)stigmatization (Link et al., 2001; West et al., 2014). Staff suggests that the anticipatory stress experienced by patients may cause increased rule-breaking, unsupervised drug use and unwelcome behavior during the preparation for – or during – the authorized leave (ten Zijthoff, personal communication, February 24, 2021), but scientific studies assessing such anticipatory stress in forensic patients seem unavailable. Nevertheless, the process of reintegration into society through authorized leave constitutes a significant transition from the confines of the clinical setting. These alterations in the patient's life and adaptations to the outside world can be considered social stressors, potentially contributing to tension and stress (Pillow et al., 1996). Stressors can be comprehensively defined as circumstances involving change, threat, challenges, demands, or structural constraints originating from the environment that are interconnected with the individual and compromise their operational integrity or well-being (Wheaton et al., 2013).Additionally, the occurrence of incidents, within correctional mental health facilities seem to hinder the authorization of leaves, prolong treatment duration, and decrease the frequency of authorized leaves granted to patients (Mevis, 2011; Ter Horst et al., 2015; Watson and Choo, 2020). This tendency to prioritize safety over freedom is commonly supported by politicians and policy-makers, due to the public safety concerns that arise following media coverage regarding incidents happening during authorized leave (Van der Wolf et al., 2020). Related concerns work their way into policy choices that focus primarily on public safety, which often hinders adequate treatment and resocialization (Ter Horst et al., 2015). Furthermore, the increasing number of forensic patients, coupled with staff shortages and budget cuts, is intensifying the need for new methods to improve treatment and streamline authorized leave applications (Jansman-Hart et al., 2011; Ter Horst et al., 2015; Wild et al., 2018; Kuosmanen et al., 2021).

It is important to prepare forensic psychiatric patients for their safe return to society by training the skills necessary for successful reintegration, and to reduce anticipatory stress (Ter Horst et al., 2015). A way to alleviate the anticipatory stress of forensic mental health patients could be to use exposure-based treatment and training. Exposure Therapy (ET) involves systematically confronting feared stimuli to teach individuals how to overcome fear responses and develop coping strategies, ultimately reducing their fear and stress levels (Kaplan and Tolin, 2011). Conventional ET, though, has two distinct drawbacks in current context: the limited transfer of skills from therapy to real-life situations, and the fact that it is not always possible to expose forensic mental health patients to certain stimuli for safety reasons.

In a non-forensic setting, the use of virtual reality (VR) for stress and anxiety reduction through ET and cognitive-behavioral therapy (CBT) holds potential as a strategy to tackle these concerns, while concurrently effectively mitigating feelings of anxiety and stress (Garcia-Palacios et al., 2001; Geraets et al., 2021). Immersive VR simulations are used as an instrument to expose patients to feared stimuli, and VR assisted therapy has been recognized as an evidence-based treatment method for a wide range of disorders (Botella et al., 2017; Easton et al., 2018; Bloch, 2021; Geraets et al., 2021). VR allows for personalized treatment by adjusting environmental factors and stimuli to meet the specific needs of the patient. This approach increases accessibility and motivation and tailors treatment to the patient's needs (Bowman and McMahan, 2007; Rehm et al., 2016; Easton et al., 2018; Ferreri et al., 2018; Ticknor, 2019; Cornet and Van Gelder, 2020; Kip and Bouman, 2020; Bloch, 2021). Furthermore, the utilization of VR enables individuals to engage in simulated scenarios within the clinical setting, specifically focusing on societal reintegration, with the aim of reflecting on behavior and developing essential skills. The adoption of VR in this context ensures a secure environment devoid of potential security risks or violation of ethical norms in real-life events (e.g., sending someone on leave to practice engaging with the outside world while still uncertain about the patient's readiness for this transition) (Cornet and Van Gelder, 2020). Given these considerations, it is plausible that the forensic patient population could profit from this approach.

To date, few studies have investigated the application of VR for the treatment of forensic psychiatric patients (Sygel and Wallinius, 2021). The existing body of evidence is currently inadequate to advocate for the immediate and widespread adoption of any particular VR intervention. Nevertheless, several interventions have demonstrated feasibility and acceptability among participants, while also offering valuable insights and serving as sources of inspiration for subsequent research and advancement (Sygel and Wallinius, 2021). The available research indicates that VR offers the opportunity to observe and monitor triggering situations and consequent behaviors, without compromising public safety (Kip et al., 2019). Moreover, a few studies on the effect of VR aggression prevention training have shown that while self-reported and staff-observed aggression had not decreased, enhanced anger control skills and less hostility and impulsivity were reported (Klein Tuente et al., 2020; Geraets et al., 2021). Finally, a study on the use of behavioral monitoring of sexual offenders against children in virtual risk situations addressed the feasibility of using VR in unsupervised privileges, showing VR risk situations provide additional information for risk management (Fromberger et al., 2018). Despite these promising developments, however, VR has not yet been deployed in the context of preparation for authorized leave.

The newly developed VReedom training – a VR-assisted authorized leave training – aids patients to (re-)learn skills needed in realistic physical environments, without the necessity of being exposed to real-world situations. The therapy is based on elements of exposure therapy and mimics real authorized leave with activities like walking outside, going to the supermarket, and interacting with a stranger (an *embodied conversational agent*) (Ferreri et al., 2018). Participants are challenged with behavioral problems and potentially stressful situations which require participants to use coping mechanisms. The goal is to gradually decrease stress levels, and subsequent unwanted behavior, by practicing challenging situations provided by personalized triggers, tailored to the participant's specific needs.

Given that VR-assisted exposure therapy is a new method of therapy within the context of forensic psychiatry and authorized leave, it is important to first examine the feasibility of both the content and

processes surrounding VReedom. In anticipation and preparation for a larger study to assess the effectiveness of VReedom, a retrospective feasibility study was conducted, to determine whether VReedom is sustainable and suitable for further examination, and what aspects of the study or training protocol need adjustment (Bowen et al., 2009). Since a feasibility study covers the initial phase of development, it can identify potential challenges that occur in preparation for and during the examination (Orsmond and Cohn, 2015). The present study therefore assesses the feasibility of VReedom, based on the objectives as described in Orsmond and Cohn (2015): recruitment capacity and resulting sample characteristics, data collection and evaluation procedures, acceptability and suitability of the training and protocol, training management and implementation, and preliminary participant results. These objectives were investigated based on retrospective, qualitative data collected on both therapist and patient level.

2. Methods

2.1. Design

Inforsa, a (clinical) forensic mental healthcare organization in Amsterdam at security level 3[1], has developed and introduced the VReedom training, targeting forensic patients who are eligible for first authorized leave. A qualitative, retrospective investigation was conducted to examine the implementation process and outcomes of this training initiative, using patient records and interviews with staff that conducted the VReedom training The study was conducted between March 1st, 2022 and November 13th, 2022, in which researchers aimed to assess overall experiences, challenges encountered, and potential areas for improvement in the execution of this training. Both the feasibility research structure and the questions are based on the article by Orsmond and Cohn (2015).

To minimize participant burden, a retrospective observational cohort study was employed using patient dossier information as the primary data source. Additionally, a dual approach to data collection was employed, utilizing both questionnaires used for treatment evaluation from the patient dossiers and focus group discussions involving clinicians. This methodology was adopted to construct a qualitative portrayal of the feasibility status of the VReedom training. Through this strategy, insights were sought directly from the sources themselves, specifically the clinicians and patients engaged in the training, concerning their perceptions of specific actions and components of the training program. Such an approach facilitated a comprehensive understanding of the training's efficacy from those closely associated with its implementation and reception. Treatment evaluations were derived from the questionnaires, while the focus groups were utilized to facilitate in-depth exploration, thereby enhancing insights.

Researchers used participant observation as the main research method. Participant observation is a qualitative research method involving active immersion in the study context, interacting closely with participants, and keenly observing their behaviors and interactions. By integrating themselves into the evaluation process, the researchers aimed to capture a comprehensive understanding of the therapists' perspectives, which could subsequently contribute to improving the quality of care provided. While the presence as a researcher may inadvertently impact participant behavior and responses, these concerns were addressed through mitigation strategies to optimize the method's validity, such as being transparent about the possible impact of the researchers' presence (Kawulich, 2005).

Since this study involved retrospective data collection without acts or interventions on subjects, it did not fall under the scope of the Medical Research Involving Human Subjects Act (WMO). Instead, the Medical Ethics Committee of the Amsterdam Academic Medical Center issued a non-WMO declaration.

2.2. Intervention

Therapists conducted VReedom training sessions with patients between March 1st, 2022 and November 13th, 2022. During these sessions, therapists supervised the training, while a technical assistant adjusted the settings of the virtual environments in the VR software based on the participant's personal information and triggers. Adverse events would have led to the immediate termination of sessions ($n = 0$). The VReedom training utilized two key features: exposure to meaningful social and environmental triggers, and gradual exposure to feared stimuli using Wander (360-degree street-view) and CleVR (interactive VR environment) software (Appendix 1 in Supplementary material). To utilize the software, the Oculus Quest 2 VR headset and the HP Spectre X360 laptop were employed. The use of Wander is advantageous due to its capacity to be applied to the precise geographical location, a capability not yet developed by CleVR. Additionally, there is the consideration of exposure to interactions, which cannot be controlled within the Wander application.

The VReedom training (ideally) consisted of a six-week program, as an individual's eligibility for leave, a minimum of 6 weeks is allocated before they are granted permission to exit the facility, with the shortest duration being 8 weeks. There were six moments of contact in total (including the introduction), allowing for an additional buffer of two weeks to accommodate any potential delays (Appendix 2 in Supplementary material). In the introductory session, patients were familiarized with the therapy procedure and the staff involved. Session 1 involved a virtual walk around the clinic and nearby supermarket without any negative triggers. In session 2, triggers were introduced in the virtual supermarket walk-through (e.g., predefined and not therapist-guided character sentences, spoken by virtual agents: "Can I ask you something?," "What time is it?," "Are you okay?"), with the level of complexity of the questions asked increasing throughout the session ("Do not be so dumb," "You've been through something!," "Do not you have a job?"). In session 3, the first round through the supermarket matched the level of triggers of session 2, with the difficulty level increasing as the session progressed. Session 4 involved medium-level confrontations in the virtual supermarket through

1 An institution belonging to security level 3 mainly houses patients who need to stay within the secure ring for a longer period of time. There is a closed setting with limited freedom of movement. Full treatment and recreation services are available within the secure ring.

role-play (e.g., a conversation between another supermarket customer, played by the therapist, and the patient, in which the supermarket customer asks where the eggs are located and inquires if they know each other). Also, in this session, the walk around the clinic was repeated. Session 5 repeated the role-play at a more difficult level (e.g., a conversation between the cashier and the patient, in which the former denies having returned the incorrect change to the latter), and practiced the trigger-rich round through the supermarket once more before the moment of first physical authorized leave.

2.3. Objectives

The feasibility of the VReedom intervention was assessed by investigating five main topics, based on Orsmond and Cohn (2015): (1) recruitment capability and resulting sample characteristics, (2) data collection and evaluation procedures, (3) training acceptability and suitability, (4) training management and implementation and (5) preliminary participant results. The VReedom training and study design are deemed feasible when there are no insurmountable issues which would prevent a successful execution of either or both.

The first objective, recruitment capability and resulting sample characteristics, pertained to the inclusion process and the suitability of the currently employed exclusion criteria. The second objective, data collection and evaluation procedures, included the comprehensibility and appropriateness of the questions asked after the VReedom session. The third objective, training acceptability and suitability, pertained to the technical applicability of the protocol, the contents of the protocol and the session duration and level of session difficulty. The fourth objective, training management and implementation, related to the assessment of the resources and ability to manage and implement the study and intervention, the opportunities offered by the employer to manage the weekly training sessions, whether the therapists possessed the necessary skills and expertise to execute the sessions, the usability of the used technology and the management of incidents. The fifth and final objective, preliminary participant results, included the preliminary reactions and experienced emotions of patients that participated in the training and the relatability and immersiveness[2] of the training in comparison the real life situations, and the suitable for different target populations. These initial findings indicate the potential effectiveness of the study, yet remain preliminary in nature and are derived from a limited sample size, preventing definitive conclusions.

2.4. Participants

2.4.1. Therapists

In preparation for the study, participating therapists underwent a comprehensive full-day training on the VR technology used in the VReedom intervention. This training was facilitated through a collaboration between CleVR, a software company, and employees of the designated mental healthcare facility Inforsa. The four therapists performing the VReedom training were employed at the same forensic mental health facility, and three therapists participated in the focus group. All were between the ages of 31 and 35 ($M = 32.33$, $SD = 1.528$), and had 2 to 12 years of therapeutic work experience ($M = 5.5$, $SD = 5.635$). All therapists were female and had treated between 1 and 5 patients with the VReedom training, predominantly training patients from their own department (70%). None of the therapists had previous experience with VR prior to the start of the VReedom training program.

2.4.2. Patients

A total of 10 forensic mental health patients with severe psychiatric problems were recruited by the four therapists to receive the VReedom training. Patients were eligible in the final phase of admission at the Forensic Psychiatric Clinic (FPC) or the Long-Term Intensive Care unit (LIC). During the preparation leading up to the leave and thus during the period in which the patients receive the VReedom training, it is determined whether an individual is actually granted leave, based on the progress observed in a patient. This is important to mention in the context of potentially desired behavior exhibited by patients.

The definitive sample was aged between 24 and 59 years ($M = 40.4$, $SD = 10.469$), and were predominantly male (9 out of 10). Within this population, 5 patients were born in the Netherlands. Others were born on the continent of Africa (3), or Asia (2). Within this sample of patients, 7 were admitted within the facility under tbs-measure (detention under a hospital order), 2 conditional tbs-measure (conditional detention under a hospital order), and 1 of the patients was submitted with a care authorization title. Patients stayed a minimum one time (current stay counting as their first), and maximum of five times in the facility ($M = 1.5$, $SD = 1.269$). Primary diagnoses included schizophrenia and other psychotic disorders (9 patients), developmental disorder in childhood or adolescence (1 patient) on Axis I, and anxiety disorder (1 patient) and substance use disorders (9 patients) on Axis II. Upon admission, patients with a conditional tbs-measure underwent a 6-week observation period to assess their eligibility for authorized leave. Patients with tbs-measure and patients with care authorization were evaluated for leave eligibility based on their progress, discussed with the relevant therapist. In case of eligibility, an introductory session was arranged with the clinician, and patients were explained the contents and purpose of the training. Patients with severe physical disabilities (such as deafness and blindness), epilepsy, insufficient understanding of the spoken and written Dutch language, or those currently experiencing a psychological crisis were excluded from the training.

2.5. Data collection procedure

The primary data source, historical dossier information, included data regarding twice daily staff-reported data on stress levels (measured using a four-point stress scale from the patient's alert plan), inpatient problematic behavior (e.g., aggression or positive drug tests), therapy compliance, visitation, medication adherence, incidents during authorized leave (e.g., absconding, criminal behavior, or drug

2 Immersive VR, also known as IVR, "is generally experienced via a head-mounted display (HMD) that shuts the user off from real-world visual input and perceptually immerses the user in the VR" (Cornet, L. J. M., & Van Gelder, J.-L. (2020). Virtual reality: a use case for criminal justice practice. Psychology, Crime & Law, 26(7), 631–647. https://doi.org/10.1080/1068316x.2019.1708357).

use), and other stress-inducing factors such as illness or family issues. This comprehensive approach allowed for the visualization of patients' stress levels over an eight-week observation period.

As part of routine treatment monitoring, patients and therapists were asked open-ended questions about their experience with and during the training session each week (Appendices 3, 4 in Supplementary material). Also, the therapists conducted weekly peer evaluations. These evaluations served as a platform to discuss their experiences and share insights among one another. In order to gather valuable data and insights, the researchers occasionally joined these sessions, observing and documenting the therapists' experiences.

Additionally, a retrospective focus group was conducted with the therapists to explore their experiences with the training and recommendations for future improvement. The focus group was led by a member of the research team and lasted one and a half hours. During the session, the experiences and recommendations from the therapists were gathered through a semi-structured questionnaire including 29 open-ended questions (Appendix 5 in Supplementary material), with room for therapists to elaborate on other topics if deemed constructive by the researcher.

2.6. Data-analysis

Quantitative data was extracted from the MijnQuarant patient dossier software and securely entered into an encrypted and anonymized data file using the VIPLive (2022) data entry program. Data was then transported to SPSS (version 27, IBM Corp), in which cumulative frequencies were calculated.

Regarding qualitative data-analysis, the audio-recording from the focus group was transcribed verbatim, and then analyzed using qualitative analysis software MAXQDA 2022 (VERBISoftware, 2021). Weekly patient questionnaires and notes from the weekly peer evaluations were analyzed using the program Microsoft Excel. Data was coded independently by two researchers through a combination of deductive and axial coding was used, by use of a process of constant comparison. After each round of coding, researchers discussed both independently coded segments, and reached consensus on the final results. In the first round of coding, five predetermined coding themes were applied to the transcript: (1) recruitment and sample, (2) data collection, (3) intervention and study procedures, (4) intervention management and implementation, and (5) preliminary results. During the second round of coding, instances where the codes differed between the two independent coders were subjected to further independent coding by the researchers. These instances were thoroughly discussed until a consensus was reached regarding the final placement of the codes. In the third round, codes were added to indicate whether the training aspect was suitable for inclusion in the new protocol, needed some adjustments, was neutral or unrelated to the training, or was a recommendation for the future development of VReedom. In the fourth round, the coded sentences were condensed into brief phrases or terms that could be easily understood when included in the results section without the original context. In the fifth and final round of coding, the five themes were compared for the last time, looking for duplicates between the themes, and theme titles were adjusted in a way that optimally represented the respective coded section. Finally, all codes and code titles were translated from Dutch to English. In the results section, it is indicated whether therapists' statements originated from the weekly evaluative questionnaires (Q) or the focus group discussions (FG). Patient responses consistently stemmed from the weekly questionnaires.

3. Results

An overview of the focus group (FG) results can be found in Appendix 6 in Supplementary material.

3.1. Recruitment capability and resulting sample characteristics

Regarding the capability of patient recruitment, it has been observed that over the course of the 8-month data collection period a total of 13 eligible patients were identified, indicating an average monthly inclusion rate of 1.63 patients. None of the approached individuals with enforced treatment ($n = 7$) declined participation. Among the group of individuals with a conditional sentence, five were initially approached, but three declined, resulting in a smaller representation of this population ($n = 2$). Within the group with care authorization, there were fewer patients with opportunities for leave during the data collection period, resulting in their underrepresentation ($n = 1$). This included individual was the only eligible candidate for inclusion. Eventually, all participants ($n = 10$) completed the training in its entirety, with no premature dropouts. It was evident, though, that there was a presence of a time gap between sessions, differing from one week delay, up to a two week delay. This delay occurred because the patients were preoccupied with other activities during those weeks or sickness, which prevented an immediate initiation of the training. During the entirety of the course of training and data collection period, no reported adverse events were reported ($n = 0$), however, the one participant with care authorization occasionally required additional post-treatment care from the therapist and the department (therapist 3, FG). This particular patient exhibited a less stable illness profile and more acute issues compared to the other participants, including suicidal tendencies. Subsequently, this was identified as an exclusion criterion for future VReedom training application or studies.

3.2. Data collection and evaluation procedures

Preliminary results extracted from the patient dossier showed low overall stress and low variability in stress levels in anticipation of authorized leave among the included patients. The retrospective chart review uncovered certain limitations associated with these (therapist) reported stress levels, which could potentially restrict the analysis possibilities. One notable limitation was the presence of a considerable number of missing data entries ($M = 26.96\%$, $SD = 22.739$). Data is most likely missing in situations when there is a low level of stress, because staff is more likely to report higher levels of stress. Moreover, a substantial portion of the entered patient data remained at a low stress level 0 or 1, with seven patients with 100% of the data entries on stress level 0 or 1, two patients with 95% stress level 0 or 1, and one patient with 63% stress level 0 or 1.

Regarding the comprehensibility of the weekly participant questionnaire, some participants encountered difficulties in understanding certain questionnaire items, emphasizing the need for improved clarity in future iterations (patient 3, 6, 8). Also, therapists stated the need to simplify the level of questioning for the patients, as evaluative questions about specific emotions were often not fully understood or misunderstood due to their broad and abstract phrasing (therapist 1, 2, FG; therapist 1, 2, Q). Emotional inquiries were recommended to be simplified by focusing on specific emotions, such as anger or sadness, rather than using more complex emotions like disgust (therapist 2, FG). Additionally, enhancing questioning precision by asking more direct and targeted questions was suggested (therapist 2, FG; therapist 2, Q). Each session included post-session evaluation and de-brief, which allowed for oversight of the treatment process and patient evaluation (therapist 1, FG). However, therapists reported a need for additional time for post-session debriefing to enable a more comprehensive discussion on the effects of the therapy between the patient and therapist (therapist 3, FG). The evaluation should also focus on patient-centered outcomes and involve comparing the questionnaire results from the first and last session, as well as discussing progress with the patient (therapist 3). Finally, therapists recommended to include session-specific questions, to accommodate the unique aspects of each session (therapist 1, FG).

3.3. Acceptability and suitability of training and protocol

Regarding the technical applicability of the protocol, the distribution of triggers during difficult sessions was found to be evenly distributed (every 20 s), and the number of avatars adjusted to the difficulty level appeared proportional and suitable (therapist 1, 2, FG; therapist 1, 4, Q). The content of spoken texts by avatars was generally effective in triggering patients and creating a realistic experience, although there were instances where the sentences felt random and disconnected from the contextual situation (therapist 1, 2, FG). Greater variety in locations was recommended to be incorporated into the virtual environment to better simulate real-life situations in the CleVR environment, as currently only the supermarket was included as a virtual environment (therapist 1, 2, 3, FG). In terms of the contents of the protocol, the session program and sequence outlined in the manual were deemed comprehensible and feasible, and the therapists adhered to the main guidelines (therapist 1, 3, FG; therapist 1, 2, 3, 4, Q). However, there was a need to enhance the opportunity for improvisation and personalization during therapy sessions, as well as to schedule dedicated sessions for role-plays tailored to the patient's current developments and concerns (therapist 1, 2, FG). This included allocating more attention to identifying the specific triggers of each patient at the beginning of the training process (therapist 1, 2, 3, FG; therapist 1, 4, Q). A more comprehensive assessment of the patient's sources of anxiety and the establishment of specific treatment goals from the initial session were also suggested (therapist 1, 2, FG). Session duration was generally manageable, with sessions rarely exceeding the allocated time (45–60 min), but there were variations depending on individual needs (therapist 1, FG). However, improvements were needed in the duration of role play modules and the supermarket walk to fully utilize their potential and provide more challenging content (therapist 1,3, FG). The progression in the difficulty level of sessions was generally appropriate, although there was a suggestion to increase the level of challenge in certain instances (therapist 1, FG; therapist 1, Q). Flexibility in adjusting the number of sessions to individual needs of the patient, based on the assessment of the therapist, was also recommended (therapist 1, FG; therapist 1, 2, 3, 4, Q).

3.4. Training management and implementation

Within the area of training management and implementation, specifically regarding the support provided by therapists and their abilities to manage the training, the therapists' assistance was considered sufficient and adequate, with participants appreciating the support they received during the sessions (patient 1, 2, 3, 7, 8). The role-playing guidance was particularly praised for its clarity and effectiveness, demonstrating the therapists' competence in managing complex situations (patient 7, 8). Moreover, technical assistance was deemed satisfactory, with minimal waiting times and supportive interactions with the technical assistant (patient 1). However, participants expressed the need for the therapists' presence during the VReedom training, as they found the VR technology to be complex and overwhelming to handle independently (patient 1, 2, 4, 5, 8, 9). During the initial months of the training period, patients were occasionally instructed to take a virtual walk around the clinic, as part of their treatment. Subsequently, a tablet was integrated with the VR headset, allowing the therapist to have a real-time view of the patient's visual experience and provide necessary guidance or adjustments as required. Furthermore, the lack of visual contact with the therapist during the VR walk-through led to feelings of loneliness and disconnection from the therapy environment (patient 4). Despite the therapist being physically present with the patient, the use of the VR headset made it more challenging for patients to establish visual contact with the therapist. This, coupled with the immersive nature of the VR experience, occasionally disrupted the desired level of contact between the patient and the therapist. This was only the case, though, for the walk-through, as during role-play this lack of visual contact was actually desired, given that the therapist would be present as a virtual agent in the virtual environment in those circumstances (therapist 1, FG). In those cases, lack of visual contact could be utilized an instrument to overcome therapeutic boundaries (therapist 1, FG).

Subsequently, findings reveal that therapists felt capable of conducting the treatment after the initial training, if technical support was available (therapist 3). Therapists expressed willingness for future sessions. This was partly due to the fact that the peripheral matters were in order. Therapist were aware of what was expected from them, and all ancillary details were well organized by the research team and the clinical team (therapist 1, 2, 3, FG). The technical support facilitated quick access and immediate work commencement (therapist 3, FG). However, more practice time was recommended to enhance therapists' proficiency and comfort with the equipment (therapist 3, FG). Tailoring treatment sessions to individual needs was important, considering personal attention and technical support, while flexible scheduling was crucial (therapist 2, FG). Incident management ensured patient well-being, interdisciplinary communication, and patient safety through therapist expertise and patient resilience (therapist 1, 2, 3, FG). The

technology and equipment were deemed user-friendly, but therapists suggested clearer instructions and reduced dependence on technical assistance (therapist 1, 2, FG). Organizational imperatives and therapist investment were crucial for successful VR treatment. Subsequently, communication with the specific departments where the patients resided could be enhanced (e.g., structural meetings in which information is shared about the contents of every specific training session, so all personnel involved has an idea of what the patient has been exposed to) (therapist 1, 2, 3, FG). Ultimately, it was advised to allocate supplementary treatment hours to allow for an adequate duration for therapists to become acquainted with the VR technology, thereby enhancing their proficiency and comfort in utilizing it. This requirement primarily arises in cases where technical support is not available for therapists (e.g., due to factors such as financial constraints or staffing shortages) (therapist 3, FG).

3.5. Preliminary participant results

The translation of VR experiences to real-life physical leave demonstrated the potential for the activities in VReedom to be comparable and applicable to real-life situations (therapist 1, 2, 3, FG; therapist 1, Q). Participants reported feeling 'good,' 'normal,' 'amazed,' 'adventurous,' and expressing happiness in being somewhere else, outside, or even in a supermarket during the VReedom training sessions (patient 1, 3, 4, 6, 7, 8). However, negative emotions were also observed, including anger (toward unfamiliar virtual agents or actions performed by these virtual agents) and feelings of being overwhelmed, grossed out, unpleasant, and tense due to encountering new and unfamiliar situations (patient 2, 4, 5, 9, 10). This display and expression of emotional expression highlighted the efficacy of personalization in addressing individual contexts, and triggering authentic emotional responses (therapist 2, 3, FG; therapist 1, 2, 3, Q).

These results were supported by the therapists' statements. The included patients demonstrated engagement, curiosity, and motivation toward the Vreedom training, and estimated the training to be successful in the reduction of stress among patients (therapist 1, 2, 3, FG). Patients with specific treatment goals, such as fear of being alone or fear of crowded places, were found to benefit the most from the treatment (therapist 1, FG). The population with tbs-measure seemed to respond well to the training and showed an urge to participate in new activities within the clinic (therapist 1, 2, FG). However, the population with conditional tbs-measure displayed lower interest in training, possibly due to their shorter disconnection from society and reduced need for preparation for life outside the clinic (therapist 2, FG). Adjustments are necessary to optimize the treatment for the population with conditional tbs-measure, including increasing its intensity and tailoring it to address specific problems (therapist 1, 2, FG; therapist 1, 2, Q). It was also recognized that the tbs-measure, conditional tbs-measure, and care authorization population should all be treated as distinct target groups (therapist 3, FG). As for example, the care authorization population followed a less structured and sequential approach, with unsupervised leave occasionally serving as the initial step instead of supervised leave. Decisions regarding the progression of leave were contingent upon the patient's progress and their ability to adapt. This approach involved a trial-and-error process, which differed from the more structured approach observed in the (conditional) tbs-measure population (therapist 3, FG).

4. Discussion

This study conducted a retrospective assessment to evaluate the feasibility of the VReedom training, which specifically targets the preparation of first authorized leave for forensic psychiatric patients. The assessment aimed to optimize and refine the VReedom treatment, treatment environment, and treatment protocol. Five objectives were employed to assess feasibility: (1) recruitment capacity and the resulting sample characteristics, (2) procedures for data collection and evaluation, (3) acceptability and suitability of the training and protocol, (4) management and implementation of the training, and (5) preliminary participant results. The analysis of these objectives predominantly indicated the feasibility of the training, as became evident from both training and research perspectives. The results concurrently revealed several areas for improvement that could enhance the content and implementation of the training in the future.

Despite the absence of a control group for comparison, the findings of this study are consistent with the existing literature, suggesting that the incorporation of VR into treatment has the potential to enhance treatment motivation (Ticknor, 2019; Kip and Bouman, 2021). Out of the 13 patients approached, only 3 declined participation, resulting in a participation rate of 10 out of 13. Importantly, there were no dropouts among the 10 participants, as they all successfully completed the full training program, and within the period of time in which the training was offered, no incidents have occurred among the participating patients. Considering how the occurrence of incidents, within correctional mental health facilities seem to hinder the authorization of leaves, prolong treatment duration, and decrease the frequency of authorized leaves granted to patients, this could be beneficial to patient progression and occupancy (Mevis, 2011; Ter Horst et al., 2015; Watson and Choo, 2020).

Subsequently, the results demonstrated the successful elicitation of emotions, as exposure to relevant stimuli triggered the patients in several ways. Emotion elicitation is a crucial mechanism for exposure based interventions, and this finding aligns with existing literature supporting the efficacy of VR Exposure therapy as a promising approach for stress and anxiety reduction (Garcia-Palacios et al., 2001; Geraets et al., 2021). These findings warrant future research aimed at both assessing the effects of VR (assisted) therapies and advancing the development of VR-assisted training and treatment modalities.

Furthermore, the present study underscored the significance of establishing favorable implementation conditions for introducing a novel intervention within a clinical setting, as the clinic initiated the development of this VReedom training program. In the current case, this was effectively achieved, as treatment providers appraised the training as valuable to provide, seemly effective in the reduction of stress, and enjoyable to engage in, expressing their commitment to future training sessions. This positive reception primarily stemmed from the support received from the organization and staff in facilitating the delivery of the training and its developmental process, thereby affirming the existence of an enabling implementation environment and substantiating the feasibility of conducting such training within a forensic clinic.

The study findings align with previous literature, highlighting the perceived value of VR in the context of forensic psychiatry (Sygel and Wallinius, 2021), and correctional rehabilitation (Cornet and Van Gelder, 2020). However, as the treatment providers already tailored the session to the specific needs of each participant to a certain extent, results indicated that there was a need for even more personalization and more location options (treatment modules) to be able to vary more with difficulty levels, especially during the roleplay. Currently, only the supermarket is included as location in the VReedom training, but during physical leave, patients also practice using public transportation, visiting parks, and shopping centers. These locations are also available in the CleVR software, thus enabling exploration of the possibility of incorporating them into a revised protocol. Subsequently, areas for improvement pertaining to the optimal treatment population were identified. There were indications suggesting that the treatment is particularly suitable for patients admitted with a tbs-measure, given the associated trajectory of leave. With appropriate modifications, like building on a more flexible and possibly shorter protocol, the training has the potential to be adapted for other patient groups, including the conditional tbs-measure and care authorization, thereby increasing its versatility and applicability.

Preliminary results have also shown that patients showed low overall stress levels in anticipation of authorized leave, and that there was little variability in their stress levels. These results could indicate that the VReedom training was successful in reducing stress in anticipation of authorized leave, but such a claim cannot be substantiated without an appropriate control group. Alternative explanations include that the (relatively small) cohort of patients were less prone to stress in anticipation of authorized leave or that the twice daily staff-reported stress-levels are not sensitive enough to capture (light) variations in stress levels.

Finally, there are several ethical challenges considering the application of virtual reality (VR) in general (including the possibility of experiencing cyber sickness) and within forensic psychiatry in particular. The utilization of VR in forensic psychiatry intersects with an ongoing debate regarding the balance between societal risk mitigation and safeguarding individual rights. While the intention of applying VR training often revolves around preemptively averting risks and decreasing anticipatory stress or unwanted behavior, the appropriateness of such an approach warrants thoughtful consideration. VR environments might alternatively be used in risk assessment, and become a prerequisite for receiving authorization for leave. Careful studies should be conducted in order to assess whether such applications are appropriate and legitimate. This perspective underscores the emphasis on facilitating individuals' reintegration into society, as opposed to primarily focusing on punitive measures.

4.1. Strengths and limitations

The current study possesses several strengths. Firstly, it serves as the pioneering attempt to investigate strategies for enhancing authorized leave treatment in forensic mental health patients. This research has been conducted conscientiously, incorporating perspectives from both healthcare providers and patients. Additionally, the study includes a diverse sample of forensic mental health patients with severe psychiatric problems, representing various demographics and diagnoses. The absence of dropouts and incidents during the training period adds to the merit of this research. Consequently, this study makes a valuable contribution to the field of VR-assisted therapy research. Moreover, the study demonstrates fruitful collaboration among multiple stakeholders, including the clinical forensic mental healthcare organization, a software company, and researchers. The recruitment of interested therapists and the establishment of ongoing communication through peer evaluations, along with the active involvement of researchers in the evaluation process, contribute to a comprehensive understanding of the intervention.

In addition, it is important to acknowledge certain limitations. As it is determined whether an individual is actually granted leave based on the progress observed in a patient during the preparation for authorized leave, and thus during the period in which the patients receive the VReedom training, this could have affected the outcomes. Nevertheless, showing desired behavior could, of course, also be seen as progress. It is also crucial to acknowledge the potential presence of a novelty effect, considering the innovative nature of the current treatment method (VR). The probability of observing any change, regardless of the content and quality of the training, is naturally conceivable.

Furthermore, the questionnaires were sometimes deemed to lingually complex for current study population. This highlights the significance of conducting a feasibility study, as it aims to identify areas for improvement. Furthermore, the reliance on retrospective data collection from patient dossiers may have introduced limitations in terms of data availability and quality. This is particularly relevant since there was no control group that did not undergo the training leading up to the first authorized leave, which weakens the study's internal validity, and the small sample size, that limits the validity of the findings. Given that the measurement of effects was not the primary focus of the current study, it is advisable to incorporate a control group in future effectiveness studies, within a more diverse sample. Also, researchers utilized historical dossier information as primary data collection method in the current study. Considering the delicate nature of the target demographic, researchers predominantly adopted a retrospective approach, evaluating patient outcomes based on their records, as introducing an additional layer of engagement, such as subjecting them to focus group discussions, could potentially impose undue demands on the participants. The rationale for this approach was to minimize any undue burden on the vulnerable population under investigation. While researchers opted not to involve patients in focus group discussions for the feasibility study, the potential relevance of this methodology in subsequent research is acknowledged. In future larger-scale studies, incorporation of focus group discussions with patients to solicit direct insights is intended. Lastly, it is important to note that the current training was based on a self-developed protocol, lacking an evidence-based foundation. Therefore, the effectiveness of the training has not been examined. While the protocol shows promise, further development is necessary.

4.2. Recommendations

Further refinement of the treatment protocol is strongly recommended, incorporating the identified shortcomings, recommendations, and modifications outlined in the results section of this article. Moreover, enhancing the clarity and precision of questionnaires and outcome measures, based on feedback received

from therapists and patients, would significantly improve the reliability and validity of the collected data. To ensure the successful implementation and evaluation of the VReedom training intervention, it is crucial to provide ongoing training and support to the therapists involved. This should include technical assistance and opportunities for skill development. Such measures would not only contribute to treatment monitoring but also foster a favorable environment for intervention delivery. Additionally, the protocol should allow for greater flexibility and customization of the sessions to meet the individual needs of the patients. This can be achieved by incorporating virtual environments that simulate authorized leave scenarios and tailoring triggers to each participant. This personalized approach would enhance engagement and effectiveness. By considering these recommendations in future research, the field of clinical practice can harness the potential of VR in the context of the first authorized leave in forensic mental healthcare.

The employed method of data collection in the present study, specifically historical dossier analysis, exhibits limitations attributed to a substantial proportion of missing values and a general lack of stress among the included patients. This result can be interpreted in several ways. On the one hand, it is possible that this particular cohort consisted of patients who experienced minimal stress prior to the leave. On the other hand, it may suggest a disparity between informal signals from staff on increased stress in anticipation of authorized leave and the information documented in the electronic patient records. Lastly, it could indicate that the VReedom training indeed has a stress-reducing effect. However, drawing definitive conclusions is challenging since this study did not include a control group. Additionally, it raises doubts about the suitability of the stress level outcome measure as a reliable indicator of the effectiveness of the VReedom training intervention. As the presence of missing values suggests potential inadequacy of stress levels reported by treatment providers, necessitating the consideration of alternative outcome measures. Conducting prospective research in the future, which allows for better control over the predetermined desired outcome data, would be advisable to facilitate the use of Single-Case Experimental Design (SCED) or Randomized Controlled Trial (RCT) designs for rigorous investigation. Secondly, the predominantly low stress levels among patients can be attributed partly to their participation in the VReedom training. However, given the absence of a control group, it is also plausible that this specific patient population experiences inherently low levels of stress prior to leave. It is recommended to explore this aspect in future research endeavors through, for instance, developing and employing a stress-specific questionnaire, or utilizing physiological measurement instruments to assess physiological stress changes.

5. Conclusion

This retrospective study assessed the feasibility of VReedom training, designed to prepare forensic psychiatric patients for their first authorized leave. The evaluation aimed to refine treatment, environment, and protocol using five objectives: recruitment, data collection, acceptability, training management, and participant results. The analysis indicated training feasibility and highlighted areas for future improvement. While lacking a control group, the findings align with existing literature suggesting VR's potential to enhance treatment motivation. With a participation rate of 10 out of 13 approached patients and no dropouts, the study suggests VReedom's positive impact on patient progression and authorized leave frequency within correctional mental health facilities, potentially counteracting incidents that hinder leave authorization and prolong treatment duration. Results indicate that the training and study protocol is generally feasible, although some suggestions have been made to improve both. Further research should focus on evaluating its effectiveness on a larger scale, using a study design more appropriate for assessing training effects. The successful implementation of this training in preparing forensic patients for leave could be beneficial not only for the patients, but also for society and healthcare professionals.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

TP, JJ, and CH conceived the entire study and developed the study designs. CH and MS performed the qualitative analysis. CH coordinated the data collection and processing and wrote the manuscript. TP, JJ, AP, and LS critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was performed within a larger research project, financially funded by ZonMw (dossiernr. 04510012110005).

Acknowledgments

The authors would like to thank the study participants and the therapists Maud Schoemaker, Roxanne Stalenhoef, Stephanie Weninger and Anneke Noorlander, and Innovation Manager Hans ten Zijthoff from Inforsa Forensic Mental Healthcare in Amsterdam, Netherlands. Also, the authors would like to thank Nick Lommerse for his valuable contributions. Furthermore, the authors would also like to thank Eline Henrotte, Sam Klaassen, and Sarah Fernandes for their assistance in data collection. Finally, the authors would like to thank the organizations involved in this study project: CleVR, Amsterdam University Medical Center and Inforsa/Arkin.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1231619/full#supplementary-material

References

Bloch, K. E. (2021). Virtual reality: prospective catalyst for restorative justice. *Am. Crim. Law Rev.* 58, 285–329.

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., and Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Curr. Psychiatry Rep.* 19, 1–13. doi: 10.1007/s11920-017-0788-4

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., et al. (2009). How we design feasibility studies. *Am. J. Prev. Med.* 36, 452–457. doi: 10.1016/j.amepre.2009.02.002

Bowman, D. A., and McMahan, R. P. (2007). Virtual reality: how much immersion is enough? *Computer* 40, 36–43. doi: 10.1109/MC.2007.257

Campagnolo, D., Furimsky, I., and Chaimowitz, G. (2019). Absconsion from forensic psychiatric institutions. *Int. J. Risk Recov.* 2, 36–50. doi: 10.15173/ijrr.v2i2.3920

Cornet, L. J. M., and Van Gelder, J.-L. (2020). Virtual reality: a use case for criminal justice practice. *Psychol. Crime Law* 26, 631–647. doi: 10.1080/1068316X.2019.1708357

Easton, C. J., Berbary, C. M., and Crane, C. A. (2018). Avatar and technology assisted platforms in the treatment of co-occurring addiction and IPV among male offenders. *Adv. Dual Diagn.* 11, 126–134. doi: 10.1108/ADD-03-2018-0003

Ferreri, F., Bourla, A., Mouchabac, S., and Karila, L. (2018). E-Addictology: an overview of new Technologies for Assessing and Intervening in addictive behaviors. *Front. Psychol.* 9:51. doi: 10.3389/fpsyt.2018.00051

Fromberger, P., Meyer, S., Jordan, K., and Müller, J. L. (2018). Behavioral monitoring of sexual offenders against children in virtual risk situations: a feasibility study. *Front. Psychol.* 9:224.

Garcia-Palacios, A., Hoffman, H. G., Kwong See, S., Tsai, A., and Botella, C. (2001). Redefining therapeutic success with virtual reality exposure therapy. *Cyberpsychol. Behav.* 4, 341–348. doi: 10.1089/109493101300210231

Geraets, C. N. W., van der Stouwe, E. C. D., Pot-Kolder, R., and Veling, W. (2021). Advances in immersive virtual reality interventions for mental disorders: a new reality? *Curr. Opin. Psychol.* 41, 40–45. doi: 10.1016/j.copsyc.2021.02.004

Jansman-Hart, E. M., Seto, M. C., Crocker, A. G., Nicholls, T. L., and Côté, G. (2011). International trends in demand for forensic mental health services. *Int. J. Forensic Ment. Health* 10, 326–336. doi: 10.1080/14999013.2011.625591

Kaplan, J. S., and Tolin, D. F. (2011). Exposure therapy for anxiety disorders: theoretical mechanisms of exposure and treatment strategies. *Psychiatr. Times* 28:33.

Kawulich, B. B. (2005). Participant observation as a data collection method. *Forum qualitative sozialforschung/forum: Qual. Soc. Res.* 6:466. doi: 10.17169/fqs-6.2.466

Kip, H., and Bouman, Y. (2020). Van denken en praten naar doen en ervaren: Op zoek naar de toegevoegde waarde van technologie voor de forensische psychiatrie. Sancties, 249–260.

Kip, H., and Bouman, Y. H. A. (2021). A perspective on the integration of eHealth in treatment of offenders: combining technology and the risk-need-responsivity model. *Front. Psych.* 12:703043. doi: 10.3389/fpsyt.2021.703043

Kip, H., Kelders, S. M., Weerink, K., Kuiper, A., Brüninghoff, I., Bouman, Y. H., et al. (2019). Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. *Front. Psychol.* 10:406. doi: 10.3389/fpsyg.2019.00406

Klein Tuente, S., Bogaerts, S., Bulten, E., Keulen-de Vos, M., Vos, M., Bokern, H., et al. (2020). Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *J. Clin. Med.* 9. doi: 10.3390/jcm9072258

Kuosmanen, A., Tiihonen, J., Repo-Tiihonen, E., Eronen, M., and Turunen, H. (2021). Nurses' views highlight a need for the systematic development of patient safety culture in forensic psychiatry nursing. *J. Patient Saf.* 17, e228–e233. doi: 10.1097/PTS.0000000000000314

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., and Phelan, J. C. (2001). Stigma as a barrier to recovery: the consequences of stigma for the self-esteem of people with mental illnesses. *Psychiatr. Serv.* 52, 1621–1626. doi: 10.1176/appi.ps.52.12.1621

Mevis, P. (2011). Opnieuw de TBS: laat svp tijd en ruimte voor rustige en bestendige ontwikkeling in de goede richting. *Delikt en Delinkwent* 2011, 339–364.

Orsmond, G. I., and Cohn, E. S. (2015). The distinctive features of a feasibility study: objectives and guiding questions. *Occup. Particip. Health* 35, 169–177. doi: 10.1177/1539449215578649

Pillow, D. R., Zautra, A. J., and Sandler, I. (1996). Major life events and minor stressors: identifying mediational links in the stress process. *J. Pers. Soc. Psychol.* 70:381. doi: 10.1037/0022-3514.70.2.381

Porporino, F. J. (2010). "Bringing sense and sensitivity to corrections: from programmes to 'fix' offenders to services to support distance" in *What Else works? Creative work with offenders*. eds. J. Brayford, F. Cowe and J. Deering (Taylor & Francis Group: Routledge).

Rehm, I. C., Foenander, E., Wallace, K., Abbott, J.-A. M., Kyrios, M., and Thomas, N. (2016). What role can avatars play in e-mental health interventions? Exploring new models of client–therapist interaction. *Front. Psych.* 7:186. doi: 10.3389/fpsyt.2016.00186

Sygel, K., and Wallinius, M. (2021). Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. *Front. Psych.* 12:673089. doi: 10.3389/fpsyt.2021.673089

Ter Horst, P., Jessen, A., Bogaerts, S., and Spreen, M. (2015). Behandelduur tot eerste toekenning tbs-verlof in een Nederlandse forensische psychiatrische kliniek; dossieronderzoek naar invloed van patiëntkenmerken en delict. *Tijdschr. Psychiatr.* 5, 314–322.

Ticknor, B. (2019). Virtual reality and correctional rehabilitation: a game changer. *Crim. Justice Behav.* 46, 1319–1336. doi: 10.1177/0093854819842588

Van der Wolf, M., Reef, J., and Wams, A. (2020). Wie zijn geschiedenis niet kent…: Een overzichtelijke tijdlijn van de stelselwijzigingen in de forensische zorg sinds 1988. (9036106184).

VERBISoftware. (2021). MAXQDA 2022 online manual Available at: maxqda.com/help-max20/welcome

Watson, T. M., and Choo, L. (2020). Understanding and reducing unauthorized leaves of absence from forensic mental health settings: a literature review. *J. Forensic Psychiatry Psychol.* 32, 181–197. doi: 10.1080/14789949.2020.1837917

West, M. L., Yanos, P. T., and Mulay, A. L. (2014). Triple stigma of forensic psychiatric patients: mental illness, race, and criminal history. *Int. J. Forensic Ment. Health* 13, 75–90. doi: 10.1080/14999013.2014.885471

Wheaton, B., Young, M., Montazer, S., and Stuart-Lahman, K. (2013). Social stress in the twenty-first century. *Handb. Sociol. Ment. Health*, 299–323.

Wild, T. S. N., Fromberger, P., Jordan, K., Muller, I., and Muller, J. L. (2018). Web-based health Services in Forensic Psychiatry: a review of the use of the internet in the treatment of child sexual abusers and child sexual exploitation material offenders. *Front. Psych.* 9:763. doi: 10.3389/fpsyt.2018.00763

frontiers | Frontiers in Psychiatry

TYPE Clinical Trial
PUBLISHED 16 November 2023
DOI 10.3389/fpsyt.2023.1239066

OPEN ACCESS

EDITED BY
Joan E. Van Horn,
De Forensische Zorgspecialisten (Netherlands), Netherlands

REVIEWED BY
Simon Chu,
University of Central Lancashire, United Kingdom
Sylvia Terbeck,
Liverpool John Moores University, United Kingdom

*CORRESPONDENCE
David Ivarsson
✉ David.ivarsson@kriminalvarden.se

RECEIVED 06 July 2023
ACCEPTED 30 October 2023
PUBLISHED 16 November 2023

CITATION
Ivarsson D, Delfin C, Enebrink P and Wallinius M (2023) Pinpointing change in virtual reality assisted treatment for violent offenders: a pilot study of Virtual Reality Aggression Prevention Training (VRAPT). *Front. Psychiatry* 14:1239066. doi: 10.3389/fpsyt.2023.1239066

Pinpointing change in virtual reality assisted treatment for violent offenders: a pilot study of Virtual Reality Aggression Prevention Training (VRAPT)

David Ivarsson[1,2,3]*, Carl Delfin[1,3], Pia Enebrink[4] and Märta Wallinius[1,3,5]

[1]Evidence-based Forensic Psychiatry, Department of Clinical Sciences Lund, Faculty of Medicine, Lund University, Lund, Sweden, [2]Swedish Prison and Probation Service, Norrköping, Sweden, [3]Centre for Ethics, Law and Mental Health, Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [4]Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden, [5]Research Department, Regional Forensic Psychiatric Clinic, Växjö, Sweden

Preventing relapse into violence and its destructive consequences among persistent re-offenders is a primary concern in forensic settings. The Risk-Need-Responsivity framework models the best current practice for offender treatment, focused on building skills and changing pro-criminal cognitions. However, treatment effects are often modest, and the forensic context can obstruct the delivery of interventions. Developing treatments for offenders should focus on the best method of delivery to make "what works work." Virtual reality (VR)-assisted treatments such as Virtual Reality Aggression Prevention Training (VRAPT) are a new and innovative approach to offender treatment. This pilot study followed 14 male violent offenders who participated in VRAPT in a Swedish prison context and measured changes from pre-treatment to post-treatment and 3-month follow-up in targeted aggression, emotion regulation, and anger. It also investigated potential impact factors (pro-criminal cognitions, externalizing behaviors, psychosocial background, and childhood adverse experiences). In Bayesian linear mixed effects models, participants showed a high probability of change from pre-treatment to post-treatment and to follow-up on all outcome measures. All outcome measures demonstrated a low probability of change from post-treatment to follow-up. Analysis of reliable change showed that participants' results ranged from recovery to deterioration. We discuss the implications of the study for VRAPT's impact on the target group, those who might benefit from the approach, and suggested foci for future studies in the field of VR-assisted offender treatment. The study was preregistered at the International Standard Randomized Controlled Trial Number registry (https://doi.org/10.1186/ISRCTN14916410).

KEYWORDS

virtual reality, offender treatment, aggression, prison, pilot study, violence

1. Introduction

Youth and adults cared for and incarcerated in various forensic institutions constitute a heterogeneous group with multifaceted problems in addition to crime, such as substance abuse and various psychological and psychiatric difficulties (1–3). Violent offenders stand out as a particularly important group to reach with risk-reducing interventions due to the destructive consequences of violence on victims and society (4). Persisting violence in high-risk offenders with early onset and complex needs such as personality disorders, substance abuse, and nonviolent criminality (5) is a challenge in forensic settings. Aggressive behaviors often persist in violent offenders with various mental disorders and problems with emotion regulation, impulsivity, and empathy. According to Smeijers et al. (6), research should focus on understanding the reciprocal relations of social information processing, emotions, and emotion regulation in violent offenders. Aggressive behaviors have been related to a lack of social problem-solving skills (7) and emotion dysregulation (8–10), which indicates various skill deficits among this group of offenders. Helping offenders with violent behavior learn anger control and interpersonal problem-solving skills may thus be especially important for reducing the risk for relapse in violent crime (11).

Interventions focused on violent offenders will be affected by target group factors, potentially confounding the treatments' impact. Such factors may be criminogenic needs, known but not properly addressed in the intervention, such as pro-criminal attitudes (12) and antisocial personality traits (13), but they can also include responsivity factors such as history of trauma (14) and psychiatric problems (3). Impact factors can also be related to how the treatment is facilitated (15), for instance the experience of presence in the virtual environment of Virtual Reality (VR)-assisted treatment (16). To increase the likelihood a treatment is effective and can handle the multifaceted problems among the group, treatment interventions should be based on the principles of risk, need, and responsivity (RNR; (15) for both adults (17, 18) and youth (19, 20). The RNR framework states that offender treatment should target individuals with the highest risk of relapse in crime (risk-principle), focus on dynamic risk factors associated with relapse (needs principle), and be adapted to general evidence of effective treatment and client-specific characteristics [responsivity principle (15)]. Understanding which needs are impacted by treatment and which needs, and responsivity factors impact treatment facilitation is crucial for treatment effectiveness.

Although cognitive behavioral treatment (CBT) programs for offenders has been supported by evidence to be effective in decreasing criminal recidivism (11, 21, 22), offenders' often complex needs place high demands on individually adapted, yet evidence-based, interventions. Some of the key components of effective treatment for violent offenders are behavioral and skills training (e.g., emotion regulation and social skills) through role-plays based on social problem-solving (22, 23). A recurring challenge in interventions provided in forensic institutions, however, is difficulties with contextual adaptation of such skills training. For both practical and safety reasons, it is difficult to create individually tailored practice situations in the forensic context. Thus, the generalization of skills is currently hampered in such institutions, presumably affecting the offenders' rehabilitation back to society. There is an urgent need to develop clinical practice in forensic settings, for example, VR technology can provide new opportunities (24).

The use and knowledge of VR as a tool to deliver psychological treatment is developing rapidly in the fields of mental illness and offender rehabilitation. Its effects, which vary in nature and degree, have been demonstrated for several psychiatric disorders such as PTSD, social phobia, schizophrenia, specific phobia, and panic disorder (25). One study focusing on assessment of reactive aggression in students using immersive VR indicated that higher self-reported aggression was correlated to shorter reaction times for aggressive behavior in VR. The VR task was also a better predictor for past violence than self-assessment. This shows promise to VR-assessment in aggression and the authors concluded that future research in the area could be used for clinical samples such as violent offenders (26). VR-assisted treatment for offenders has been described as promising; adding VR as a complementary method to existing treatments creates opportunities for both adapted treatment (27) and controlled research (28). In a recent systematic review, the authors stated that immersive VR-assisted assessment and treatment is feasible and acceptable for offenders, but the evidence for implementing any specific VR intervention remains insufficient (29). In addition to being a tool for interventions, VR is also unique and powerful in creating immersive experiences that can lead to adaptations in responsivity. The two concepts at the heart of understanding the responsive nature of VR are immersion and presence (30). Immersion is best understood as the VR system's ability to support natural contingencies for perception. Presence is the combination of *place illusion,* the sense of being in the virtual environment, and *plausibility illusion*, the sense that virtual events are actually happening (30, 31). The experiences of spatial presence, involvement, and realness are key factors in measuring presence in an immersive VR experience (16).

Virtual reality-focused research has approached offender treatment from various informative angles. In a study presenting a protocol for aggressive impulse management using the VR-GAIME system (32), Aggression Replacement Therapy (33) was evaluated in a randomized controlled trial with VR added on to impact approach and avoidance behaviors in provoking situations. In VR-GAIME, the participants receive training in decision-making with the task of avoiding disagreeable avatars and approaching agreeable avatars. This study thus aimed at investigating the effect of a motivational intervention on social threat, impacting automatic approach behaviors displayed by individuals high in trait anger (32). Another study investigated criminal expertise (34) using VR, comparing offenders with and without burglary experience with nonoffending community participants, showing that burglars demonstrated a distinct set of burglary skills in relation to the comparison groups. The study could have implications for offender treatment and further reveal the automatic and habitual nature of expertise in decision making (35). Virtual Reality Aggression Prevention Training (VRAPT) is an example of a newly developed VR-assisted treatment (36, 37) aimed to reduce reactive aggression in offenders. The program is CBT based and consists of 16 individual treatment sessions delivered once or twice a week, making the program 8–16 weeks long. VRAPT focuses on skills training in emotion recognition, emotion differentiation, problem-solving, communication, and self-control of impulses and pro-violent cognitions. All sessions in VRAPT but the last include some sort of VR experience. The VR experience in VRAPT is expected to provide; skills training in environments not naturally found in the prison context, more intensive training sessions due to the immersive experience and tailored skills training addressing the needs of the participants to a higher degree than standard CBT-programs for offenders. All in all, the suggestion is that VR has the potential to make offender treatment more precise, intensive and resource efficient. The program starts with an introduction

to VRAPT and VR (Session 1), continues with assessment, emotion recognition, and differentiation (Sessions 2–6), skills training in role-play (Sessions 6–15), and ends with an evaluation of the treatment (Session 16). Each session lasts 45–60 min, with 10–40 min per session in VR according to the manual (38, 39). In addition to in-session activities, VRAPT now also includes between-session assignments (40).

Virtual Reality Aggression Prevention Training was evaluated for forensic psychiatric patients in a multicenter RCT (37). The authors showed positive post-treatment effects on self-reported aggression and hostility, anger control skills, anger expression, and impulsiveness, but no effect on staff-reported aggression or long-term effects at the 3-month follow-up. Possible reasons for the non-persistent results, the authors suggest, could be that the model was based on the social information processing model, which does not consider trauma history; that the target group was heterogeneous in psychiatric disorders; that there was a lack of generalization in skills because homework was not assigned, and the scenarios in VR did not match the patients' everyday life (37). In addition, self-assessment has limitations in a target group with cognitive deficits, behavioral skills were not explicitly measured, and the observational tools may not have been utilized optimally on the wards (37). The authors, however, recommended that future research focus on VR aggression treatment in other forensic populations such as clients in prison with aggressive behavior (37). VRAPT has subsequently been revised to address the initial RCT findings (40). It seems that VR-assisted interventions such as VRAPT can contribute to safer, ecologically valid, and effective interventions for violent offenders in forensic settings. Much work remains, however, to understand how VR-assisted offender treatment should be optimized.

1.1. Study design and research questions

The current study is a pilot study of the newly revised VRAPT (40) implemented in a prison setting. The study has a case-series within-group design with pre-treatment, post-treatment, and 12-week post-treatment follow-up measures. The overarching aim is to investigate the impact of VRAPT on key criminogenic needs related to aggressive behaviors, while highlighting important factors that may impact treatment outcomes, with the following specific research questions:

I. How do emotion regulation abilities and strategies, aggression, and anger change over time in imprisoned violent offenders participating in VRAPT?
II. Which important factors (e.g., experience of presence in the virtual environment, psychosocial background, psychiatric characteristics, pro-criminal attitudes, and prevalence of other externalizing behaviors including substance use) may impact the observed change over time in violent offenders?

2. Materials and methods

2.1. Sample

Participants were recruited from two medium- and high-security prisons in the Swedish prison and probation service (SPPS). To be included, possible participants had to (1) have a history of violent crime, (2) be sentenced to prison, (3) have been assessed with an increased risk (medium to high) of criminal recidivism, and (4) have an indicated need for treatment of aggression. Aggression was screened using pooled items from the risk and assessment tool Risk, Behov och Mottaglighetsbedömning (RBM_B) (41) with a cutoff value ≥8 indicating the need for treatment. The maximum value for the pooled items was 24, and the within-group range was 8–19 ($M = 14$, $SD = 3.76$). Exclusion criteria were (1) inability to understand and provide informed consent, (2) major deficits in understanding the Swedish language preventing active participation, (3) epilepsy, (4) indications of acute psychosis, (5) intellectual disabilities (IQ < 70), (6) acute suicide risk, (7) current and serious security risks preventing safe participation, and (8) less than 10 weeks prison time remaining. The inclusion and exclusion processes were part of regular sentence planning, and investigative staff identified the potential candidates for participation.

A total of 18 male offenders were recruited to the study during the years 2020–2022. Before treatment started, one participant dropped out, and three more dropped out during treatment. Drop-outs were client-initiated ($n = 3$) or administrative ($n = 1$) due to the participant's sudden transfer to a lower security prison where VRAPT was unavailable. The final sample thus consisted of 14 participants from the high-security ($n = 6$) and medium-security ($n = 8$) prisons. The participants were all violent offenders who had been assessed with a medium ($n = 2$) or high risk ($n = 12$) of relapse to criminality as measured by RBM-B. All participants had violence prevention programs as part of their prison treatment plan and were assessed by the study coordinator (first author DI) as eligible to participate in VRAPT.

Data on (1) current and past aggression, violence, and crime, (2) sociodemographic and psychosocial background, and (3) psychiatric problems, were collected from participants self-reports and structured data collection from file material.

2.2. Procedure

2.2.1. Data collection

Data were collected at four time points: (T0) Inclusion screening, (T1) pre-treatment (administered approximately 1 day to 1 week before start of treatment, adjusted after the participants' possibility to begin), (T2) post-treatment, and (T3) 12-week post-treatment follow-up. See Table 1 for an overview of data collection sources at the different time points. To ensure the reliability of the data, participants were offered support by research staff in answering self-assessment forms. This was done in order to mitigate impact of potential responsivity factors (e.g., impulsivity, reading disabilities, or attention deficits) on answering performance, for example by portioning the text in the questions for better readability.

2.2.2. Virtual reality aggression prevention training

The mean amount of treatment weeks for the current VRAPT study was 17.5 weeks (SD 11.6, median 13.5), ranging from 7 to 48 weeks. Only half of the VRAPT treatments followed the VRAPT protocol of 8–16 weeks of sessions due to the COVID-19 pandemic, general incidents at the involved prisons, and various logistical reasons. Two of the treatments were shorter than the stipulated

TABLE 1 Missing data on study measures at data collection points.

Measures	Missing at pre-treatment	Missing at post-treatment, *n*	Missing at follow-up, *n*
AQ-RSV (Pre-treatment–follow-up)	N/A	2	3
DERS (Pre-treatment–follow-up)	N/A	2	3
STAXI-2-S (Pre-treatment–follow-up)	N/A	2	3
ESI-BF (Pre-treatment)	N/A	N/A	N/A
CTQ-SF (Pre-treatment)	N/A	N/A	N/A
DSM-XC (Pre-treatment)	N/A	N/A	N/A
MCAA—part B (Pre-treatment)	N/A	N/A	N/A
IPQ (Post-treatment)	N/A	2	N/A
IPQ subscales	Internal missing, *n*		
Full scale	3		
Spatial presence	3		
Involvement	3		
Realness	2		
Global presence	3		

protocol (7 weeks) and 5 were longer, ranging 21–48 weeks. All participants received the same number of treatment sessions in accordance with the treatment protocol. Seven program facilitators were trained for the study, each of whom treated 1–5 participants, with a median of two participants per facilitator.

The VR environment was created with the software Social Worlds by CleVR using Oculus rift VR glasses, a high-performance laptop computer, a touch pad for the control of the VR environment, and a microphone and headphones for communicating with the participant in the VR environment. The software included three different functions that were used in VRAPT. The first function was "Walking around," where the participant could get acquainted with the virtual world. The second function included two parts: "Emotion recognition" and "Emotion differentiation," where the participant was introduced to avatars displaying different kinds of emotions. The participant was instructed to identify the correct emotions for the avatars. The third and final function that was used was real-time role-play where the program facilitator controlled the avatar's speech using voice distortion, body language, and emotional responses. Each virtual environment (e.g., park, supermarket, home, office, and prison) had several different situations where the participant could meet between 1 and 3 avatars.

2.2.3. Measures/instruments

Background data (e.g., psychosocial factors, criminal history, risk assessment, description of criminogenic needs, reports of misconduct, and individual plan for ongoing sentence) were collected from SPPS file materials. As a part of file data from RBM-B, we used the Drug Use Disorder Identification Test (42) and the Alcohol Use Disorder Identification Test (43) and the following self-assessment instruments for the different time points from pre-treatment to follow-up (see Table 1).

2.2.3.1. Aggression questionnaire-revised Swedish version

Aggression questionnaire is a self-assessment tool of aggression and hostile behavior containing 29 items spanning over four different factors: Physical aggression (PA), Verbal aggression (VA), Anger (AN), and Hostility (HS) (44). The items in the AQ-RSV version are measured on a five-point Likert scale (1 = least characteristic; 5 = most characteristic) (45). AQ is a highly used self-assessment instrument for aggression and has shown good psychometric properties [global internal consistency: alpha = 0.89 (44); and good generalizability to both the general population (46) and to prison samples (47)]. AQ-RSV has been shown as robust in translation and in a Swedish context (45).

2.2.3.2. Difficulties in emotion regulation scale

Difficulties in emotion regulation scale is a self-assessment tool containing 36 items that measure 6 dimensions of emotion recognition and regulation: Nonacceptance, Goals, Impulse, Awareness, Strategies, and Clarity (48). The items are measured on a five-point Likert scale (1 = almost never, 2 = sometimes, 3 = half the time, 4 = often, and 5 = almost always). The psychometric properties have shown high internal consistency, good test–retest reliability, adequate construct and predictive validity (48), good internal consistency and clinical and predictive utility when used with an adult sample with emotional disorders (49).

2.2.3.3. State–trait anger expression inventory-2-S

State–Trait Anger expression Inventory is a self-assessment of anger, both current and habitual, containing 57 items on the State–Trait anger scale (STAS) and the Anger expression scale (AX) (50). The items are measured on a four-point Likert scale for both State Anger items (1 = not at all, 2 = a little, 3 = rather, and 4 = very) and Trait Anger items (1 = almost never, 2 = sometimes, 3 = often, and 4 = almost always). STAXI-2 is one of the most commonly used instruments for assessing anger (51). The instrument, which is a revision of the STAXI (52), has excellent psychometric qualities in assessing anger (53). The adapted Swedish version, STAXI-2-S, has demonstrated good construct validity and appropriate reliability (54).

2.2.3.4. Externalizing spectrum inventory-brief form

Externalizing spectrum inventory-brief form provides a self-assessment of lifetime externalizing behaviors and contains 160 questions. ESI-BF (55) was developed from the conceptualization of the

externalizing spectrum to provide a more fine-grained assessment of impulsiveness/recklessness, substance abuse, and antisocial/aggressive behaviors (56). The items are measured on a four-point Likert scale (1 = true, 2 = partly true, 3 = partly false, and 4 = false) and summarized across the whole scale and on the subscales General Disinhibition (GD), Callous Aggression (CA), and Substance Abuse (SA). The factor structure for the ESI-BF was not confirmed when looking at a sample in a Dutch context, which the authors concluded could be due to cultural differences (56). The criterion validity analysis indicates that the ESI-BF could be more useful as a tool for prediction than as a measurement (56). A study on Swedish forensic psychiatric patients found that ESI-BF showed good to adequate reliability and internal consistency and good criterion validity, but an unclear structural fit (57).

2.2.3.5. Childhood trauma questionnaire-short form

Childhood trauma questionnaire-short form is a self-assessment of traumatic childhood circumstances, containing 28 items covering several types of childhood maltreatment and abuse (58). The items are measured on a five-point Likert scale (1 = never true, 2 = seldom true, 3 = sometimes true, 4 = often true, and 5 = very often true) and organized into five trauma types: emotional abuse, physical abuse, sexual abuse, emotional neglect, and physical neglect, which are assessed as either none (minimal); low to moderate; moderate to severe; or severe to extreme (58). CTQ is a well-established test of childhood trauma showing good psychometric properties (59), with research needed on test–retest reliability, measurement error, and criterion validity (60). CTQ demonstrates a strong level of evidence regarding adequate internal consistency, reliability, content validity, structural validity, and convergent validity, with CTQ-SF as a good alternative (61).

2.2.3.6. DSM-5 self-rated level 1 cross-cutting symptom measure

DSM-5 self-rated level 1 cross-cutting symptom measure provides a 23-item self-assessment of 13 domains of mental illness important to psychiatric diagnostics (62). Each item focuses on how often during the last 2 weeks; the participant has been bothered by the symptoms. The items are measured on a five-point Likert scale (0 = none or not at all; 1 = slight or rare, less than a day or two; 2 = mild or several days; 3 = moderate or more than half the days; and 4 = severe or nearly every day). When using DSM-XC, a rating of mild (i.e., 2) or greater on any item in 10 of the domains (depression, anger, mania, anxiety, somatic symptoms, sleep problems, memory, repetitive thoughts and behaviors, dissociation, and personality functioning) may guide decisions about additional assessments (62). For substance use, suicidal ideation, and psychosis, a rating of slight (i.e., 1) or greater on any item within the domain may serve the same purpose (62). The psychometric properties of DSM-XC were evaluated through the field trials of DSM-5 (63–65) and found to have adequate test–retest reliability for all items except the two on mania (64). When evaluated as a screening tool on a sample of healthy adults, the conclusion was that the DSM-XC proved to have a good specificity (66). However, DSM-XC was not developed as a screening tool; it can, however, be a good instrument for transdiagnostic assessment in research and clinical use (66).

2.2.3.7. Measures of criminal attitudes and associates part B

Measures of criminal attitudes and associates gives a self-assessment of pro-criminal attitudes, containing 46 items measured across the areas of Violence, Entitlement, Antisocial Intent, and Attitudes Toward Associates (12). The items are measured on a four-point Likert scale (1 = disagree, 2 = undecided, 3 = agree, and 4 = agree completely) in the Swedish version of the questionnaire (67). The scale has shown acceptable levels of reliability and validity in a sample of incarcerated males (12), and when translated to Swedish and evaluated with an offender sample and a public sample, it showed satisfactory psychometric properties (67). MCAA has shown good predictive validity for relapse to both general and violent crime (68), and it is useful in understanding the dynamic risk factor of criminal attitudes (12, 68).

2.2.3.8. Igroup presence questionnaire

Igroup presence questionnaire, used as a self-assessment of presence in the virtual environment, contains 14 items across three subscales: Spatial presence, Involvement, and Realness (16). The items are measured on a seven-point Likert scale, with higher ratings indicating a higher degree of experienced presence in VR. IPQ has shown good internal consistency across several translations (16, 69, 70).

2.3. Statistical analysis

R (version 4.2.1) was used for all statistical analysis. We analyzed change over time using robust Bayesian linear mixed effects models, with participant ID as the random effect. All Bayesian statistical models were specified using the R package brms (71), interfacing R with the Stan probabilistic programming language (72). Robustness was achieved using Student *t* likelihood (73), which alleviates the impact of potential outliers. Furthermore, all priors were chosen to be weakly informative and to thus having negligible impact on obtained estimates while still providing moderate regularization (74). Model sampling using Markov Chain Monte Carlo (MCMC) was conducted using four chains with 4,000 iterations each. All models converged well, with Gelman-Rubin diagnostics (R-hat) of 1.00 (75).

Results from Bayesian analyses are presented as the median posterior estimate of change between the time points, with the associated 90% highest density interval (HDI) presented within square brackets. The 90% HDI may be interpreted such that it has a 90% probability of containing the actual value. Since there is no notion of statistical significance in Bayesian statistics, we followed guidelines suggesting that a probability of 90% or higher can be considered very likely (76). We therefore considered an estimated change as robust if the 90% HDI did not contain zero and was very likely different from zero. In addition, we also calculated the probability of direction (*PD*), which is the probability, ranging from 50 to 100%, that the estimated change is either positive or negative (77). Since we were interested in lowered scores for all outcomes, we present the probability of the estimated change being negative. The *PD* has a 1:1 numerical correspondence with frequentist *p* values such that $P\text{two}-\text{sided} = 2 \times (1 - PD)$.

We investigated the impact of potential confounding factors separately by including each potential confounder as a covariate for each outcome. The leave-one-out cross-validated expected log predictive density ($ELPD_{LOO}$) (78); was then used to quantify and compare model fit. The relative model fit measure $ELPD_{LOO}$ provides an estimate of predictive accuracy for the model's out-of-sample fit compared to another model fit on the same data, but with a different set of variables. We multiplied obtained values by 1, so that lower

values of $ELPD_{LOO}$ indicated better model fit. Differences in $ELPD_{LOO}$ of <4 points are generally considered unreliable and of no clear predictive advantage, thus favoring the least complex model (79).

Finally, we conducted supplementary analysis using the reliable change index (RCI) developed by Jacobson and Truax (80) to further explore the direction of individual change. RCI is the individual post-treatment or follow-up measurement subtracted by the individual pre-treatment measurement divided by the standard error of the change between measures and is calculated by the following formulas:

$$RCI = \frac{X2 - X1}{S_{diff}}$$

$$S_{diff} = \sqrt{2(S_E)^2}$$

Standard error (SE) was calculated by multiplying the standard deviation for a normal population times the square of 1 – the instrument's internal consistency:

$$SE = S_1\sqrt{1 - r_{xx}}$$

According to the authors, a reliable change beyond ±1.96 is unlikely with an alpha level of $p < 0.05$ without a real change, therefore this indicates reliable change. The best cutoff for change indicating a move from a pathological level to a normal level uses the C criteria (80). Cutoff C is calculated using the SD for the clinical group multiplied by the mean for the normal population and adding this with the SD for the normal population multiplied by the mean for the clinical group divided by the added SDs from the normal population and clinical group. The calculation uses the following formula:

$$CutoffC = \frac{(SD_{clinical} x M_{non-clinical}) + (SD_{non-clinical} x M_{clinical})}{(SD_{clinical} + SD_{non-clinical})}$$

Based on RCI and cutoff, participants were divided into the categories *Recovered* (those who passed the cutoff and made a reliable change), *Improved* (those who did not pass the cutoff but made reliable change), *Unchanged* (neither passed the cutoff nor made a reliable change), and *Deteriorated* (made a reliable change, but in the wrong direction) (80). Participants below cutoff at pre-treatment, but who showed a reliable change were assigned to the *Improved* category. Missing values at post-treatment and follow-up assigned the participant to the *Unchanged* category. Due to missing data, reliable change was true for different participants for different time points.

3. Results

3.1. Sample characteristics

The mean age of the participants was 29 years (SD = 8.1, median = 28.5, range 20–49). The majority (71%, $n = 10$) were 20–30 years old. Most (64%, $n = 9$) had a high school degree, while 21% ($n = 3$) lacked that qualification. Two participants had a degree from senior high school. Fewer than half (43%, $n = 6$) owned their own homes at the time of imprisonment, and an equal number (43%, $n = 6$) were homeless. Two participants had temporary living conditions outside of prison. Almost all participants (93%, $n = 13$) had no work outside prison; only one worked or studied part-time. Foster home placement prior to age 18 was uncommon (valid for only 21%, $n = 3$). Repeated misbehavior before age 15 was, however, common (57%, $n = 8$), with occasional misbehavior being the second most common category (21%, $n = 3$), followed by mainly well-behaved (14%, $n = 2$) and unknown (7%, $n = 1$).

For self-reported mental health, overall and across all DSM-XC domains, 71% ($n = 10$) of participants were eligible for some sort of additional assessment. The number of domains above the cutoff for additional assessment ranged 1–11 for the whole sample, with a median of four domains above cutoff per participant. The most common domains above cutoff were depression and mania (50%, $n = 7$). Four participants had at least one domain reported as severe, with a range of 1–8 domains. See Table 2 for individual variety across domains.

File data on mental health showed that 50% ($n = 7$) had a diagnosis of mental disorder, with the most common being ADHD (43%, $n = 6$). Other indexed diagnoses were personality disorder (7%, $n = 1$), personality disorder due to organic brain damage (7%, $n = 1$), post-traumatic stress syndrome (7%, $n = 1$), and depression (7%, $n = 1$). In addition, various undiagnosed aspects of mental illness registered were anxiety and panic attacks, depressive mood, recurring nightmares, and a history of suicidal ideation. A total of 36% ($n = 5$) had no registered mental illness, and 14% ($n = 2$) had unspecified (depressive, anxiety) mental illness. Most of the group (79%, $n = 11$) had no history of suicide attempts and none of the participants had attempted suicide within the last year. Diagnoses of substance abuse disorders (29%, $n = 4$) and alcohol abuse disorder (7%, $n = 1$) were uncommon in the sample. However, 64% ($n = 9$) of the participants scored >0 on the DUDIT measure (81), with a mean of 20 (SD = 13.03 and median = 17) and a range of 3–40. A score of <6 on the DUDIT is an indication of substance use problems, and a score of ≥25 indicates that substance abuse syndrome is probable (82). The most common drug used in the entire sample was cannabis (64%, $n = 9$), followed by cocaine (43%, $n = 6$). Many participants had a history of using various drug (57%, $n = 8$). A total of 64% ($n = 9$) scored >0 on the AUDIT measure (83) with a mean of 8.1 (SD = 5.37; median = 10) and a range of 1–18. The mean for the subsample is in the zone indicating risky alcohol consumption (a score of 8–15) and the range carries over to the zone of problematic alcohol consumption (a score of 16–19) (82).

The participants' index crimes comprised 1–8 different offenses, covering a total of 31 different categories of offenses (e.g., attempted murder, murder, robbery, aggravated robbery, drunk driving, theft, and minor and major drug offenses). The most common index crime was robbery. The length of prison sentence ranged from 7 months to life imprisonment, and the range of prior prosecutions was 1–25 (M = 10.78, median = 10). For a summary of the participants' additional antisocial history and behaviors, see Table 3.

Misconduct during imprisonment was measured both before and during the VRAPT trial. A large proportion of the sample (79%, $n = 11$) had reported or suspected acts of misconduct prior to the pre-treatment measure, with ranges of 0–10 for reported misconduct (M = 3, SD = 3.11) and 0–8 for suspected misconducts (M = 2.85, SD = 2.68). Together, the range of reported and

TABLE 2 DSM-XC variety in domains.

DSM-XC domains	None, *n*	Slight, *n*	Mild*, *n*	Moderate, *n*	Severe, *n*	Total, *N*	Above cutoff, %
Depression	5	2	N/A	4	3	14	50
Anger	2	6	2	1	3	14	43
Mania	4	3	2	5	N/A	14	50
Anxiety	6	3	4	N/A	1	14	36
Somatic symptoms	7	3	1	2	1	14	29
Sleep problems	5	3	3	2	1	14	43
Memory	9	2	3	N/A	N/A	14	21
Repetitive thoughts and behaviors	8	3	1	1	1	14	21
Dissociation	7	5	N/A	N/A	2	14	14
Personality functioning	10	2	N/A	1	1	14	14
DSM-XC domains	**None**	**Slight****	**Mild**	**Moderate**	**Severe**	**Total**	**Above cutoff**
Substance use	N/A	N/A	N/A	N/A	N/A	14	N/A
Suicidal ideation	N/A	N/A	N/A	N/A	N/A	14	N/A
Psychosis	N/A	N/A	N/A	1	N/A	14	7%

*Cutoffs indicate depression, anger, mania, anxiety, somatic symptoms, sleep problems, memory, repetitive thoughts and behaviors, dissociation, personality functioning. **Cutoffs indicate substance use, suicidal ideation, and psychosis.

TABLE 3 Antisocial history and behaviors.

Antisocial history and behaviors (*N* = 14)			
Family members convicted of a violent crime: n	≥2: 3	1: 2	0: 9
Prior SPPS sentence: n	Yes: 9	No: 5	
Prior prosecutions: n	≥2: 9	1: 3	0: 2
violent crimes: n	≥2: 8	1: 2	0: 4
Age of 1st prosecution, years: n	<18: 8	18–20: 3	>21: 3
-age of first prosecution for violent offense, years: n	<18: 3	18–20: 4	>21: 7
Institutionalizations for misconduct: n	>2: 2	1: 6	0: 6

suspected misconducts was 0–18 (M = 5.85, SD = 5.05). No criminal acts were committed during imprisonment, and the number of participants with reports of violence (14%, *n* = 2) or suspected violence (14%, *n* = 2) was small. One participant, however, was responsible for nine acts of reported violence prior to the pre-treatment measure and another had three acts of suspected violence.

Externalizing behaviors, childhood adverse events, and pro-criminal attitudes as measured by the ESI-BF, CTQ-SF, and MCAA are presented in Table 4.

3.1.1. Time spent and presence in the virtual environment

The mean time spent in the virtual environment in a session was estimated by the participants as 24.5 min (range 10–45, SD = 10.12). Because the experience of presence as measured by the IPQ was severely skewed, we used the median and interquartile range (IQR) for this measure. The results were highly variable and ranged from −1.5 to 6. Average score for spatial presence was 6 (IQR = 0.5–8), general presence 2 (IQR = 1–3), involvement 1 (IQR = −4.5–4.5), and experienced realism −1.5 (IQR = −3–2).

3.2. Change in aggression, emotion regulation, and anger following VRAPT

3.2.1. Aggression

The estimated change in AQ-RSV scores from pre-treatment to post-treatment was 8.5 [1.16, 15.88], and the estimated change from pre-treatment to follow-up was 9.03 [1.4, 17.05]. There was no robust difference between post-treatment and follow-up (0.38 [−7.67, 8.25]). The probability that AQ-RSV scores were lower at post-treatment than at pre-treatment was 97.45%, and the probability that AQ-RSV scores were lower at follow-up than at pre-treatment was 96.74%. Figure 1 demonstrates changes in aggression as measured by AQ-RSV for the whole sample over all measure points.

The AQ-RSV trajectories of individual participants from pre-treatment to follow-up can be described as either decreasing at each measure point (*n* = 2), increasing at each measure point (*n* = 1), or showing variability in increase and decrease (*n* = 6). Two participants lacked post-treatment data but showed a general, decreasing curve, while three others also lacked follow-up data, but showed a decreasing trend between pre-treatment and post-treatment.

TABLE 4 Outcome measures and impacting factors at pre-treatment for the whole sample ($N = 14$).

Measures	Mean	SD	Range
AQ-RSV	94.4	23.4	39–124
DERS	92.9	23.5	44–120
STAXI-State	24	11.9	15–51
STAXI-Trait	24.7	8.1	11–39
STAXI-AX	53.9	17.9	13–76
ESI-BF	254.4	71	110–343
General disinhibition	36	12.7	13–51
Callous aggression	31.4	9.2	15–44
Substance abuse	25.9	8.9	12–38
CTQ-SF	60.5	6.9	50–72
Emotional abuse	8.1 (none–low)	3.7	5–15 (none–moderate)
Physical abuse	9.5 (low–moderate)	3.9	5–16 (none–severe)
Sexual abuse	-	-	-
Emotional neglect	17.1 (moderate–severe)	4.9	8–25 (none–severe)
Physical neglect	13 (severe)	2	10–17 (low–severe)
MCAA	128.1	29.4	70–160

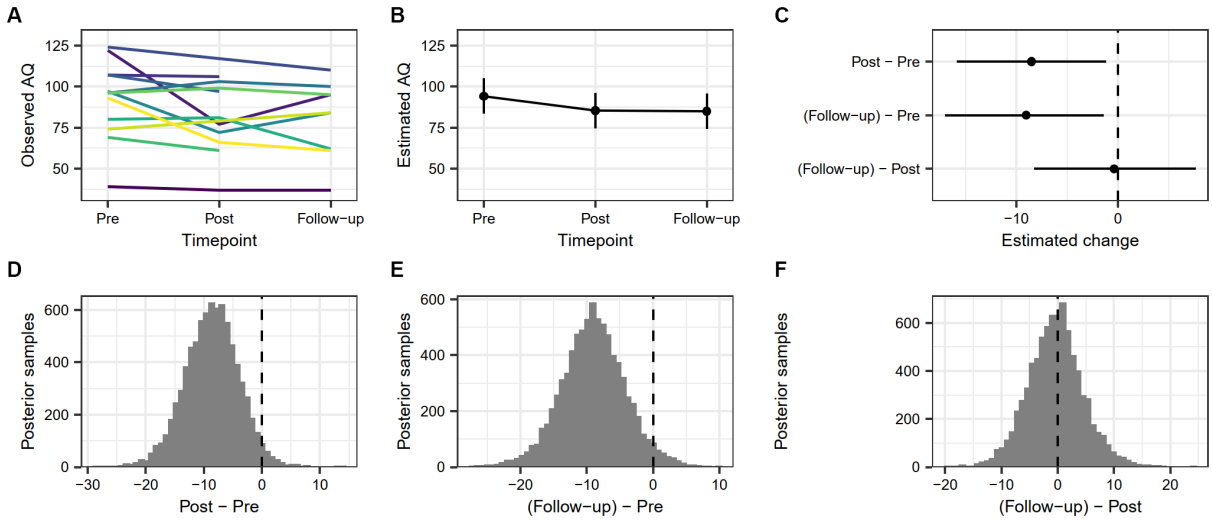


FIGURE 1
(A) Individual trajectories showing observed values of AQ-RSV score across timepoints. **(B)** Model-based estimate of AQ-RSV score across timepoints, with associated 90% highest density interval. **(C)** Model-based change in AQ-RSV score between timepoints, with associated 90% highest density interval. **(D)** Posterior distribution of estimated difference in AQ-RSV score from pre-treatment to post-treatment. **(E)** Posterior distribution of estimated difference in AQ-RSV score from pre-treatment to follow-up. **(F)** Posterior distribution of estimated difference in AQ-RSV score from post-treatment to follow-up.

Three participants showed higher aggression levels at follow-up than at pre-treatment. The rest of the sample ($n = 11$) had either lower aggression at follow-up than at pre-treatment ($n = 8$) or incomplete data for follow-up ($n = 3$). Stated differently, 73% of the participants with valid data showed decreased aggression between pre-treatment and follow-up.

Reliable change (RC) between pre-treatment and post-treatment as indicated by a value greater than ±1.96 was true for seven participants with complete pre-treatment and post-treatment measures. Three of these participants had a post-treatment value under the calculated cutoff value (84.7), indicating a status of recovered, and three showed an RC but did not cross the cutoff, indicating an improved status. One participant was in the Deteriorated category. The rest of the sample ($n = 7$) was unchanged. See Table 5 for complete RC data between pre-treatment to post-treatment.

Reliable change from pre-treatment to follow-up was true for eight participants with complete pre-treatment and follow-up data. Of these, 2 were classified as Recovered, 4 as Improved, and 1 as Deteriorated. The rest of the sample was Unchanged ($n = 7$). See

TABLE 5 Reliable change and cutoffs on outcome measures between pre- and post-treatment.

Measure	RC >±1.96*	Cutoff	Recovered, *n* (%)	Improved, *n* (%)	Unchanged***, *n* (%)	Deteriorated *n* (%)	Below cutoff at pre-treatment *n* (%)
AQ-RSV	−17.3–2.9	84.7	3 (21)	3 (21)	7 (50)	1 (7)	4 (29)
DERS	−30.7–3.0	85	4 (29)	5 (36)	4 (29)	1 (7)	4 (29)
STAXI-Trait	−15.2–(−2.2)	20.9	3 (21)	7 (50)	4 (29)	N/A	4 (29)
STAXI-AX	−24.5–3.2	42.2	6 (42)	4 (29)	3 (21)	1 (7)	3 (21)

*SE and S_{diff}: AQ-RSV = 1.85 and 2.62 (IC = 0.89**), DERS = 1.45 and 2.05 (IC = 0.93**), STAXI-Trait = 0.65 and 0.92 (IC = 0.88**), and STAXI-Anger Expression Index = 1.57 and 2.2 (IC = 0.88**). **AQ-RSV (44), DERS (48), and STAXI-2 (50). ***Missing Post-treatment data: *n* = 2.

TABLE 6 Reliable change and cutoffs on outcome measures between pre-treatment and follow-up.

Measure	RC >±1.96*	Cutoff	Recovered *n* (%)	Improved *n* (%)	Unchanged*** *n* (%)	Deteri-orated *n* (%)	Below cutoff at pre-treatment *n* (%)
AQ-RSV	−10.3–8.4	84.7	2 (14)	4 (29)	7 (50)	1 (7)	4 (29)
DERS	−18.5–2.9	85	4 (29)	4 (29)	5 (36)	1 (7)	4 (29)
STAXI-trait	−11.9–5.4	20.9	3 (21)	3 (21)	7 (50)	1 (7)	4 (29)
STAXI-AX	−16.4–5	42.2	4 (29)	4 (29)	5 (36)	1 (7)	3 (21)

*SE and S_{diff}: AQ-RSV = 1.85 and 2.62 (IC = 0.89**), DERS = 1.45 and 2.05 (IC = 0.93**), STAXI-Trait = 0.65 and 0.92 (IC = 0.88**), STAXI-Anger Expression Index = 1.57 and 2.2 (IC = 0.88**). **AQ-RSV (44), DERS (48), and STAXI-2 (50). ***Missing Follow-up data: *n* = 3.

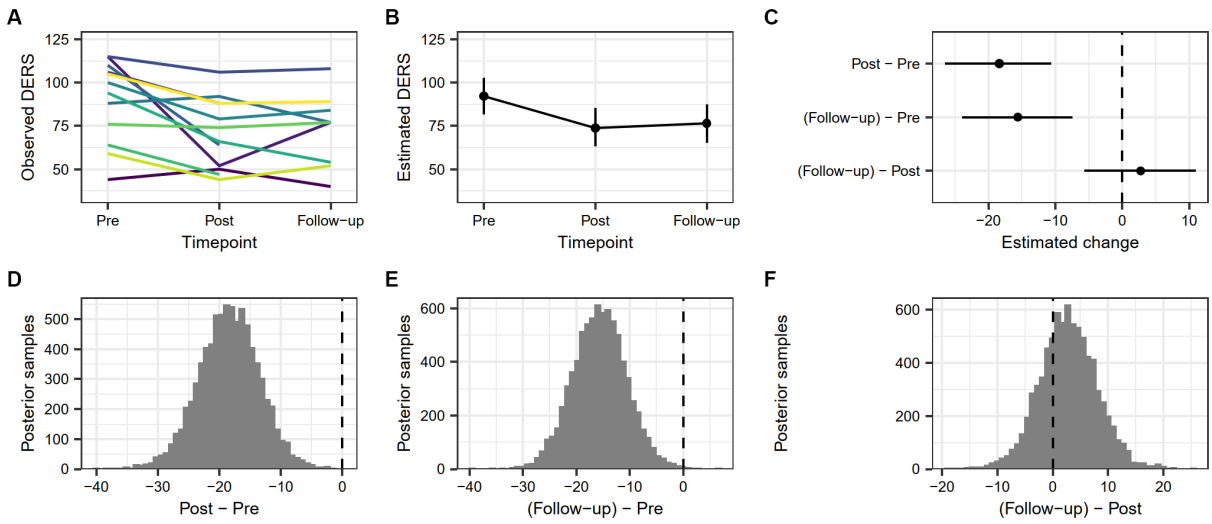


FIGURE 2
(A) Individual trajectories showing observed values of DERS score across timepoints. **(B)** Model-based estimate of DERS score across timepoints, with associated 90% highest density interval. **(C)** Model-based change in DERS score between timepoints, with associated 90% highest density interval. **(D)** Posterior distribution of estimated difference in DERS score from pre-treatment to post-treatment. **(E)** Posterior distribution of estimated difference in DERS score from pre-treatment to follow-up. **(F)** Posterior distribution of estimated difference in DERS score from post-treatment to follow-up.

Table 6 for complete AQ-RSV RC data from pre-treatment to follow-up.

3.2.2. Difficulties in emotion regulation

The estimated change in DERS score from pre-treatment to post-treatment was 18.37 [10.65, 26.53], and the estimated change from pre-treatment to follow-up was 15.6 [7.39, 23.94]. There was no robust difference between post-treatment and follow-up (−2.76 [−11.02, 5.67]). The probability that DERS scores were lower at post-treatment than at pre-treatment was 99.99%, and the probability that DERS scores were lower at follow-up than at pre-treatment was 99.89%. Change in emotion regulation difficulties for the sample as measured by DERS is demonstrated in Figure 2.

Different individual trends between pre-treatment to follow-up were evident, with emotion dysregulation either decreasing at every measure point (*n* = 1) or showing variability in increase and decrease (*n* = 8). Of two participants lacking post-treatment data, one showed a decrease and the other an increase in emotion dysregulation from pre-treatment to follow-up, and three others lacked follow-up data but showed a decreasing trend from pre-treatment to post-treatment.

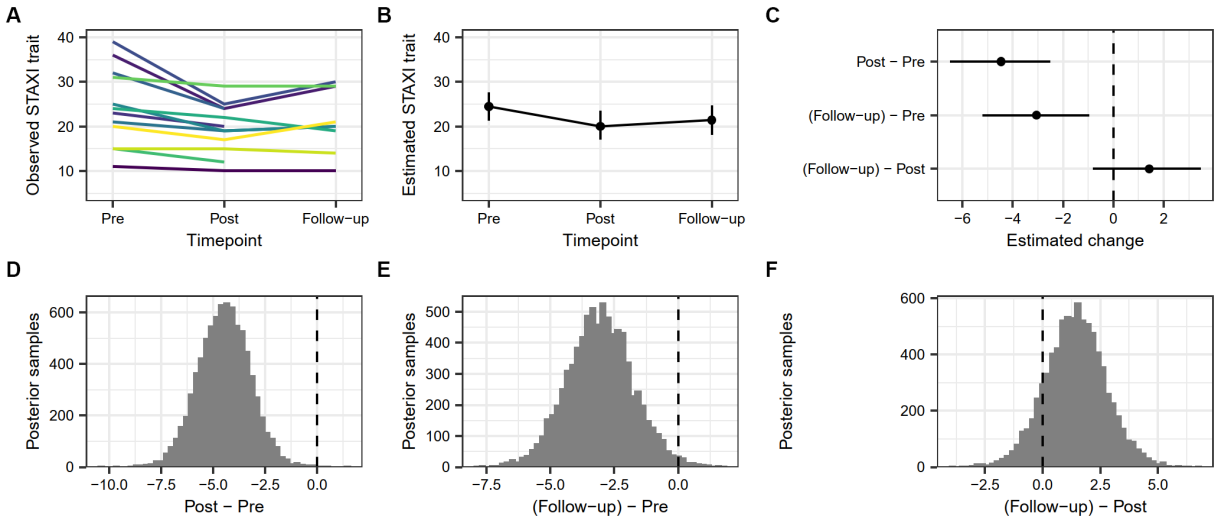


FIGURE 3
(A) Individual trajectories showing observed values of STAXI-Trait score across timepoints. **(B)** Model-based estimate of STAXI-Trait score across timepoints, with associated 90% highest density interval. **(C)** Model-based change in STAXI-Trait score between timepoints, with associated 90% highest density interval. **(D)** Posterior distribution of estimated difference in STAXI-Trait score from pre-treatment to post-treatment. **(E)** Posterior distribution of estimated difference in STAXI-Trait score from pre-treatment to follow-up. **(F)** Posterior distribution of estimated difference in STAXI-Trait score from post-treatment to follow-up.

Overall, two participants showed an increased level of emotion dysregulation from pre-treatment to follow-up. The rest of the sample ($n = 12$) had either incomplete data at follow-up ($n = 3$) or lower ($n = 9$) emotion dysregulation at follow-up than at pre-treatment. Thus, approximately 82% of the participants with valid follow-up data showed a decrease in emotion dysregulation from pre-treatment to follow-up.

Reliable change between pre-treatment and post-treatment as indicated by a score greater than ±1.96 was true for 10 participants with complete pre-treatment and post-treatment measures. Four of these participants had a post-treatment value under the calculated cutoff (84), indicating a Recovered status and five had a RC value above cutoff, indicating an Improved status. One participant had a Deteriorated status from pre-treatment to post-treatment. The rest of the sample ($n = 4$) was Unchanged. See Table 5 for complete data on DERS RC between pre-treatment and post-treatment.

Reliable change from pre-treatment to follow-up was true for nine participants with complete pre-treatment and follow-up measures. Of these, four were classified as Recovered, four as Improved, and one as Deteriorated. The rest of the sample was Unchanged ($n = 5$). See Table 6 for complete data on DERS RC between pre-treatment and follow-up.

3.2.3. Anger

3.2.3.1. State anger

Due to the low variance in the STAXI-State score, this subscale could not be used in the mixed effects models and was therefore left out of the analysis on probability of change between time points. For consistency, the STAXI-State score was also kept out of the RCI analyses.

Between pre-treatment and post-treatment, anger scores decreased at every time point ($n = 1$), varied between increasing and decreasing ($n = 6$), or showed no variability at all ($n = 2$). Of the three participants who lacked follow-up data, two showed a decrease in anger from pre-treatment to post-treatment, and one showed no trend. Of the two participants who lacked post-treatment data, both showed no trend between pre-treatment and follow-up. At follow-up, the sample had either incomplete data ($n = 3$), increased state anger ($n = 1$), no change ($n = 3$), or decreased state anger ($n = 5$) since pre-treatment. In other words, 45% of the sample with valid follow-up data showed decreased levels of state anger between pre-treatment and follow-up.

3.2.3.2. Trait anger

The estimated change in STAXI-Trait score from pre-treatment to post-treatment was 4.46 [2.51, 6.51], and the estimated change from pre-treatment to follow-up was 3.05 [0.97, 5.21]. There was no robust difference between post-treatment and follow-up (−1.43 [−3.48, 0.81]). The probability that STAXI-Trait scores were lower at post-treatment than at pre-treatment was 99.95%, and the probability that STAXI-Trait scores were lower at follow-up than at pre-treatment was 98.92%. Figure 3 demonstrates changes in trait anger as measured by STAXI-trait for the whole sample over all measure points.

Between pre-treatment and follow-up, members in the sample were either decreasing in trait anger at every measure point ($n = 1$) or showed variability in increase and decrease ($n = 8$). Three participants lacked follow-up data, and all of them showed a decrease between pre-treatment and post-treatment. Two participants lacked post-treatment data, with one showing an increasing trend and the other a decreasing trend between pre-treatment and follow-up. The sample had either incomplete data for follow-up ($n = 3$) or increased ($n = 2$) or decreased ($n = 9$) levels of state anger at follow-up than at pre-treatment. In other words, 64% of the sample with valid follow-up data showed decreased levels of state anger between pre-treatment and follow-up.

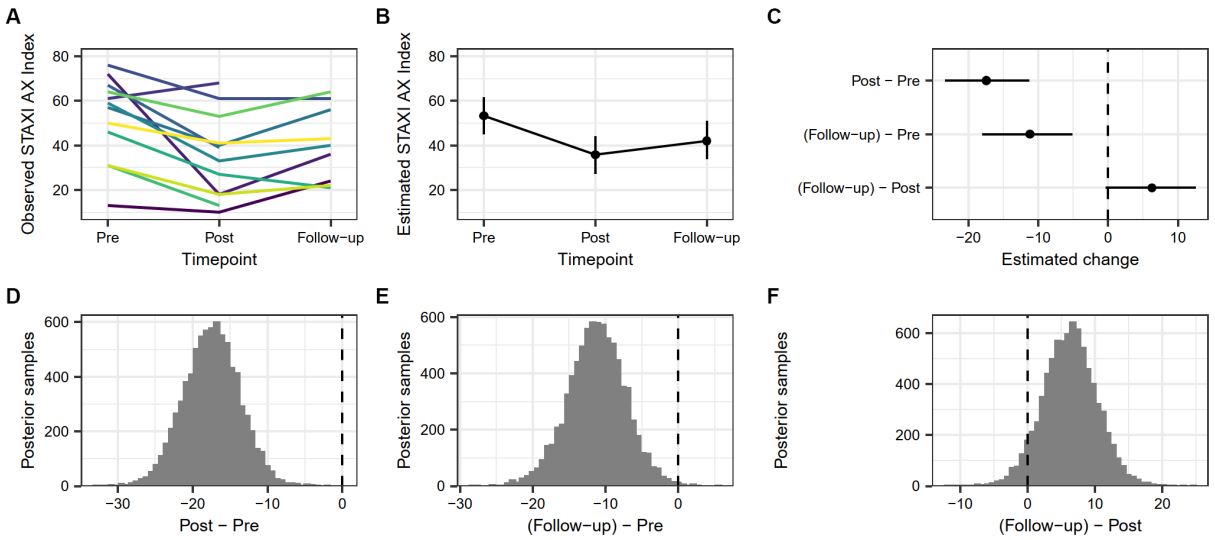


FIGURE 4
(A) Individual trajectories showing observed values of STAXI AX Index score across timepoints. **(B)** Model-based estimate of STAXI AX Index score across timepoints, with associated 90% highest density interval. **(C)** Model-based change in STAXI AX Index score between timepoints, with associated 90% highest density interval. **(D)** Posterior distribution of estimated difference in STAXI AX Index score from pre-treatment to post-treatment. **(E)** Posterior distribution of estimated difference in STAXI AX Index score from pre-treatment to follow-up. **(F)** Posterior distribution of estimated difference in STAXI AX Index score from post-treatment to follow-up.

Reliable change between pre-treatment and post-treatment as indicated by a score greater than ±1.96 was true for 10 participants with complete pre-treatment and post-treatment measures. Three of these showed a change below the cutoff (20.9), indicating a Recovered status after the treatment, and seven showed a change that did not drop below the cutoff, indicating an Improved status. The rest of the sample ($n = 4$) was Unchanged. All the unchanged participants were under the calculated cutoff at pre-treatment. No participant deteriorated in trait anger between pre-treatment and follow-up. See Table 5 for complete RC data between pre-treatment and post-treatment.

Reliable change between pre-treatment and follow-up was true for seven participants with complete pre-treatment and follow-up measures. Of these, three participants were classified as Recovered, three as Improved, and one as Deteriorated. The rest of the sample was Unchanged ($n = 7$). See Table 6 for complete RC data between pre-treatment and follow-up.

3.2.3.3. Anger expression

The estimated change in STAXI-AX Index score from pre-treatment to post-treatment was 17.43 [11.28, 23.39], and the estimated change from pre-treatment to follow-up was 11.18 [5.08, 18.02]. There was no robust difference between post-treatment and follow-up (−6.3 [−12.58, 0.37]). The probability that STAXI-AX Index scores were lower at post-treatment compared to pre-treatment was >99.99%, and the probability that STAXI-AX Index scores were lower at follow-up compared to pre-treatment was 99.75%. Figure 4 demonstrates changes in anger expression as measured by STAXI-AX Index for the whole sample over all measure points.

Different trends between pre-treatment and follow-up in the sample showed a decrease every measure point ($n = 1$) or variability between increase and decrease ($n = 8$). Of the three participants who lacked follow-up data, two showed a decrease between pre-treatment and post-treatment, and a single participant showed an increasing trend. Two participants lacked post-treatment data but showed a decreasing trend from pre-treatment to follow-up. The sample had incomplete data for follow-up ($n = 3$), increased state anger ($n = 1$), no change ($n = 2$), or decreased state anger ($n = 8$) between pre-treatment and follow-up. In other words, 57% of the sample with valid follow-up data showed decreased levels of anger expression between pre-treatment and follow-up.

Reliable change between pre-treatment and post-treatment as indicated by a score greater than ±1.96 was true for 11 participants with complete pre-treatment and post-treatment measures. Six of these showed a change below the cutoff (42.2), indicating a Recovered status post-treatment, four showed no change below cutoff and were Improved, and one participant was Deteriorated. The rest of the sample ($n = 3$) was Unchanged. Three participants were under the calculated cutoff at pre-treatment. See Table 5 for complete RC data.

Reliable change between pre-treatment and follow-up was true for nine participants with complete pre-treatment and follow-up measures. Of these, four participants were Recovered, four were Improved, and one was Deteriorated. The rest of the sample was Unchanged ($n = 5$). See Table 6 for complete RC data between pre-treatment and follow-up.

3.3. Possible factors impacting outcome from VRAPT

Analysis of potential confounding factors revealed no impact on the model's predictive performance for any confounding factor, regardless of outcome, with none of the $ELPD_{LOO}$ values showing a change of at least **four** points. The base model, without any impacting

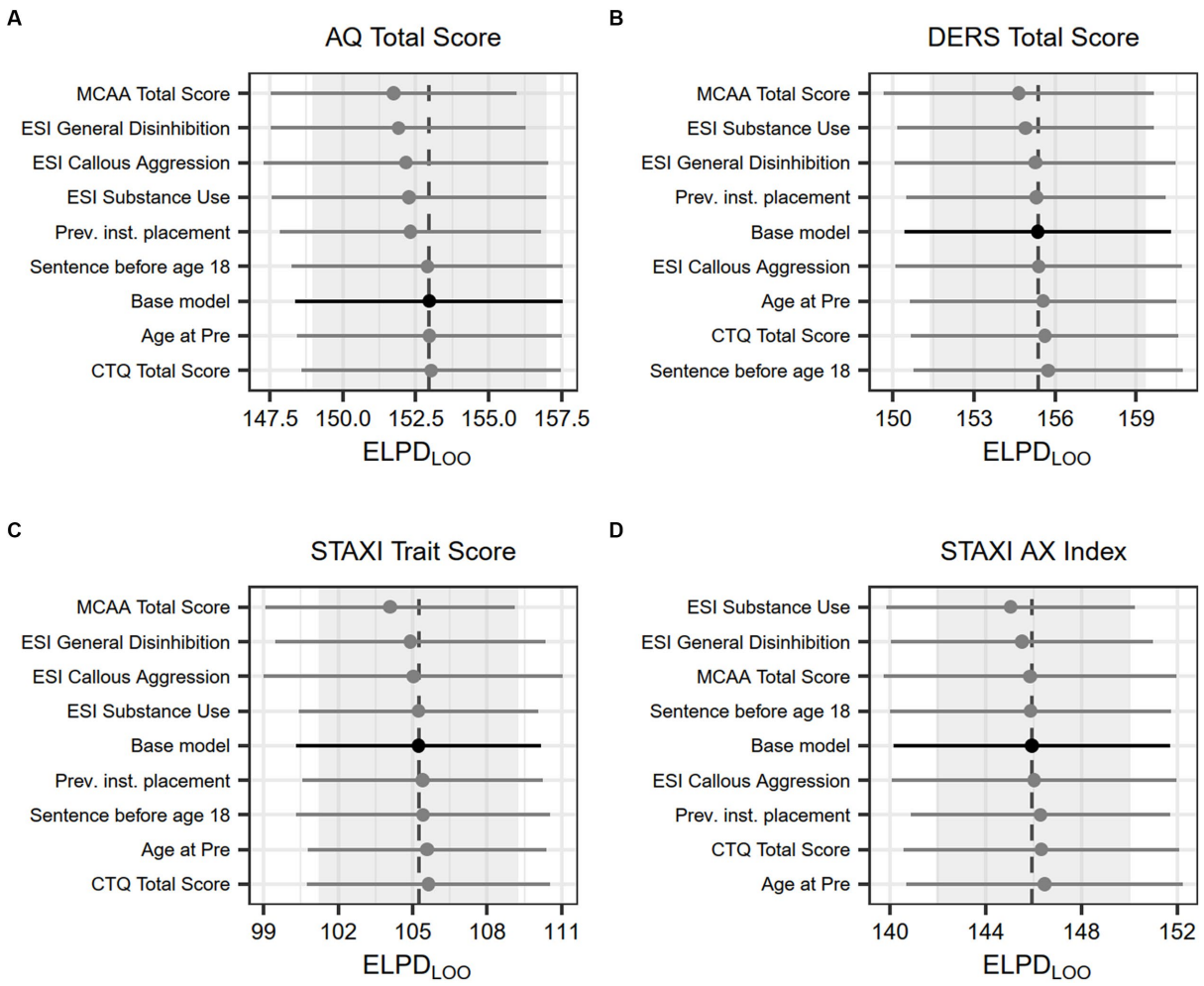


FIGURE 5
Effect of confounding factors on model fit. Dots show the average leave-one-out cross-validated expected log predictive density, lines show the associated standard error. A difference of four or less is considered unreliable and of no clear predictive advantage, thus favoring the base model.

factors, thus exhibited the best predictive performance. Results are visualized in Figure 5.

Reliable change analyses identified participants as either Recovered, Improved, Unchanged, or Deteriorated. In terms of impacting factors, the group that Deteriorated on at least one outcome measure ($n = 4$, 29% of N) was assessed as interesting for further exploratory analysis. The Deterioration group showed lower values across all measures except STAXI-State anger (equal) and ESI-BF SA and CTQ-SF (both higher). The CTQ-SF subscales showed that the Deterioration group had less experience of physical abuse (none–low) and more experience of emotional and physical neglect (moderate–severe) than the rest of the sample. See Table 7 for outcome measures and factors impacting the deterioration group.

4. Discussion

In this pilot study on violent offenders undergoing VRAPT treatment, we could estimate that a decrease in emotion dysregulation, aggression, and trait anger as well as anger expression was very likely (> 95% probability across all outcome measures) to occur between pre-treatment to post-treatment and follow-up. We could also show RC at a Recovered or Improved level for just under half or slightly more than two-thirds of the sample, respectively. However, no robust difference was demonstrated between post-treatment and follow-up and no robust impact of potential impacting factors on the model were demonstrated for any of the outcome measures.

Considering the first research question, "How do emotion regulation abilities and strategies, aggression, and anger change over time in imprisoned, violent offenders participating in VRAPT?" RC analyses revealed that 42% (6/14) of the sample recovered or improved in terms of aggression and as many as 65% (9/14) either recovered or improved in terms of emotion regulation between pre-treatment and post-treatment. For trait anger 71% (10/14) either recovered or improved, which was also true for anger expression. RC analyses of the period between pre-treatment and follow-up produced similar results, but with fewer participants showing RC, recovery, or improvement, and more participants staying unchanged, in line with the upward slope trend between post-treatment to follow-up shown in Figures 1–4. Consistently across outcome measures, a majority of the unchanged group was below the calculated cutoff C for RC at pre-treatment, possibly indicating a floor effect. RNR-based research

TABLE 7 Outcome measure and impacting factors for the deterioration group ($n = 4$).

Measures	Mean	SD	Range
AQ-RSV	84.2	30.6	39–107
DERS	89.5	33.0	44–120
STAXI—state	24	18.0	15–51
STAXI—trait	20.5	6.8	11–27
STAXI—AX	48	23.4	13–61
ESI-BF	250.2	94.4	110–311
General disinhibition	33.8	14.4	13–46
Callous aggression	27.3	10.0	15–39
Substance abuse	29.3	11.7	12–38
CTQ	65.3	4.6	60–70
Emotional abuse	7.5 (none)	5.0	5–15 (none–moderate)
Physical abuse	8.3 (none–low)	2.5	5–11 (none–low)
Sexual abuse	5 (none)	N/A	5–5 (none)
Emotional neglect	20.3 (severe)	5.0	15–25 (moderate–severe)
Physical neglect	15 (severe)	1.8	13–17 (severe)
MCAA	123.2	37.4	70–153

states that individuals with higher risk for relapse tend to benefit more from treatment (15, 85) and that offender treatment should focus on the appropriate criminogenic needs of the individual (15). The sample consisted of offenders with a medium to high risk of relapse into crime, making them all eligible for the treatment. However, it is possible that the outcome measures did not accurately assess some of the unchanged individuals' crimonogenic needs. It is possible that the unchanged participants could be different to the other participants, and that their changing might depend on impacting factors, even though our models could not show this. The unchanged individuals varied between different outcome measures, which highlights the heterogeneity of the offender population. More research is needed into individual offenders' various backgrounds and needs and treatment impact.

Four participants demonstrated some deterioration between pre-treatment and post-treatment and/or pre-treatment to follow-up, two of these on AQ-RSV and two on either DERS or STAXI. This is an important finding to highlight, since any potentially counterproductive results of a treatment must be known, especially in such a novel treatment as VRAPT. Although deterioration following treatment must be taken seriously, three of these semi-deteriorating participants also seemed to benefit from VRAPT and demonstrated positive changes on other outcome measures. No participant deteriorated across all outcome measures or crossed the cutoff C in the wrong direction. The fourth participant lacked data from post-treatment but deteriorated on all outcome measures except anger expression at follow-up. Given the small sample size, no firm conclusions can be drawn from this, but the findings highlight the importance of matching the treatment to offenders' individual risks and needs. An important area that this study could not address is how to identify offenders who potentially should not take part in VRAPT and to learn more about the potential hazards and unbeneficial aspects of VR-assisted treatment for offenders. While no in-depth analyses of this is possible with our study data, it is notable that the participants with partial deterioration all scored in the moderate to severe category for emotional neglect and in the severe category for physical neglect according to the CTQ-SF. This could mean that offenders with a history of childhood trauma including physical or emotional neglect might need to be treated differently—or not at all—in VRAPT. Clearly, this needs to be further investigated in future studies as also recommended by Klein Tuente et al. (37).

To further translate the findings from our outcome measures into clinical meaning, they should be placed in a larger context. We found that the sample mean on emotion regulation was below that of a normal population (48) at both post-treatment and follow-up. This was also true for the post-treatment measure of state anger compared with another normal population (50). STAXI in the study sample ranged from the 35 to the 45th percentile for state anger, between the 55 and 70th percentiles for trait anger, and the 55 and 75th percentiles for anger expression at post-treatment and follow-up, putting the sample within a normal population range (50). The study design does not permit conclusions about the effect of the intervention due to the lack of control group, and the findings in this study must be interpreted with much care due to the obvious limitations in the study design and small sample size, but the participants seemed to change on aggression, anger, and emotion dysregulation both during VRAPT and over time.

Several factors may impact the results of this study. For instance, half the sample had participated in treatment programs in the SPPS before entering VRAPT, which could have led to their greater ability to change during the time period for VRAPT due to skills acquired in previous treatments. No analyses of previous treatment in relation to VRAPT outcome were possible in this study, but this highlights a potential way to utilize VR-assisted offender treatment as a means of boosting or rehearsing specific skills and behaviors. Another aspect which must be considered relates to the generalizability of the current results. The current sample—14 violent offenders with high to medium risk of criminal recidivism—was like other offender samples in terms of complexity of needs and background factors commonly found in violent offender populations. Our sample consisted of young individuals lacking in post high school education and work

experience, who had a history of externalizing behaviors before age 15. Psychiatric problems and a history of adverse childhood experiences including prior convictions were prominent, as was a history of substance or alcohol abuse. Pro-criminal cognitions and externalizing behaviors over the life span were common. Aggression at baseline as measured by AQ-RSV (M = 94.4, SD = 23.4) was similar to other forensic populations (M = 95.5, SD = 20.4) (86) but differed from the general population (M = 77.8, SD = 16.5) (44). Emotion regulation as measured by DERS (M = 92.9, SD = 23.5 at baseline) was comparable to adults with emotional disorders (M = 89.3, SD = 22.6) (49) but higher than in a general population (M = 78, SD = 20.7) (48). State anger (M = 24, SD = 11.9), trait anger (M = 24.7, SD = 8.1), and anger expression (M = 53.9, SD = 17.9) as measured by STAXI-2-S showed similar levels for state anger (M = 21.3, SD = 1.2) but not trait anger (M = 19.2, SD = 0.84) as a forensic outpatient group (51) but higher than a general population (state anger M = 19.3, SD = 6.9; trait anger M = 18.4, SD = 5.4; anger expression M = 33.5, SD = 13.1) (50). In a general population age group of 20–29, scores between the 25 and the 75th percentiles can be seen as normal (50); the study sample, however, scored in the 80th percentile for state anger, in the 85th percentile for trait anger, and in the 95 to 97th percentiles for anger expression at pre-treatment. The study sample therefore differed from normal populations and demonstrated similarities to clinical samples on the outcome measures. This sample cannot, however, be claimed as representative of violent offenders in general.

Since VR technology is an innovative and novel tool in offender treatment, its impact on offenders is not yet clear. Ethical considerations are important when using persuasive technology such as VR with vulnerable persons (87), and the participants in this study constitute a vulnerable group considering their complex needs, trauma history, and residence under the power dynamics of a correctional facility. The ethical dilemma is how to take vulnerability into consideration, avoid harm, and resist using persuasive technology coercively or manipulatively (87) while still developing better treatment for offenders to reduce their risk of criminal recidivism. A recent systematic review (29), concluded that VR-assisted assessment and treatment does not seem to pose any harm when applied in forensic settings, although much is still unknown of its effects. Thus, utilizing VR within offender treatment is feasible, but we still need to be mindful of its potential impacts and consequences.

Another ethical dilemma involves the risk of not providing potentially beneficial treatment simply because of its unknown characteristics. Standing on the frontlines of treatment development is always foggy, and steps forward need to be taken one at a time before we pick up speed. This study is a small step forward in a small group in dire need of effective treatment. The ethical dilemmas were managed through informed consent and participants' ability to opt-out (or simply return to the real world by removing the VR goggles) at any time, which were meant to strengthen participant autonomy. All participants received VRAPT as voluntary treatment and were informed that dropping out would not impact their sentence planning or earned prison privileges in any way. The first VRAPT session allows the participant to try the VR environment without any specific purpose except to get acquainted with it. This slow start is important, not only to build trust in the technology, but also to learn more about participants' reactions. It might be better to let participants try out the VR environment before enrolling in treatment, so they can make a fully informed decision about participating. Another benefit of such an approach would be learning beforehand about participants' reactions, thoughts, and feelings about VR. This knowledge could be helpful in tailoring the experience better to participants having trouble experiencing immersion and presence in the virtual world and mitigating participants' fears, attitudes, or prejudice about using VR in treatment.

4.1. Strengths and limitations

This study is rare in that it targets imprisoned violent offenders with VR-assisted relapse-preventive treatment, a novel methodology in forensic contexts. The study provides important insights into target group impact and feasibility in an innovative field that has great potential to improve the effectiveness of offender treatment. Specifically, the study provides clinical data on individual change, important knowledge that could be helpful in designing future VR-assisted treatment protocols and studies. The main strengths of the study lie in its novelty and exploratory aspect.

Given the current study's design as a pilot study, the sample size was small and there was no control group, limiting possible analyses, and conclusions from analyses. When introducing a new treatment protocol based on a novel technology such as VR, pilot studies performed in smaller samples are recommended to limit the number of participants exposed to the intervention in recognition of the importance of testing with care (84). The current analyses were selected considering this, and the combination of both a nomothetic and an idiographic approach with descriptive and statistical analysis filled the explorative purpose. For instance, the small sample made it hard to answer the second research question with inferential statistics on the group level, but the descriptive statistical analysis and analysis of RC made it possible to discuss potential impact factors in line with the aim of the study. To help address the limitation concerning sample size, we report missing post-treatment and follow-up data for different participants. Although this method of linear mixed models can accommodate missing data (88), the fact remains that some analyses consisted of only 11 of 12 participants, which is unfortunate and must be considered when interpreting the results.

The hybrid Bayesian and frequentist approach might also be seen as a limitation in that a more consistent approach would have been preferable. However, from an explorative perspective, the combination of Bayesian linear mixed effects models that could handle the nomothetic approach of group data with a small sample, and the well-established and useful frequentist approach of RCI (89) for the idiographic approach allowed us to utilize the strengths of both Bayesian and frequentist statistics. As stated by Bayarri and Berger (90), statisticians should use both Bayesian and frequentist ideas, and there are several situations where a combination is highly useful.

Another limitation lies in deviations taken from the recommended VRAPT methodology. Only half of the VRAPT treatments were conducted at the recommended intensity, possibly influencing the study's reliability. Treatment length was, however, both shorter and longer than the recommendations. The prison context is riddled with obstacles for treatment in general due to logistical issues such as overcrowding, staff turnover, competing planned activities, transportation distances, difficulties finding treatment rooms, or client misconduct leading to ward lockdowns. This study was conducted during the years 2020–2022, with a pandemic paralyzing much of the

everyday work with both staff and client infections, staff shortages, and ward lockdowns due to infections. In light of both the usual prison constraints and those of the COVID-19 pandemic, impact on treatment length was unavoidable in this study. However, despite these limitations, given the robust RC for many participants in this study, it seems plausible that VRAPT can be adjusted to be implement at different treatment intensities. The strong emphasis on self-reported data to measure change is another limitation in the study. Misconduct could not be used as a change measure mainly due to large differences in time served between participants before VRAPT. Staff ratings of participant behavior which was used in the randomized controlled trial by Klein Tuente et al. (37) was abandoned early in this project due to the overcrowding strain that staff was under at the time, making staff ratings impossible. More precise data on misconduct and staff ratings of participant behavior would certainly have added value to the exploratory aim of the study, providing valuable information on target group impact.

During the study, we found a skew in the number of VRAPT treatments conducted by the individual VRAPT facilitators; some conducted only one treatment, while one conducted as many as 5. Even though all VRAPT facilitators received basic training and supervision in using VR and the VRAPT protocol, it is reasonable to assume that facilitators performing more VRAPT treatments would excel in using the VR technology. With proficiency in using the technology, it is probably easier to focus on alliance and treatment in general. On the other hand, all facilitators were selected because of their experience in offender treatment, CBT methodology, and working with complex violent offenders. Learning a VR-assisted offender treatment program would probably be a much larger challenge for an unexperienced facilitator. Whether treatment effectiveness differs according to the alliance-building and methodological competencies of individual VRAPT facilitators is beyond the scope of this study. It may, however, impact treatment outcomes and should therefore be investigated in further studies since it raises questions about whether VR technology itself is better than the hand wielding it.

Finally, a possible impacting factor that could not be thoroughly investigated in this study was the impact of presence in VR. Our measure of presence in VR, IPQ, unfortunately provided data that was hard to interpret, with large differences in averages between the different subscales, possibly due in part to both general and internal missing data. Also, the IPQ results did not match the feedback provided from the participants during the study, most of whom seemed to describe both proper immersion and presence. Future studies on VR-assisted treatment need to explore further ways to measure presence and immersion, since these concepts are at the core of the VR experience (30).

4.2. Future research

Virtual reality-assisted offender treatment is still in its infancy and VRAPT is only one of many potential VR-assisted offender treatment programs. Much is unknown about how, when, with whom, why, and even whether we should use this technology. Randomized controlled trials are needed to evaluate various treatment protocol's effect on the risk of criminal recidivism. Continued explorative studies on the impact of VR-assisted offender programs and whether and how VR-assisted treatment might trigger trauma responses in offenders and impact treatment responsivity are also needed. Other aims could include determining who might and might not benefit from VR-assisted treatment and whether VR treatment can be tailored to address responsivity issues related to the technology itself such as presence and immersion. Bridging the gap between the participant and the technology raises yet other research questions, and appropriate amounts of VR exposure should also be investigated to establish the most effective treatment intensity. Finally, research is needed on how VR can fit into existing treatment protocols and how clinicians' experiences using the technology can affect treatment response in participants.

4.3. Conclusion

This pilot study provides an early glimpse of how the VR-assisted aggression treatment VRAPT might impact the target group of incarcerated violent offenders. No conclusions on the effect of VRAPT can be made due to the study's lack of a comparison group, and results must be interpreted with caution due to its small sample size. However, the results indicated that for many participants in this study, a highly probable change in core criminogenic needs related to the risk of relapse in crime occurred during the time for enrollment in VRAPT, and that change was largely maintained over a 3-month period after treatment ended. Some participants did not seem to benefit much from VRAPT treatment, however, and a minority even deteriorated on the outcome measures. From both an ethical and relapse-prevention perspective, further investigations on identifying appropriate VRAPT target groups remain key avenues for future research. Although VRAPT should be considered an early adaptation in a field that will develop rapidly during the years to come, this pilot demonstrates that taking part in VRAPT may be associated with change on important outcomes for imprisoned violent offenders.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The Swedish ethical review authority. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DI participated in planning the study, collecting, and analyzing the data, and writing and revising the article. CD participated in the data analysis and in writing and revising the article. PE participated in revising the article, and MW planned the study and revised the article. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Swedish Prison and Probation Service, the Regional Forensic Psychiatric Clinic in Växjö, and the Swedish National Council for Crime Prevention (5.1-0348/22) and the Swedish Research Council for Health, Working Life and Welfare (2018–01409).

Acknowledgments

The study was made possible through the joint venture of the Swedish Prison and Probation Service and Regional Forensic Psychiatric Clinic in Växjö, Region Kronoberg. The study was preregistered at the International Standard Randomized Controlled Trial Number registry (ISRCTN14916410).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor JH is currently organizing a Research Topic with the author MW.

Publisher's note

References

1. Ståhlberg O, Anckarsäter H, Nilsson T. Mental health problems in youths committed to juvenile institutions: prevalences and treatment needs. *Eur Child Adolesc Psychiatry*. (2010) 19:893–903. doi: 10.1007/s00787-010-0137-1

2. Wallinius M, Nilsson T, Hofvander B, Anckarsäter H, Stålenheim G. Facets of psychopathy among mentally disordered offenders: clinical comorbidity patterns and prediction of violent and criminal behavior. *Psychiatry Res*. (2012) 198:279–84. doi: 10.1016/j.psychres.2012.01.005

3. Fazel S, Hayes AJ, Bartellas K, Clerici M, Trestman R. Mental health of prisoners: prevalence, adverse outcomes, and interventions. *Lancet Psychiatry*. (2016) 3:871–81. doi: 10.1016/S2215-0366(16)30142-0

4. Mikton CR, Butchart A, Dahlberg LL, Krug EG. Global status report on violence prevention 2014. *Am J Prev Med*. (2016) 50:652–9. doi: 10.1016/j.amepre.2015.10.007

5. Falk O, Wallinius M, Lundstrom S, Frisell T, Anckarsater H, Kerekes N. The 1% of the population accountable for 63% of all violent crime convictions. *Soc Psychiatry Psychiatr Epidemiol*. (2014) 49:559–71. doi: 10.1007/s00127-013-0783-y

6. Smeijers D, Benbouriche M, Garofalo C. The association between emotion, social information processing, and aggressive behavior: a systematic review. *Eur Psychol*. (2020) 25:81–91. doi: 10.1027/1016-9040/a000395

7. Kelty SF. You have to hit some people! Why problem-solving skills are more criminogenic than hostile attributional biases for adult male violent offenders. *Psychiatry Psychol Law*. (2013) 20:713–34. doi: 10.1080/13218719.2012.729022

8. Calvete E, Orue I. The role of emotion regulation in the predictive association between social information processing and aggressive behavior in adolescents. *Int J Behav Dev*. (2012) 36:338–47. doi: 10.1177/0165025412444079

9. Helmsen J, Koglin U, Petermann F. Emotion regulation and aggressive behavior in preschoolers: the mediating role of social information processing. *Child Psychiatry Hum Dev*. (2012) 43:87–101. doi: 10.1007/s10578-011-0252-3

10. Gagnon J, McDuff P, Daelman S, Fournier S. Is hostile attributional bias associated with negative urgency and impulsive behaviors? A social-cognitive conceptualization of impulsivity. *Personal Individ Differ*. (2015) 72:18–23. doi: 10.1016/j.paid.2014.08.011

11. Landenberger NA, Lipsey MW. The positive effects of cognitive–behavioral programs for offenders: a meta-analysis of factors associated with effective treatment. *J Exp Criminol*. (2005) 1:451–76. doi: 10.1007/s11292-005-3541-7

12. Mills JF, Kroner DG, Forth AE. Measures of criminal attitudes and associates (MCAA): development, factor structure, reliability, and validity. *Assessment*. (2002) 9:240–53. doi: 10.1177/1073191102009003003

13. Krueger RF, Markon KE, Patrick CJ, Benning SD, Kramer MD. Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. *J Abnorm Psychol*. (2007) 116:645–66. doi: 10.1037/0021-843X.116.4.645

14. Sindicich N, Mills KL, Barrett EL, Indig D, Sunjic S, Sannibale C, et al. Offenders as victims: post-traumatic stress disorder and substance use disorder among male prisoners. *J Forensic Psychiatry Psychol*. (2014) 25:44–60. doi: 10.1080/14789949.2013.877516

15. Bonta J, Andrews DA. *The Psychology of Criminal Conduct*. *6th* ed New York: Routledge (2017).

16. Schubert T, Friedmann F, Regenbrecht H. The experience of PRESENCE: factor analytic insights. *Presence Teleoperat Virtual Environ*. (2001) 10:266–81. doi: 10.1162/105474601300343603

17. McGuire J. A review of effective interventions for reducing aggression and violence. *Philos Trans R Soc B*. (2008) 363:2577–97. doi: 10.1098/rstb.2008.0035

18. McGuire J. "What works" to reduce re-offending: 18 years on In: Craig LA, Dixon L, Gannon TA, editors. *What Works in Offender Rehabilitation: An Evidence-Based Approach to Assessment and Treatment*: Wiley Blackwell (2013). 20–49.

19. Koehler JA, Lösel F, Akoensi TD, Humphreys DK. A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *J Exp Criminol*. (2013) 9:19–43. doi: 10.1007/s11292-012-9159-7

20. Brogan L, Haney-Caron E, NeMoyer A, DeMatteo D. Applying the risk-needs-responsivity (RNR) model to juvenile justice. *Crim Justice Rev*. (2015) 40:277–302. doi: 10.1177/0734016814567312

21. Lipsey MW, Landenberger NA, Wilson SJ. Effects of cognitive-behavioral programs for criminal offenders. *Campbell Syst Rev*. (2007) 3:1–21. doi: 10.1002/CL2.42

22. Papalia N, Spivak B, Daffern M, Ogloff JRP. A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clin Psychol Sci Pract*. (2019) 26:12282. doi: 10.1111/cpsp.12282

23. Jolliffe D, Farrington DP (2009). Effectiveness of interventions with adult male violent offenders: report prepared for the Swedish National Council for crime prevention. Brottsförebyggande rådet. Available at: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat07147a&AN=lub.1874799&site=eds-live&scope=site (Accessed Nov 25, 2021).

24. Kip H, Bouman YHA, Kelders SM, van Gemert-Pijnen LJEWC. eHealth in treatment of offenders in forensic mental health: a review of the current state. *Front Psychiatry*. (2018) 9:42. doi: 10.3389/fpsyt.2018.00042

25. Valmaggia LR, Latif L, Kempton MJ, Rus-Calafell M. Virtual reality in the psychological treatment for mental health problems: an systematic review of recent evidence. *Psychiatry Res*. (2016) 236:189–95. doi: 10.1016/j.psychres.2016.01.015

26. Terbeck S, Case C, Turner J, Spencer V, Bacon A, Howard C, et al. Assessing reactive violence using immersive virtual reality. *PLoS One*. (2022) 17:e0268191. doi: 10.1371/journal.pone.0268191

27. Ticknor B. Virtual reality and correctional rehabilitation: a game changer. *Crim Justice Behav*. (2019) 46:1319–36. doi: 10.1177/0093854819842588

28. Ticknor B, Tillinghast S. Virtual reality and the criminal justice system: new possibilities for research, training, and rehabilitation. *J Virtual Worlds Res*. (2011) 4:4–44. doi: 10.4101/jvwr.v4i2.2071

29. Sygel K, Wallinius M. Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. *Front Psychol*. (2021) 12:673089. doi: 10.3389/fpsyt.2021.673089

30. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos Trans R Soc Lond Ser B Biol Sci*. (2009) 364:3549–57. doi: 10.1098/rstb.2009.0138

31. Slater M. Immersion and the illusion of presence in virtual reality. *Br J Psychol*. (2018) 109:431–3. doi: 10.1111/bjop.12305

32. Smeijers D, Koole SL. Testing the effects of a virtual reality game for aggressive impulse management (VR-GAIME): study protocol. *Front Psychol*. (2019) 10:83. doi: 10.3389/fpsyt.2019.00083

33. Goldstein AP, Glick B, Reiner S, Zimmerman D, Coultry TM, Gold D. Aggression replacement training: a comprehensive intervention for the acting-out delinquent. *J Correct Educ*. (1986) 37:120–6.

34. Nee C, Ward T. Expertise and offending: themes for future research. *Aggress Violent Behav*. (2015) 20:92–3. doi: 10.1016/j.avb.2014.12.009

35. Nee C. Understanding expertise in burglars: from pre-conscious scanning to action and beyond. *Aggress Violent Behav*. (2015) 20:53–61. doi: 10.1016/j.avb.2014.12.006

36. Klein Tuente S, Bogaerts S, van IJzendoorn S. Effect of virtual reality aggression prevention training for forensic psychiatric patients (VRAPT): study protocol of a multi-center RCT. *BMC Psychiatry*. (2018) 18:251. doi: 10.1186/s12888-018-1830-8

37. Klein Tuente S, Bogaerts S, Bulten E, Keulen-de Vos M, Vos M, Bokern H, et al. Virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *JCMM*. (2020) 9:2258. doi: 10.3390/jcm9072258

38. Klein Tuente S, van JS, Vos M, Bogaerts S, Veling W. *VRAPT Manual*, vol. v2 (Eds.) (Edsl) Klein Tuente S, Wallinius M, Moraga FG and veling M (2017).

39. Klein Tuente S, Gonzales WM, Moraga F, Veling W. *VRAPT Manual*, vol. v3 (Edsl) Klein Tuente S, Wallinius M, Moraga FG and veling M (2020).

40. González Moraga FR, Klein Tuente S, Perrin S, Enebrink P, Sygel K, Veling W, et al. New developments in virtual reality-assisted treatment of aggression in forensic settings: the Case of VRAPT. *Front Virtual Real*. (2022) 2:675004. doi: 10.3389/frvir.2021.675004

41. Kriminalvården (2020). Manual för RBM-B. Risk-, Behov Och Mottaglighetsbedömning Version 1.1 Inom Ramen för Kriminalvårdens Verkställighetsplanering

42. Berman AH, Bergman H, Palmstierna T, Schlyter F. DUDIT – The Drug Use Disorders Identification Test, MANUAL Version 1.0. Karolinska institutet, Department of Clinical Neuroscience, Section for Alcohol and Drug Dependence Research (2003).

43. Bohn MJ, Babor TF, Kranzler HR. The Alcohol Use Disorders Identification Test (AUDIT): validation of a screening instrument for use in medical settings. *J Stud Alcohol.*. (1995) 56:423–432.

44. Buss AH, Perry M. The aggression questionnaire. *J Pers Soc Psychol*. (1992) 63:452–9. doi: 10.1037/0022-3514.63.3.452

45. Prochazka H, Ågren H. Aggression in the general Swedish population, measured with a new self-rating inventory: the aggression questionnaire—revised Swedish version (AQ-RSV). *Nord J Psychiatry*. (2001) 55:17–23. doi: 10.1080/080394801750093661

46. Gerevich J, Bácskai E, Czobor P. The generalizability of the Buss-Perry aggression questionnaire. *Int J Methods Psychiatr Res*. (2007) 16:124–36. doi: 10.1002/mpr.221

47. García-León A, Reyes GA, Vila J, Pérez N, Robles H, Ramos MM. The aggression questionnaire: a validation study in student samples. *Span J Psychol*. (2002) 5:45–53. doi: 10.1017/S1138741600005825

48. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J Psychopathol Behav Assess*. (2004) 26:41–54. doi: 10.1023/B:JOBA.0000007455.08539.94

49. Hallion LS, Steinman SA, Tolin DF, Diefenbach GJ. Psychometric properties of the difficulties in emotion regulation scale (DERS) and its short forms in adults with emotional disorders. *Front Psychol*. (2018) 9:539. doi: 10.3389/fpsyg.2018.00539

50. Spielberger C. *State-Trait Anger Expression Inventory-2 Professional Manual*. Odessa, Florida: Psychological Assessment Resources (1999).

51. Lievaart M, Franken IHA, Hovens JE. Anger assessment in clinical and nonclinical populations: further validation of the state-trait anger expression Inventory-2: further validation of the STAXI-2. *J Clin Psychol*. (2016) 72:263–78. doi: 10.1002/jclp.22253

52. Spielberger CD. *State-Trait Anger Expression Inventory Professional Manual*. Odessa, Florida: Psychological Assessment Resources (1988).

53. Eckhardt C, Norlander B, Deffenbacher J. The assessment of anger and hostility: a critical review. *Aggress Violent Behav*. (2004) 9:17–43. doi: 10.1016/S1359-1789(02)00116-7

54. Lindqvist JK, Dåderman AM, Hellström Å. Swedish adaptations of the novaco anger scale-1998, the provocation inventory, and the state-trait anger expression inventory-2. *Soc Behav Pers*. (2003) 31:773–88. doi: 10.2224/sbp.2003.31.8.773

55. Patrick CJ, Kramer MD, Krueger RF, Markon KE. Optimizing efficiency of psychopathology assessment through quantitative modeling: development of a brief form of the externalizing Spectrum inventory. *Psychol Assess*. (2013) 25:1332–48. doi: 10.1037/a0034864

56. Soe-Agnie SE, Paap MCS, Nijman HLI, De Jong CAJ. Psychometric properties of the externalizing Spectrum inventory: replication and extension across clinical and non-clinical samples. *J Pers Assess*. (2021) 103:332–41. doi: 10.1080/00223891.2020.1753752

57. Berlin J, Wallinius M, Nilsson T, Karlén MH, Delfin C. Exploring the psychometric properties of the externalizing spectrum inventory-brief form in a Swedish forensic psychiatric inpatient sample. *BMC Psychiatry*. (2023) 23:184. doi: 10.1186/s12888-023-04609-y

58. Bernstein DP, Fink L. *Childhood Trauma Questionnaire: A Retrospective Self-Report Manual*. San Antonio, Texas: The Psychological Corporation (1998).

59. Bernstein DP, Stein JA, Newcomb MD, Walker E, Pogge D, Ahluvalia T, et al. Development and validation of a brief screening version of the childhood trauma questionnaire. *Child Abuse Negl*. (2003) 27:169–90. doi: 10.1016/S0145-2134(02)00541-0

60. Georgieva S, Tomas JM, Navarro-Pérez JJ. Systematic review and critical appraisal of childhood trauma questionnaire—short form (CTQ-SF). *Child Abuse Negl*. (2021) 120:105223. doi: 10.1016/j.chiabu.2021.105223

61. Saini SM, Hoffmann CR, Pantelis C, Everall IP, Bousman CA. Systematic review and critical appraisal of child abuse measurement instruments. *Psychiatry Res*. (2019) 272:106–13. doi: 10.1016/j.psychres.2018.12.068

62. American Psychiatric Association (2019). DSM-5 self-rated level 1 cross-cutting symptom measure American Psychiatric Association [WWW document]. Available at: https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures (Accessed Feb 1, 2020).

63. Clarke DE, Narrow WE, Regier DA, Kuramoto SJ, Kupfer DJ, Kuhl EA, et al. DSM-5 Field trials in the United States and Canada, part I: study design, sampling strategy, implementation, and analytic approaches. *AJP*. (2013) 170:43–58. doi: 10.1176/appi.ajp.2012.12070998

64. Narrow WE, Clarke DE, Kuramoto SJ, Kraemer HC, Kupfer DJ, Greiner L, et al. DSM-5 Field trials in the United States and Canada, part III: development and reliability testing of a cross-cutting symptom assessment for DSM-5. *AJP*. (2013) 170:71–82. doi: 10.1176/appi.ajp.2012.12071000

65. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, et al. DSM-5 Field trials in the United States and Canada, part II: test-retest reliability of selected categorical diagnoses. *AJP*. (2013) 170:59–70. doi: 10.1176/appi.ajp.2012.12070999

66. Mahoney MR, Farmer C, Sinclair S, Sung S, Dehaut K, Chung JY. Utilization of the DSM-5 self-rated level 1 cross-cutting symptom measure-adult to screen healthy volunteers for research studies. *Psychiatry Res*. (2020) 286:112822. doi: 10.1016/j.psychres.2020.112822

67. Bäckström M, Björklund F. The measures of criminal attitudes and associates (MCAA): further testing of structural and criterion-related validity. *Crim Justice Behav*. (2008) 35:1398–410. doi: 10.1177/0093854808322239

68. Mills JF, Kroner DG, Hemmati T. The measures of criminal attitudes and associates (MCAA): the prediction of general and violent recidivism. *Crim Justice Behav*. (2004) 31:717–33. doi: 10.1177/0093854804268755

69. Chicago Panahi SM, Fathi AA, Azad FP, Montazer GA. Reliability and validity of Igroup presence questionnaire (IPQ). *J Behav Sci*. (2009) 3:27–34.

70. Vasconcelos-Raposo J, Bessa M, Melo M, Barbosa L, Rodrigues R, Teixeira CM, et al. Adaptation and validation of the Igroup presence questionnaire (IPQ) in a Portuguese sample. *Presence Teleop Virt*. (2016) 25:191–203. doi: 10.1162/PRES_a_00261

71. Bürkner PC. Brms: an *R* package for Bayesian multilevel models using *Stan*. *J Stat Softw*. (2017) 80:1–28. doi: 10.18637/jss.v080.i01

72. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. *J Stat Softw*. (2017) 76:1–32. doi: 10.18637/jss.v076.i01

73. Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using the t distribution. *J Am Stat Assoc*. (1989) 84:881–96. doi: 10.2307/2290063

74. Gelman A, Simpson D, Betancourt M. The prior can often only be understood in the context of the likelihood. *Entropy*. (2017) 19:555. doi: 10.3390/e19100555

75. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. (1992) 7:457–72. doi: 10.1214/ss/1177011136

76. Mastrandrea MD, Mach KJ, Plattner GK, Edenhofer O, Stocker TF, Field CB, et al. The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Clim Chang*. (2011) 108:675–91. doi: 10.1007/s10584-011-0178-6

77. Makowski D, Ben-Shachar MS, Chen SHA, Lüdecke D. Indices of effect existence and significance in the Bayesian framework. *Front Psychol*. (2019) 10:2767. doi: 10.3389/fpsyg.2019.02767

78. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. (2017) 27:1413–32. doi: 10.1007/s11222-016-9696-4

79. Gelman A, Hill J, Vehtari A (2020). *Regression and Other Stories*. Cambridge University Press.

80. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*. (1991) 59:12–9. doi: 10.1037/0022-006X.59.1.12

81. Berman AH, Bergman H, Palmstierna T, Schlyter F (2003). DUDIT—the drug use disorders identification test, MANUAL version 1.0. Karolinska institutet, Department of Clinical Neuroscience, Section for Alcohol and Drug Dependence Research.

82. Berman AH, Wennberg P, Källmen H. *AUDIT & DUDIT*. Gothia förlag: Identifiera problem med alkohol och droger (2012).

83. Bohn MJ, Babor TF, Kranzler HR. The alcohol use disorders identification test (AUDIT): validation of a screening instrument for use in medical settings. *J Stud Alcohol*. (1995) 56:423–32. doi: 10.15288/jsa.1995.56.423

84. Marloth M, Chandler J, Vogeley K. Psychiatric interventions in virtual reality: why we need an ethical framework. *Camb Q Healthc Ethics*. (2020) 29:574–84. doi: 10.1017/S0963180120000328

85. Andrews DA, Dowden C. Risk principle of Case classification in correctional treatment: a meta-analytic investigation. *Int J Offender Ther Comp Criminol*. (2006) 50:88–100. doi: 10.1177/0306624X05282556

86. Berlin J, Tärnhäll A, Hofvander B, Wallinius M. Self-report versus clinician-ratings in the assessment of aggression in violent offenders. *Crim Behav Ment Health*. (2021) 31:198–210. doi: 10.1002/cbm.2201

87. Jacobs N. Two ethical concerns about the use of persuasive technology for vulnerable people. *Bioethics*. (2020) 34:519–26. doi: 10.1111/bioe.12683

88. Magezi DA. Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Front Psychol*. (2015) 6:2. doi: 10.3389/fpsyg.2015.00002

89. Blampied NM. Reliable change and the reliable change index: still useful after all these years? *tCBT*. (2022) 15:e50. doi: 10.1017/S1754470X22000484

90. Bayarri MJ, Berger JO. The interplay of Bayesian and frequentist analysis. *Stat Sci*. (2004) 19:58–80. doi: 10.1214/088342304000000116.full

frontiers | Frontiers in Psychology

TYPE Clinical Trial
PUBLISHED 14 November 2023
DOI 10.3389/fpsyg.2023.1235808

OPEN ACCESS

EDITED BY
Josanne van Dongen,
Erasmus University Rotterdam, Netherlands

REVIEWED BY
Carmen-Silva Sergiou,
Leiden University, Netherlands
Alexander Hochdorn,
University of Brasilia, Brazil

*CORRESPONDENCE
Chris N. W. Geraets
✉ chris.geraets@med.lu.se

RECEIVED 07 June 2023
ACCEPTED 25 October 2023
PUBLISHED 14 November 2023

CITATION
Woicik K, Geraets CNW, Klein Tuente S, Masthoff E and Veling W (2023) Virtual reality aggression prevention treatment in a Dutch prison-based population: a pilot study. *Front. Psychol.* 14:1235808. doi: 10.3389/fpsyg.2023.1235808

Virtual reality aggression prevention treatment in a Dutch prison-based population: a pilot study

Kasja Woicik[1,2], Chris N. W. Geraets[2,3]*, Stéphanie Klein Tuente[2], Erik Masthoff[4,5] and Wim Veling[2]

[1]Penitentiary Institution Vught, Vught, Netherlands, [2]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, Netherlands, [3]Lund Clinical Research on Externalizing and Developmental Psychopathology, Child and Adolescent Psychiatry, Department of Clinical Sciences Lund, Lund University, Lund, Sweden, [4]Fivoor Science and Treatment Innovation, Rotterdam, Netherlands, [5]Department of Developmental Psychology, Tilburg University, Tilburg, Netherlands

Background: Treating violent behavior in prisons comes with challenges, such as the inability to practice safely with triggering situations and motivational issues. A solution may be the use of Virtual Reality (VR). With VR, specific conditions or needs can be tailored for individual practice, it can enhance motivation and VR has proven to be a safe and effective tool in mental health treatment.

Objective: A pilot study was conducted to test the acceptability, feasibility, and preliminary effects of VR Aggression Prevention Treatment (VRAPT) in a prison-based population.

Methods: In total 17 detainees with aggressive behavior were included in this single-group pilot study. Acceptability and feasibility were assessed using qualitative measures for participants and therapists. Preliminary treatment effects were measured with self-report and observational measures on aggression, anger, emotion regulation, and impulsiveness.

Results: Participants and therapists were predominantly positive about VRAPT. Participants rated the sessions with an average satisfaction score of 9.2 out of 10 (SD = 0.3). Qualitative data showed that participants reported having learned to respond more adequately to aggressive behavior and gained insights into their own and others' triggers and tension. The combination of VR and theory was experienced as a strength of the treatment, as well as the ability to trigger aggression in VR which provided insights into aggression. However, the theoretical framework was found to be too complex, and more aggressive and personal scenarios should be incorporated into the sessions. Self-reported aggression, anger, provocation, emotion regulation, and observed verbal aggression decreased and seemed to stabilize after the treatment ended, with small to medium effect sizes.

Conclusion: VRAPT proved feasible and acceptable for most participants and therapists. An adapted treatment protocol called Virtual Reality Treatment for Aggression Control (VR-TrAC), will be used in a future RCT to investigate the effects of the treatment in a prison-based population.

KEYWORDS

aggression, virtual reality, prison, behavioral training, forensic

1 Introduction

Offenders of violent crimes involving force or causing injury (such as assault, homicide or armed robbery) recidivate more often than non-violent offenders (Cuervo et al., 2018; Recidivism Among Federal Violent Offenders, 2019). Furthermore, in prison violent offenders show high rates of aggression and violence, causing harm to staff, fellow detainees, and their living environment (Mcguire, 2018; Mcneeley, 2020; Wildra, 2020). Aggressive behavior during imprisonment has also been related to higher rates of crime recidivism (Mooney and Daffern, 2015). Given these effects of aggression on both society and the prison environment, treating individuals with aggression problems during detention is of critical need. Whereas social interventions such as supporting work and education have been found to be related to reduced recidivism, also psychological treatment forms are an important strategy for decreasing recidivism and violent behavior (Lipsey and Cullen, 2007; Walk et al., 2021).

Research has shown that, in general, psychological aggression treatment can be successful in reducing such behavior with small to moderate effect sizes (McGuire, 2008). However, findings for prison-based populations are still inconclusive. A recent meta-analytic review concluded that more high-quality research is needed to understand the specific factors contributing to effective treatment (Papalia et al., 2019). Research on prisoners has shown that, in general, psychological therapies that combine more than one method, seem most effective (Auty et al., 2017; Papalia et al., 2019). Such treatments mostly consist of multimodal cognitive behavioral methods that focus on role-play, relapse prevention, reshaping cognitions, improving problem-solving, exposure, and/or training skills (Papalia et al., 2019).

One of the best-known and studied frameworks for treating prison-based populations is the Risk-Need -Responsivity model (RNR). This model states that interventions should be personalized based on the risk of recidivism (Risk), adjusted to the factors that predict criminogenic behavior (behavior directly related to recidivism; Need) and should fit the motivation and abilities of the offender (Responsivity; Polaschek, 2012). A common limitation concerning this "Need" principle in treating violent behavior in prison-based populations is the inability to practice safely with challenging and triggering real-life situations. Furthermore, motivational issues are common in such populations, for example, because individuals may lack problem awareness or may have followed therapy previously without success, demotivating them to follow therapy again (Jochems et al., 2012; Smeijers et al., 2018). Also, interventions often have a cognitive and theoretical approach, which may not be the best fit for prison-based populations in which intellectual abilities below average are common, limiting responsivity (Muñoz García-Largo et al., 2020).

A solution to the above-encountered problems may be the use of Virtual Reality (VR; Kip et al., 2018, 2019). VR uses computer-generated, interactive environments to imitate real-world situations. VR makes it possible to practice situations in a virtual environment and therefore individuals can practice aggression-inducing interactions safely. Virtual situations can also be tailored to fit the specific conditions or needs of an individual to practice (Freeman et al., 2017; Kip et al., 2019). Furthermore, adding VR technology as a tool in treatment may enhance motivation, as it is new, interesting, and interactive technology. In general, VR has been proven to be a safe and effective tool in the treatment of several disorders, such as anxiety and psychotic disorders (Freeman et al., 2017; Geraets et al., 2021). However, as a recent review on VR in forensic settings discussed, studies using VR in the treatment of behavior (such as aggression) show promising results but the number of studies is still very limited and further research is needed (Sygel and Wallinius, 2021).

A recent multicenter randomized controlled trial (RCT) by Klein Tuente et al. investigated the first VR aggression prevention training (VRAPT) in forensic inpatients (Klein Tuente et al., 2020). As a theoretic underpinning the social information processing theory (SIP) was used which is based on the social-cognitive theory (Dodge and Crick, 2007; Klein Tuente et al., 2018). To understand aggressive behavior, the SIP describes several steps in which an individual processes social and situational information, based on which a behavioral response is enacted (Crick and Dodge, 1994). Within the framework of the SIP model, aggressive behavior can result from aberrant or biased interpretation of situations, but also from aggression-inducing goal framing, a learning history impacted by trauma or aggressive role models, and limited resources to respond adequately (for more details of the SIP model see the methods section).

The participants were long-term forensic inpatients (average duration since index offense was 8 years) with a special judicial measure, called 'TBS-order' which is a measure for the court establishing a relationship between the committed crime and a psychiatric disorder. The forensic inpatients were positive about VRAPT and motivated to participate in the RCT, which was reflected by high inclusion rates (Klein Tuente et al., 2020). Furthermore, interviews revealed that participants were able to recall what they had learned (e.g., recognizing arousal and insights in triggers). Whereas no decrease in staff-rated or self-reported aggression was found in this RCT, self-reported hostility, anger and impulsiveness did improve after VRAPT compared to the waiting list condition. The lack of effects on aggression might be explained by the study population as it concerned participants with severe and long-lasting mental illness, who had been treated for many years. Given the long-term treatment history and persistent psychiatric problems, the included population may have had limited abilities to change.

In the current study, we aimed to test the VRAPT protocol in a prison-based population. To our knowledge, this is the first study examining VR treatment for aggression problems in a prison-based setting. Although no significant changes in aggression were found in the first VRAPT study, the intervention fits well within the RNR framework for prison-based populations. Though similar levels of motivation are expected, higher responsivity is expected as prisoners with aggression regulation problems likely have no, shorter-term and/or less severe psychopathology than long-stay forensic inpatients. Therefore, with this pilot, we aimed to test the acceptability, applicability and feasibility of VRAPT and identify potential points of improvement in the treatment protocol, as a base for a future RCT. The secondary objective was to explore the preliminary effects of VRAPT on aggression, anger, emotion regulation and impulsive behavior directly after treatment and at two-month follow-up.

2 Methods

2.1 Participants

Participants were male detainees residing at the Penitentiary Institution Vught in the Netherlands. On the units where participants

were recruited, only male detainees were imprisoned. The study was announced through flyers and a video on the prison tv-channel, and detainees were made aware of the study when aggressive behavior was noticed in them by the staff. Detainees who wanted to participate applied themselves by asking the staff for contact with the researchers.

Inclusion criteria were aged 18 or older and aggression regulation problems within the last month as indicated with the Aggression Questionnaire (AQ), with a score > 70 (Buss and Perry, 1992). Exclusion criteria were an indication of an intellectual disability (measured with the Screener for intelligence and mild intellectual disability (SCIL) score < 15 which is indicative of an IQ below 70; van Esch et al., 2020), acute suicidality, current psychotic episode, occurrence of epileptic seizures within the past year, insufficient command and understanding of the Dutch language, and estimated remaining imprisonment shorter than 5 months.

We aimed to include approximately 15 participants. The final sample consisted of 17 participants of whom the average age was 32 years (SD 8.4). Participants were of different ethnicity, foremost being Dutch (59%), but also Moroccan (12%), Colombian (6%), Antillean (6%), Congolese (6%), Belgian (6%), and Surinamese (6%). Participants had different educational levels: five participants (31%) had none or a lower education level, four (25%) had a vocational educational level, five (31%) had a secondary vocational level and two (12%) had a higher educational level. The majority of participants were single and had children (44%), 38% were single and did not have children, 17% had children and lived with a partner. Further characteristics of the sample are shown in Table 1.

TABLE 1 Baseline characteristics of the sample ($N = 17$).

	M(SD) or *N* (%)	
	Completers $N = 10$	Drop-out $N = 7$
Age	32.2 (8.2)	32.4 (9.4)
Education level		
None or primary	4 (40%)	1 (14.3%)
Vocational	3 (30%)	1 (14.3%)
Secondary vocational	1 (10%)	4 (57.1%)
Higher	1 (10%)	1 (14.3%)
Convictions		
Manslaughter	3 (30%)	1 (14.3%)
Property crimes (with violence)	2 (20%)	4 (57.1%)
(Heavy) violent crimes	2 (20%)	1 (14.3%)
Homicide	2 (20%)	0
Property crime (without violence)	1 (10%)	1 (14.3%)
Arson	1 (10%)	0
Destruction (property)	1 (10%)	0
Traffic violation	1 (10%)	0
Adverse childhood experiences		
Emotional abuse	1 (10%)	4 (57.1%)
Physical abuse	2 (20%))	4 (57.1%)
Sexual abuse	3 (30%)	1 (14.3%)
Emotional neglect	2 (20%)	4 (57.1%)
Physical neglect	0	2 (28.6%)
Parental separation or divorce	4 (40%)	5 (71.4%)
Mother treated violently	1 (10%)	2 (28.6%)
Household substance abuse	0	4 (57.1%)
Mental illness in household	0	2 (28.6%)
Criminal household member	3 (30%)	3 (42.9%)
Substance abuse		
Alcohol	3 (30%)	1 (16.7%)
Tobacco	2 (20%)	1 (16.7%)
Cannabis	4 (40%)	2 (33.3%)
Cocaine	0	1 (16.7%)

2.2 Design and procedure

This was an uncontrolled pilot intervention study with three measurement moments. This study was approved by the Medical Ethical Committee of the University Medical Center Groningen (METC number: 2019/381). Participants were not compensated for participating.

When a detainee was interested in the study, a researcher visited the detainee and provided verbal and written information on the study. If the detainee was willing to participate, informed consent was signed. For the screening, the SCIL score was checked from their file and the AQ was administered (van Esch et al., 2020).

After informed consent was obtained, observations by the staff started and continued until 4 weeks after the last VRAPT session had taken place. Four weeks after the start of the observations, the baseline assessment was performed. Then the treatment took place and the post-treatment assessment was performed after the final session. In case of treatment dropout, the post-treatment assessment was performed 2 months after the baseline assessment. The follow-up assessment was performed 2 months after post-treatment. All participants received care as usual, when necessary.

2.3 VR system

Participants were exposed to simulated virtual environments by wearing an Oculus Rift S headset and noise-canceling headphones, see Figure 1. Therapists operated the VR surroundings with a tablet, and on a second screen the therapist saw what the participant viewed. The 'Social Worlds' software, created with Unity by CleVR BV was used in this study, which was also used in the first VRAPT RCT (Klein Tuente et al., 2020). The following three modules of the software were used: (1) the emotion recognition task, (2) the aggression catwalk, and (3) the interactive scenarios. See Figure 2 for screenshots of the software.

During the emotion recognition task, participants navigated the virtual street by changing their body orientation and operating a joystick enabling forward and backward movement. Avatars were

FIGURE 1
The VR set-up. Reproduced with permission from Sander Martens.

FIGURE 2
Impression of the **(A)** VR emotion recognition task, **(B)** aggression catwalk, and **(C)** an interactive scenario. Images of the VR environment are reproduced with permission from CleVR BV.

standing at random locations in the VR street. When a participant moved within a two-meter radius, the avatar oriented towards the participant and displayed an emotion (anger, disgust, fear, happiness, sadness, surprise, or neutral), and the correct emotion had to be chosen from a pop-up menu with four options with the joystick. Six pre-installed levels were available to enable customizing.

In the aggression catwalk, participants were approached by avatars showing neutral to aggressive behavior. This was shown through facial emotions, body language and verbal expressions. Participants rated the level of aggression (from level 1 not aggressive to level 4 very aggressive).

During interactive scenarios, therapists wore headphones with a microphone and voice morphing. The virtual environments (e.g., a store, bar or prison) could be adapted to the specific needs of the participants, for example by choosing specific avatars (e.g., a security guard, a group of females or males with different ethnic backgrounds) to be present in the VR environment, as well as the number and type of avatars in the background. Furthermore, the therapist controlled the emotions, gestures and speech of the avatar(s) with whom the participant was interacting.

2.4 Intervention

The treatment consisted of 16 twice-weekly individual sessions with a maximum duration of 60 min per session. In practice, participants complete VRAPT on average in 13 weeks (range 8–26 weeks), due to practical reasons such as sickness of the participant or therapist, COVID restrictions, no-show of the participants or vacation of the therapist. Sessions were planned in

consultation with the participant. Four qualified psychologists received a one-day training and monthly group supervision.

The treatment protocol used the SIP model as a theoretical framework, which describes how problems with social information processing are linked to aggressive behavior (Crick and Dodge, 1994). It describes six steps in which an individual processes social and situational information leading to behavioral responses. The early stages involve the identification of (1) what is happening and (2) what it means to me. The late stages match the outcomes of (1) and (2) to (3) what goals am I trying to achieve, (4) what options do I have to react, (5) what am I going to do, eventually culminating in (6) the reaction or behavior. The steps are interrelated and can influence each other. During the treatment, each step of the SIP model is discussed, and related exercises are performed to improve social information processing and practice new behavior.

The treatment consisted of two parts. Part one focused on the early stages of social information processing, related to emotion recognition (sessions 2–6). Part two focused on the late information processing stages with interactive scenarios (sessions 7–15).

Each session started with a short recap of the previous session. During this recap, participants were asked if they discussed the theory with their mentor (every prisoner has an individual mentor) if they experienced any situations relevant to VRAPT, and if they applied any learned skills. Next, the theory and goal of the current session were explained and VR exercises were performed. The session ended with discussing how the participant could apply the learned theory and skills in the upcoming week. Participants received a workbook that contained exercises and a summary of the theory discussed in each session. From session 6 onwards, physiological measures were performed during the session. Below an overview of the sessions is given.

Session 1: included a general introduction and a simplified version of the SIP model was explained. At the end of the session, participants got acquainted with the VR system by exploring the VR street by utilizing the joystick, for maximally 10 min.

Session 2: SMART treatment goals were formulated and expectations were discussed. The first two steps of the SIP model and the relation between emotion recognition and emotions were discussed shortly as it is known that individuals with aggression problems are more likely to interpret the behavior of others as aggressive (especially when ambiguous or unpredictable; Coccaro et al., 2017). Then participants practiced with three levels of the emotion recognition task. This was customized based on the performance of the participant on the task.

Sessions 3: the information of session 2 was summarized. Then participants practiced the emotion recognition task again. The difficulty level could be adjusted to the skills of the participant.

Session 4: focused on recognizing different levels of aggression. After discussing this topic, participants practiced with the aggression catwalk in VR. Participants rated approximately 10 to 15 avatars on their level of aggression. Scores were evaluated and the participant repeated the task.

Session 5: learning goals were evaluated and the theory from sessions 1 to 4 was repeated. The therapist discussed with the participant where repetition was needed, and exercises from the previous session were repeated (i.e., the emotion recognition task or aggression catwalk).

Sessions 6: information on steps 3 to 6 of the SIP model was repeated (they were explained in session 1) and the concept of physical arousal and physiological measurements were introduced (heart rate measured with an electrocardiogram (ECG), and galvanic skin response (GSR) measured with a finger sensor). The goal was to use these measures to teach participants to recognize physical signs of arousal. Real-time graphs of the physiological measures, the VR, and the therapist's and participant's voices were recorded to enable rewatching and discussing physiological responses with the participant. While participants wore the sensors, they were approached aggressively by avatars during the aggression catwalk. Participants were asked not to react to the avatars but to pay attention to signals in their body. This was discussed after the assignment. Next, participants were approached again by aggressive avatars on the aggression catwalk, but now they had to react like they normally would. The experience was discussed.

Session 7: the goal was to learn that there are different responses to aggression-provoking situations. This was practiced with two pre-scripted interactive scenarios. Scenario 1 took place in the bar, where the participant is accused of louring at the girlfriend of an avatar. The participant was asked to react like he would normally do. After the scenario, different types of reactions were discussed (sub-assertive, assertive, and aggressive), and the participant's reaction was classified. The assertive way was discussed in more detail (telling your message in an I-formulated message). Next, three interactive scenarios were played in which the participants spilled coffee on the shoes of the avatar in the bar. The therapist (enacting the avatar) demonstrated the different types of reactions in each scenario to give more insight into the different responses and what they provoke.

Session 8: different responses were practiced in three interactive scenarios in the supermarket. The participant wanted to enter the supermarket, but the security denied his entrance due to closing time. However, during the scenario the participant saw how another avatar entered the supermarket without the security noticing it. In the first scenario, it was asked to react in a sub-assertive way, in the second in an assertive way and the third in an aggressive way. The different ways of reacting were then discussed and evaluated.

Session 9: different ways and tools were discussed to help the participant react more assertively, including strategies such as ignoring, helping thoughts, counting to 10, focusing on breathing, or a time-out. In this session, three prescripted scenarios were played. In scenario 1 he was accused by a police officer of something he did not do, in scenario 2 the participant was not allowed to call his lawyer and in scenario 3 he had an argument with friends. Before each scenario, it was discussed which strategy the participant wanted to use. The scenarios were discussed and evaluated afterward.

Session 10: three different pre-scripted scenarios were played to practice new skills and reduce tension. In the first the participant was accused of stealing something, in the second scenario the participant wanted to order a drink but the bartender refuses and in the third scenario, the participant was accused of using drugs in prison. The participant practiced with the strategies that were discussed in session 9.

Session 11: goals were evaluated and a brief recap of the different kinds of responses and strategies was given. One or two interactive scenarios were practiced. The scenario could be chosen from a list of pre-scripted scenarios, or a personalized scenario could be practiced

(with a personalized environment, number and type of avatars as well as general content of the interaction).

Session 12–15: Three personalized or pre-scripted scenarios per session were performed. Each scenario was discussed and evaluated.

Session 16: the treatment was evaluated, and when necessary, previous topics were repeated or trained with an interactive scenario.

2.5 Measures

2.5.1 Sociodemographic and clinical characteristics

Sociodemographic characteristics were collected on age, cultural background, education, family status, and conviction history. Lifetime substance dependence and abuse were measured with the *Measurement in the Addiction for Triage & Evaluation* (*MATE*; Schippers et al., 2011). The MATE has good psychometric standards, with satisfactory inter-rater reliability (ICC range 0.75–0.92). Concurrent validity is good, with correlations above 0.50 (Schippers et al., 2010). Childhood trauma was measured with the *Adverse Childhood Experiences* (*ACE*; van der Feltz-Cornelis et al., 2019). Construct reliability is acceptable (ω = 0 0.91; Mei et al., 2022).

2.5.2 Self-report measures

At baseline, post-treatment, and follow-up the following questionnaires were administered.

Aggression was measured with the *Aggression Questionnaire (AQ)* which was the primary outcome measure (Buss and Perry, 1992). The AQ consists of 29 items measuring aggression on four different scales: physical aggression, verbal aggression, anger, and hostility. Items are rated on a 5-point Likert scale (ranging from 'disagree a lot' to 'agree a lot'). Test–retest reliability of the AQ is good (0.72), as well as the internal consistency (Cronbach's α = 0.83). This also applies to the validity of the total score of the AQ (the AQ correlated positively with alternative questionnaires measuring aggression; Hornsveld et al., 2009).

Anger was assessed with the *Novaco Anger Scale and Provocation Inventory* (*NAS-PI*; Hornsveld et al., 2011), which consists of two parts. The NAS part contains 48 questions and measures three factors of anger: cognitive, arousal, and behavior. Items are measured on a 3-point Likert scale (ranging from 'never true' to 'always true'). The PI part contains 25 items assessing provocation in response to anger-eliciting situations rated on a 4-point Likert scale (ranging from 'not angry at all' to 'very angry'). The internal consistency of the NAS and PI is excellent (Cronbach's α NAS = 0.92, and Cronbach's α PI = 0.90, the test–retest reliability of the NAS is good (r = 0.80) and the validity of the NAS and PI is good (the NAS-PI correlated positively with alternative questionnaires measuring anger and personality; Hornsveld et al., 2011).

Reactive and proactive aggression was measured with the *Reactive-Proactive Questionnaire* (*RPQ*; Cima et al., 2013). The RPQ consists of 23 items rated on a 3-point Likert scale (ranging from 'never' to 'a lot'); 11 items on reactive aggression and 12 items on proactive aggression. The RPQ has excellent internal consistency (Cronbach's α = 0.91), test–retest reliability is good, (all ICCs>0.41 at 3-year follow-up) and the convergent validity is adequate (with significant positive correlations with several other aggression measures; Cima et al., 2013).

Emotion regulation was assessed with the *Difficulties in emotion regulation* (*DERS*; Gratz and Roemer, 2004). The DERS consists of 36 items measured on a 5-point Likert scale (ranging from 'almost never' to 'almost always'). The DERS has high internal consistency (Cronbach's α = 0.93 for the total score), good test–retest reliability (r = 0.88; Gratz and Roemer, 2004).

Impulsive behavior was measured with the *Baratt Impulsiveness Scale* (*BIS-11*; Stanford et al., 2009). The BIS-11 has 30 items, assessing different personality and behavioral constructs of impulsiveness, rated on a 4-point Likert scale (ranging from 'rarely/never' to 'almost always'). The BIS-11 showed good internal consistency (Cronbach's α = 0.83) and test–retest reliability (Spearman's Rho = 0.83; Stanford et al., 2009).

2.5.3 Staff-rated measure

Aggressive behavior was scored by prison staff with the *Social Dysfunction and Aggression Scale* (*SDAS-9*; Wistedt et al., 1990; Kobes et al., 2012). The SDAS-9 is a behavior-observatory questionnaire consisting of 9 items measuring the extent of outward physical and verbally aggressive behavior in the past week. It is rated on a 5-point Likert scale, ranging from 'not present' to 'very serious'. Internal consistency (Cronbach's α = 0.82) and convergent validity are good (r = 0.73 with the staff observation aggression scale revised and interobserver reliability is moderate (ICC = 0.50; Kobes et al., 2012).

2.5.4 VR session measures

At the end of each session, participants completed the *Session Rating Scale (SRS)* on session satisfaction (Duncan et al., 2003) which includes the therapeutic alliance. The SRS consists of four items measuring the relationship (from "I did not feel heard, understood, and respected" to "I did feel heard, understood, and respected"), goals and topics (from "We did not work on or talk about what I wanted to work on and talk about" to "We did work on or talk about what I wanted to work on and talk about"), approach or method (from "The therapist's approach is not a good fit for me" to "The therapist's approach is a good fit for me") and the fourth item requires the participant to generally evaluate the session. Items are scored on a 10-centimeter visual analog scale, ranging from 0 to 10. The total score was analyzed. The SRS shows moderate to high internal consistency (ranging from 0.70 to 0.97) and low to moderate validity (ranging from 0.29 to 0.48; Murphy et al., 2020).

The sense of presence experience in VR was measured with *The Igroup Presence Questionnaire (IPQ)* and was conducted at the end of treatment (Schubert et al., 2001). The IPQ is rated on a 7-point Likert scale, ranging from negative statements about the VR world (such as 'the VR world felt like an imaginary world) to positive statements about the VR world (such as 'it could not be differentiated from the real world'). Internal consistency reliability is good (Cronbach's α = 0.85; Igroup, n.d.).

2.5.5 Qualitative measurements

Therapists and participants completed open questions about the treatment after every session. The questions concerned how they generally experienced the session, how well the session content fitted the participant's aggression regulation problem, critical feedback about the session, if there were any difficulties, how they experienced the use of the VR equipment, the exercises and roleplays, and the usage of the treatment protocol. During the first session, additionally

participants were specifically asked whether they experienced cybersickness, in the following sessions this was only noted when reported by the participant.

2.6 Analyses

Qualitative data consisted of answers to open questions that were first grouped into topics by the first author. Then, similar answers and categories were grouped in an iterative process between the first and second authors. After sessions 2 to 15, the therapists rated how well the session content fitted the participant's aggression problem. Answers were coded into a good fit, reasonable fit, or did not fit.

For quantitative data, descriptive statistics were calculated, i.e., means and standard deviations or count and percentages. Total scores were calculated if maximally 2 items were missing, for subscales maximally 1 item was allowed to be missing. Missing items were replaced by the scale mean item score. Effect sizes were calculated based on the mean and standard deviation for parametric data and the 'Hedges' g correction was used because of the small sample size ($n < 20$; 0.20–0.49 is a small effect, 0.50–0.79 medium effect and 0.8 a large effect). For non-parametric data effect sizes were based on the Wilcoxon signed rank test for which effect sizes 0.10–0.39 are considered a small effect, 0.30–0.49 a medium effect and 0.50 or higher a large effect.

3 Results

3.1 Feasibility and acceptability

The study was completed between November 2019 and May 2021. In total, 32 detainees self-referred to the study. After the screening, 17 participants were included, see Figure 3. The baseline characteristics of the sample are shown in Table 1. Reasons for not participating after self-referral were: AQ score below 70 ($n = 3$), leaving detention before the start of the study ($n = 3$), transferal to another prison ($n = 2$), not motivated ($n = 1$), too busy ($n = 1$), not feeling well ($n = 1$), scared of potential side effects ($n = 1$), unwilling for physiological measurements to be recorded ($n = 1$), a SCIL score below 15 ($n = 1$), and no reason given ($n = 1$).

Of the 17 participants, 10 completed all 16 sessions. Reasons for dropout were cybersickness ($n = 1$ stopped after session 2), stress factors on the unit ($n = 1$ stopped after 6 sessions), loss of motivation ($n = 2$ stopped after session 7 and 12 respectively), feeling that the VR exercises did not help ($n = 1$ stopped after session 8), no reason given ($n = 1$ stopped after session 11), and transferal ($n = 1$ stopped after session 12).

Post-treatment measures were completed by all participants who finished the treatment, and by three of the participants who dropped out. Three of the 10 participants who completed VRAPT did not complete the follow-up assessment because of a lack of motivation, leaving detention and no specific reason; only one of the treatment dropouts completed the follow-up measure.

Feelings of presence in VR were slightly above the theoretical average of 3.5 on all subscales and were interpreted as acceptable (range 1–7): spatial presence was scored on average 4.0 (SD = 0.6), general presence 4.6 (SD = 2.1), involvement 4.2 (SD = 0.9) and experienced realism 3.6 (SD = 1.1). After the first session, cybersickness symptoms were evaluated. Four participants reported that they experienced cybersickness, three of them described a feeling of dizziness, and one mentioned a feeling of 'car-sickness'.

After each session the SRS was completed. Participants were predominantly positive about the VR exercises, sessions were rated with an average satisfaction score of 9.2 out of 10 (SD = 0.3; all data were included, also from participants who dropped out in a later session). Therapists indicated that overall, there was a good fit between the session content and the participants, see Table 2. A positive trend was noticeable, with almost only good ratings from session 7 onwards, during which the scenarios were more personalized due to interactive scenarios. Sessions with lower scores were session 3 and 6. In session 3 the theory of session 2 was repeated. In session 6 physiological measurements started, using the 'aggression catwalk' as exercise (which was also used in an earlier session). Therapists indicated that there were technical issues with the VR software or hardware in 22 sessions, and problems with the physiological measurements in six sessions.

In Table 3 the qualitative evaluation of the 17 participants and four therapists is presented. The three main topics included (1) what participants learned, (2) strengths of the treatment, and (3) points of improvement. Participants learned to respond more assertively and take more time to think about different reactions. They also learned to use more appropriate reactions in aggression-triggering situations. Participants gained more insight into the internal processes leading to aggressive behavior (such as triggers for aggression and estimating their level of tension), and insight into ascending aggression in others (e.g., in estimating aggression, emotions, and facial expressions of others). Participants as well as therapists were predominantly positive in their general opinion about the sessions. As a strength of the treatment, therapists mentioned that the scenarios played during the sessions fitted treatment goals well and were sufficiently challenging. Therapists mentioned that treatment provided different insights, e.g., in different forms of reacting, insight into their behavior and triggers. Points of improvement were also made. Therapists indicated that the theoretical part was too difficult to understand for some participants (such as the SIP model or the theory on emotions). Participants mentioned they wanted to practice more with the interactive scenarios and that these could be more personalized (such as more relatable scenarios, more challenging situations and customized to their circumstances). As for the physiological measurements, there were technical problems and results were found hard to interpret. As for the hardware and software, participants mentioned that the resolution and graphical realism could be improved, the walking speed could be faster and the ability to move more was missed (such as walking away by the participant).

3.2 Intervention effects

Results of the self-report and observational measures over time are shown in Table 4. The mean scores of aggression, anger, provocation and emotion regulation decreased, with small to medium effect sizes. These improvements in mean levels were maintained at follow-up. Observational measurements showed a slight decrease in physical aggression at post-treatment, with a small effect size. This decrease was not maintained at follow-up. For verbal aggression, there

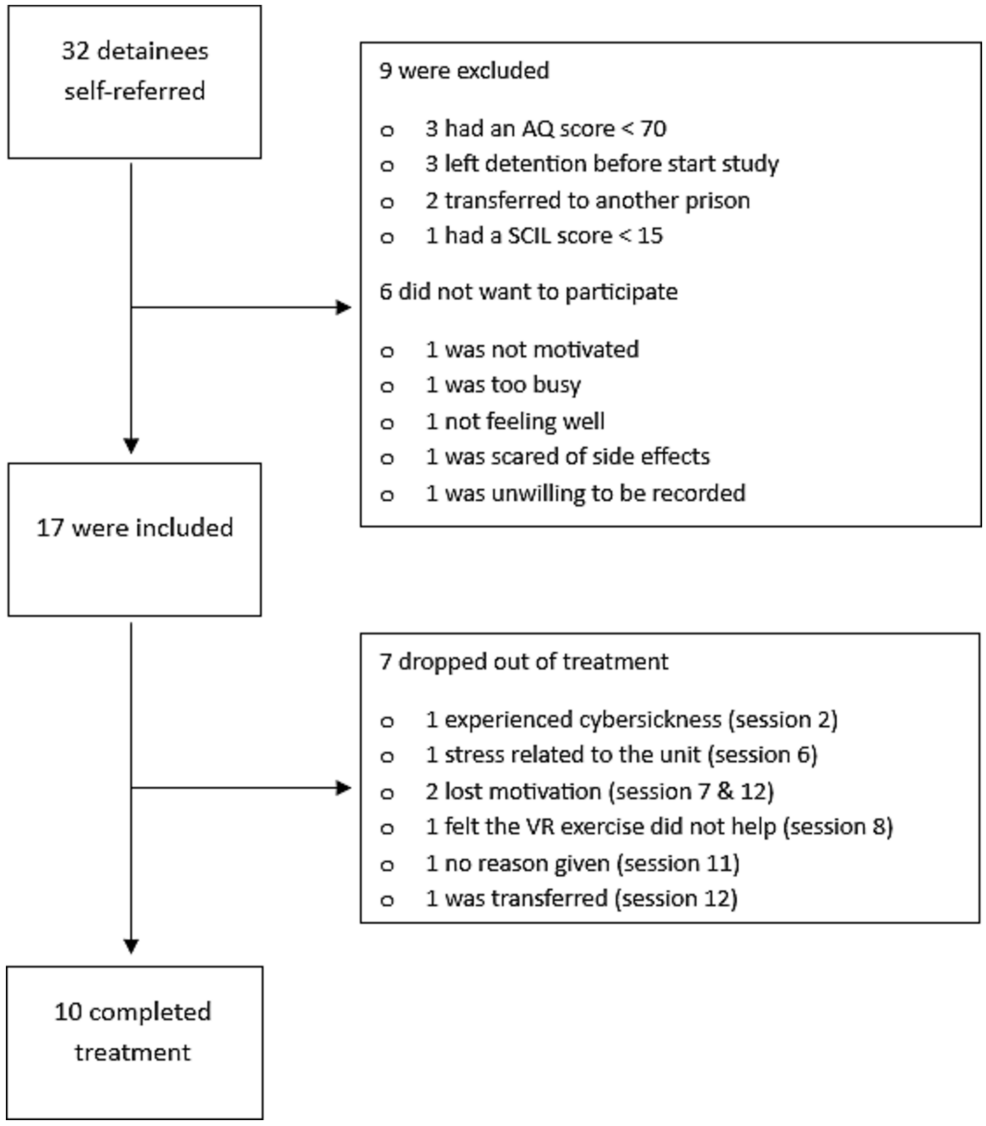


FIGURE 3
Flow chart of the pilot.

was a small decrease for post-treatment as well as for follow-up, with a small effect size.

4 Discussion

We aimed to test the feasibility and acceptability of VRAPT, a VR treatment for aggression, in a prison-based population. Overall, therapists and participants were predominantly positive about the intervention and found it to be acceptable and feasible. The most commonly named strengths were the interactive roleplays which provided new insights in aggression of themselves and others. Points of improvement for the intervention were identified; the theoretical framework was too complex and sessions needed to be more customized to the needs of the participants. Results of the questionnaires and staff observations tentatively suggested improvements in aggression, anger, provocation and emotion regulation following the treatment.

4.1 Acceptability and feasibility

Overall, the recruitment went relatively quickly, with 32 self-referrals of detainees, even though no compensation was provided. Nonparticipation after self-referral occurred mainly because of external factors such as transfers, leaving detention or not meeting the criteria. These findings are in line with the first study on VRAPT, which also showed relatively easy recruitment (Klein Tuente et al., 2020).

Regarding treatment dropouts, two participants quit because of reasons unrelated to the intervention and four participants quit due to reasons either directly or possibly related to the intervention. In total 59% of the participants completed the treatment as intended. Although this number is relatively low, high dropout rates are common in this setting (Smeijers et al., 2018). Except for one participant, an inspection of the individual session satisfaction scores showed that even dropouts rated sessions between 7.5 and 10,

10.3389/fpsyg.2023.1235808

TABLE 2 Match between session content and the participants aggression problems (therapist-rated).

	Session number													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N	17	14	16	14	14	15	13	13	12	12	11	10	10	10
Good fit	82%	64%	88%	79%	64%	86%	100%	100%	83%	100%	100%	100%	100%	100%
Reasonable fit	0%	29%	12%	21%	21%	7%	0%	0%	17%	0%	0%	0%	0%	0%
No fit	18%	7%	0%	0%	14%	7%	0%	0%	0%	0%	0%	0%	0%	0%

TABLE 3 Qualitative evaluation of the treatment obtained from the workbooks by the therapists ($n = 4$) and participants ($n = 17$).

Topic	Explanation	Illustrative quotes
1. What participants learned		
1. Responding adequately & various ways of reacting	Participants reported that they learned during the treatment to respond assertively, take the time to think about a different reaction (different than aggression), deploy de-escalating behavior, and try to seek for a solution instead of discussing/win the argument.	'That I also can react in a calm matter, I do not always need to react angry' (*participant*) 'That there are different ways of reacting in an assertive way. Walking away or taking your distance is not always sub-assertive' *(participant)*
2. Coping strategies	Participants reported they learned about different coping strategies. In the role-plays, they frequently applied strategies such as staying calm, counting to ten, helping thoughts, and walking away.	'Focus on your breath and walk away earlier from a situation' *(participant)* 'That I need to think about the consequences; what do I win by not reacting' *(participant)*
3. Insight in self –triggers	Participants reported they gained more insight into their triggers for anger/aggression and what different reactions can evoke in others.	'That there are various situations that trigger my anger, but it's mostly about authority' *(participant)* 'I can react differently than I am used to; my reaction can evoke aggression with someone else; I was not aware of that' *(participant)*
4. Insight in self – recognizing signs of aggression	Participants reported that they gained more insight into estimating aggression levels of themselves and more insight into their own tension levels.	'It is important to pay attention to signals in your body, so tension does not raise' *(participant)* 'To come to a solution, you first need to calm down' *(participant)*
5. Insight in others	Participants reported they gained more insight in estimating aggression levels, emotions, and facial expressions of others.	'You need to listen and look carefully and estimate if someone really wants to hurt you' *(participant)* 'Someone's physique does not automatically say something about his emotional state *(participant)*
2. Strengths of the treatment		
1. Scenarios triggered aggression	Therapists reported that the scenarios practiced in the treatment fitted well and were challenging enough for participants.	'Participant did not bring in own scenarios, but said these were recognizable and fitted well' *(therapist)*
2. Treatment provided insights in aggression	Therapists reported that sessions provided different forms of insight, as well as in aggression in general but also the theory provided per session.	'It's nice to see how he participates in the VR roleplays; he reacts fiercely on aggression. That's the reason he gets in trouble often … he seems to understand well what we are doing in the treatment and why' *(therapist)* 'He seems to get more insight in where his aggression is stemming from' *(therapist)*
3. Theory could be applied in VR	Therapists as well as participants were predominantly positive in their general opinion on how the sessions progressed and how the theory was applied in the VR role plays.	'Participant managed to react in the way he had intended to' *(therapist)* 'Session went well, participant makes an effort to react in an assertive way and internalize it' *(participant)*
3. Points of improvement		
1. Theory too difficult	Therapists reported that for some the SIP model was too hard to explain and apply, as well as the theory on emotions.	'Difficult to explain facial expressions when someone really does not understand facial emotions' *(therapist)* 'Nice, but ill at ease. SIP model is hard to explain' *(therapist)*

(Continued)

TABLE 3 (Continued)

Topic	Explanation	Illustrative quotes
2. Practicing more with interactive scenarios	The interactive scenarios were seen as very useful by the participants, and this should be done more during the treatment and also in earlier sessions.	'I wished to see more different reactions in the avatars, now they sometimes felt into repetition and that was less effective' *(participant)*
3. More personalized scenarios	Participants reported that they would like to practice with more recognizable scenarios in the VR role plays.	'It went well, but maybe more personalized things can be used' *(participant)* 'Add more recognizable situations' *(participant)*
4. More challenging/ aggressive scenarios	Participants reported they would like to practice more with challenging and aggressive scenarios in the VR role plays.	'There were no avatars who challenged me' *(participant)*' 'It could be more aggressive: say something about my family, be more personal: 'come to my cell, we can work it out there or everybody will be informed about the address of your family' *(participant)*
5. Personalizing sessions	Some participants mentioned that repetition was useful whereas others found it not useful. Some participants needed more sessions, some needed relaxation after the session.	'Maybe a fun exercise to lose the tension after the session. I do not want to bring it to the unit' *(participant)* 'This session could be skipped, maybe it could be an optional session, so it could be rated per person if repetition is useful' *(participant)*
6. Physiological measures	After session six it was asked therapists and participants what their experience with the physiological measurements was. Overall, it was concluded that the physiological measurements did not always work and results were hard to interpret.	'Physiological measurements were difficult to understand (graphics) *(therapist)* 'It's difficult to pay attention to the heart rate and give feedback at the same time *(therapist)*. 'Leave the ECG stickers out of the physiological measurements, they are annoying' *(participant)*
7. Hardware/software	The hardware and software could be improved; it was reported that the resolution and realism could be higher, the walking speed was experienced as too low by some and that during interactive scenarios the ability to was missed.	'Sometimes the VR world did not feel real' *(participant)*

suggesting that the dropouts due to motivation issues were unrelated to the intervention. Furthermore, it was noteworthy that drop-outs seemed to have experienced more trauma than completers of the treatment, they also seemed to be higher educated. However, these findings need to be further explored in future research with larger samples.

In general, participants and therapists were positive about VRAPT, as was shown by high session satisfaction scores by participants and good ratings by therapists on the fit between the session content and participants' needs. Through the VR exercises, participants reported to have learned to respond more adequately to aggressive behavior, gained new insights into triggers and tension, and or gained insights into aggression and (facial) emotions of others.

Although the qualitative data showed that participants acquired several skills, it was also reported that not all scenarios were provocative enough and were sometimes hard to relate to. This indicates that therapists were sometimes too careful in acting out aggressive scenarios, reflecting a different point of view between the participants and therapists when it comes to aggressive behavior. What was seen as highly aggressive by therapists can be mildly aggressive for participants. Furthermore, there needs to be more room for discussing personal situations where aggression occurs so that more relevant scenes can be roleplayed.

Therapists needed some time to adjust to the software and hardware, but over time they got more experienced and familiar with it. Extensive training and practice are needed when first starting to use VR. Some technical problems were mentioned, but these could be fixed in a relatively short time with the support of the helpdesk or by restarting the software. Therapists also reported some issues with physiological measurements. They found it difficult to play interactive scenarios and simultaneously understand and give feedback on heart rate and skin response, resulting in less usage than planned in the protocol.

Although feelings of realism and presence in VR were moderate, this did not seem to have immediate implications for the treatment. Concerning realism, no feedback was given by participants on improving the VR environments. There are indications that other factors are of more importance in how VR is perceived, such as the involvement of emotion or arousal in VR (Ling et al., 2014; Diemer et al., 2015). In accordance with this, several studies in the last two decades have used VR software that was less realistic but nonetheless effective (Freeman et al., 2017; Pot-Kolder et al., 2018).

4.2 Treatment effects

While demonstrating the efficacy of VRAPT was not an aim of this pilot, nearly all effects were in the expected direction showing improvements between baseline and post-treatment (except for impulsiveness). Self-reported aggression improved between baseline and posttreatment with medium effect size, and the effect was maintained at follow-up. Further, outcomes on anger, provocation,

TABLE 4 Means, standard deviations, and test results of outcomes over time.

	Baseline		Post-treatment			Follow-up		
	n	Mean (SD)	*n*	Mean (SD)	Effect size	*n*	Mean (SD)	Effect size
Self-report measures								
Aggression (AQ) total score	17	98.6 (16.0)	14	90.7 (11.0)	0.37*	8	90.1 (16.9)	0.39*
Physical aggression	17	35.6 (6.7)	14	32.3 (6.9)	0.29	8	31.1 (8.0)	0.40
Verbal aggression	17	17.0 (2.7)	14	16.5 (2.7)	0.07	8	16.6 (4.3)	0.25
Anger	16	22.7 (4.9)	14	20.4 (4.3)	0.46	8	22.1 (4.6)	0.04
Hostility	17	23.1 (6.9)	14	21.6 (5.9)	0.09	8	20.3 (7.6)	0.56
Reactive & proactive aggression (RPQ) total score	17	23.1 (9.1)	14	21.5 (7.9)	0.29*	8	22.4 (11.7)	0.05
Reactive aggression	17	13.2 (4.3)	14	12.7 (4.1)	−0.01	8	12.8 (5.3)	0.14
Proactive aggression	17	9.9 (5.6)	14	8.9 (4.5)	0.21*	8	9.6 (6.5)	0.18*
Anger (NAS)	16	103.7 (14.8)	14	94.9 (15.5)	0.51	8	96.8 (13.2)	0.46
Cognitive	16	34.9 (4.7)	14	33.1 (5.0)	0.30	8	32.7 (3.0)	0.47
Arousal	16	34.7 (6.5)	14	30.5 (6.4)	0.66	8	32.4 (4.3)	0.30
Behavior	16	34.1 (5.6)	14	31.3 (5.7)	0.29	8	31.8 (6.9)	0.56
Provocation (PI)	15	60.7 (11.9)	14	55.3 (11.4)	0.31	8	53.4 (14.0)	0.48
Impulsiveness (BIS-11)	17	68.7 (13.9)	14	69.8 (14.8)	−0.46*	8	70.1 (17.1)	−0.11*
Emotion regulation (DERS)	16	94.3 (27.5)	13	85.9 (21.6)	0.28	7	69.1 (11.8)	0.58
Observational measures								
Physical aggression (SDAS)	17	0.35 (0.79)	17	0.20 (0.32)	0.24*	12	0.22 (0.47)	0.05*
Verbal aggression (SDAS)	17	4.95 (4.70)	17	4.07 (3.87)	0.19*	12	2.58 (2.56)	0.17*

Conventions effect size Hedges' g correction for paired samples (0.20–0.49), medium (0.50–0.79), and large effect (≥0.80).
*Conventions effect size Wilcoxon signed-rank test effect size: small (0.10–0.29), medium (0.30–0.49), and large effect (≥0.50).

and emotion regulation decreased with small to medium effect sizes. It is important to notice that questionnaires concerning aggression focused on trait aggression, which is also affected by other factors such as childhood trauma (Sarchiapone et al., 2009), and may influence effects. Small improvements in staff-observed aggression were found, especially in verbal aggression. SDAS observations only focus on aggressive behavior, and newly learned positive (coping) behavior is thus not scored.

Whereas scores of most measures stabilized between posttreatment and follow-up, emotion regulation improved further in the period after the treatment. This finding seems to fit well with the treatment targets, and converges with the qualitative findings which reported that participants learned new coping skills and gained insights in estimating signs of their aggression which is a precursor for being able to regulate such emotions. Maladaptive emotion regulation is common in offenders (Roberton et al., 2014). A recent study showed that emotion regulation moderates the effect between anger and aggression in aggressive offenders (Xie et al., 2023), further stressing the importance of intervening on emotion regulation. This could also indicate that longer follow-up periods are relevant for aggression outcomes, as in response to improved emotion regulation, aggression might decrease over time as well. As such, it would be interesting to explore in future research whether emotion regulation mediates or moderates treatment effects on aggression.

The results of the current study are in line with the earlier VRAPT study of Klein Tuente et al. (2020), as mean scores on aggression, anger and provocation (as measured with the AQ and NAS-PI) changed in similarly in both studies. It is noticeable that mean baseline scores of aggression on both the AQ and RPQ were overall higher in this study than the study of Klein Tuente and colleagues, which may result in detecting small changes in aggression easier in this prisoner population.

4.3 Adjustments to the treatment protocol

Based on this pilot, several adjustments were made to the treatment protocol. First, the theory explained during the treatment was minimized and simplified as criticism by the therapists indicated the theory was hard to understand and apply for participants. Instead of explaining SIP steps in each session in-depth, the model is now used as an underpinning for the therapists and is explained explicitly to participants during session 1 only. As a more practical replacement, elements from cognitive behavioral therapy (CBT) are used, providing insights in the relation

between thoughts, feelings, behavior, and the consequences of behavior in a specific situation. Also, theory was linked directly to the different exercises to make it more practical and understandable.

Second, homework assignments were added to both increase treatment efficacy and enable creating more personalized and challenging VR scenarios. Participants were initially encouraged to apply what they learned in each session in daily life. However, this turned out to be difficult for most of them. To give more guidance, participants have to complete forms on the think-feel-act-consequence (the CBT-related exercise) as a homework exercise about aggressive or stressful events that week (Beck, 2011). The forms will also be used as input for creating personal scenarios. In the pilot, participants found it hard to come up with concrete examples for personal scenarios, and the new homework assignment can help with this.

Third, to enable more personalized sessions and practice more in VR during the treatment, we replaced two sessions (sessions 4 and 9) with two sessions in which the therapist and participant can freely choose what to practice. In this way, the treatment can be tailored more to the individual and specific assignments with which the participant has difficulty can be practiced.

Finally, physiological measurements were removed from the treatment protocol as the data was too difficult to monitor and interpret for therapists during roleplays (which resulted in minimal usage). Furthermore, the ECG was experienced as uncomfortable by participants. However, as it was relevant to discuss experienced tension, an 'anger thermometer' was added to session 5 and 8 which is a common tool for discussing tension (Rose et al., 2008).

4.4 Limitations

This study had several limitations. First, participation was based on self-referral, which may have led to selection bias. Participants may have been more motivated for (innovative) treatment than the average prison population. However, treatment in Dutch prisons is always on a voluntary basis. No further documentation was kept on the reasons why participants wanted to participate, which is a limitation. Also, only males were included in this study, so we do not know whether the findings would be the same or different in females.

Data were missing for some questionnaires, and questions concerning feedback on the protocol and study were open-ended questions in the workbook of the therapist that were not always fully completed. Also, it is unknown which other forms of care participants received during the study period.

Furthermore, participants completed the qualitative questions in the workbooks in the presence of the therapist, which may have caused a positive bias as therapists were not blinded and participants may have been less critical and may have given more socially desirable answers due to the presence of the therapist. To minimize positive bias it was emphasized that therapists shared no content-related information with any other parties and that participating or dropping out could not influence (positive or negative) ongoing trajectories in detention in any way, to clarify for participants that there were no further gains from participating. Also, during the study period, VRAPT therapists only had contact with participants for VRAPT and not for any other reason.

Finally, only about half of the participants completed the follow-up measurement. We checked whether participants who did not complete measurements differed in the main outcome from completers at the start of the study. This was not the case, baseline aggression total scores revealed that both participants who completed the follow-up measure ($M = 97.9$, $SD = 15.9$) and who did not complete the measure ($M = 99.2$, $SD = 16.9$) had similar levels of aggression at the start of the study. Thus, this does not seem to have caused a bias in aggression outcomes.

4.5 Conclusion

Our findings indicate that VRAPT is an acceptable and feasible intervention for both detainees and therapists to train multiple skills for reducing aggressive behavior. Furthermore, preliminary positive findings on aggression, anger, and emotion regulation suggest that this treatment has potential in a prison-based population. Implementing VRAPT in a larger-scale RCT requires several adjustments, such as simplifying the theoretical framework and roleplaying with more personalized scenarios. Based on our findings we have adjusted the treatment protocol to a new version specifically for detainees called Virtual Reality-Treatment for Aggression Control (VR-TrAC). Our next step will be to test VR-TrAC in an RCT.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors upon request, without undue reservation.

Ethics statement

The studies involving humans were approved by Medical Ethical review board UMC Groningen PO Box 300,019,700 RB GRONINGEN. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

WV, EM, and SK designed the study. KW collected the data. KW and CG wrote the first draft of the manuscript. KW and CG carried out the analysis of the results. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Auty, K. M., Cope, A., and Liebling, A. (2017). Psychoeducational programs for reducing prison violence: a systematic review. *Aggress. Violent Behav.* 33, 126–143. doi: 10.1016/j.avb.2017.01.018

Beck, J. S. (2011). Cognitive behavior therapy, second edition: Basics and beyond | enhanced reader.

Buss, A. H., and Perry, M. (1992). The aggression questionnaire. *J. Pers. Soc. Psychol.* 63, 452–459. doi: 10.1037/0022-3514.63.3.452

Cima, M., Raine, A., Meesters, C., and Popma, A. (2013). Validation of the Dutch reactive proactive questionnaire (RPQ): differential correlates of reactive and proactive aggression from childhood to adulthood. *Aggress. Behav.* 39, 99–113. doi: 10.1002/AB.21458

Coccaro, E. F., Fanning, J., and Lee, R. (2017). Development of a social emotional information processing assessment for adults (SEIP-Q). *Aggress. Behav.* 43, 47–59. doi: 10.1002/ab.21661

Crick, N. R., and Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in Children's social adjustment. *Psychol. Bull.* 115, 74–101. doi: 10.1037/0033-2909.115.1.74

Cuervo, K., Villanueva, L., Born, M., and Gavray, C. (2018). Analysis of violent and non-violent versatility in self-reported juvenile delinquency. *Psychiatry Psychol. Law* 25, 72–85. doi: 10.1080/13218719.2017.1347935

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., and Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Front. Psychol.* 6, 1–9. doi: 10.3389/fpsyg.2015.00026

Dodge, K. A., and Crick, N. R. (2007). Social information-processing bases of aggressive behavior in children. *Personal. Soc. Psychol. Bull.* 16, 8–22. doi: 10.1177/0146167290161002

Duncan, B., Miller, S., Sparks, J., Claud, D. A., Reynolds, L., Brown, J., et al. (2003). The session rating scale: preliminary psychometric properties of a "working" Alliance measure. *J. Brief Therapy* 3, 3–12.

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychol. Med.* 47, 2393–2400. doi: 10.1017/S003329171700040X

Geraets, C. N. W., van der Stouwe, E. C. D., Pot-Kolder, R., and Veling, W. (2021). Advances in immersive virtual reality interventions for mental disorders: a new reality? *Curr. Opin. Psychol.* 41, 40–45. doi: 10.1016/j.copsyc.2021.02.004

Gratz, K. L., and Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J. Psychopathol. Behav. Assess.* 26, 41–54. doi: 10.1023/B:JOBA.0000007455.08539.94

Hornsveld, R. H. J., Muris, P., and Kraaimaat, F. W. (2011). The Novaco anger scale-provocation inventory (1994 version) in Dutch forensic psychiatric patients. *Psychol. Assess.* 23, 937–944. doi: 10.1037/A0024018

Hornsveld, R. H. J., Muris, P., Kraaimaat, F. W., and Meesters, C. (2009). Psychometric properties of the aggression questionnaire in Dutch violent forensic psychiatric patients and secondary vocational students. *Assessment* 16, 181–192. doi: 10.1177/1073191108325894

Igroup (n.d.). Igroup presence questionnaire (IPQ) factor analysis. Available at: http://www.igroup.org/pq/ipq/factor.php [Accessed August 17, 2023].

Jochems, E. C., Mulder, C. L., van Dam, A., Duivenvoorden, H. J., Scheffer, S. C. M., van der Spek, W., et al. (2012). Motivation and treatment engagement intervention trial (MotivaTe-IT): the effects of motivation feedback to clinicians on treatment engagement in patients with severe mental illness. *BMC Psychiatry* 12:209. doi: 10.1186/1471-244X-12-209

Kip, H., Bouman, Y. H. A., Kelders, S. M., and van Gemert-Pijnen, L. J. E. W. C. (2018). eHealth in treatment of offenders in forensic mental health: a review of the current state. *Front. Psychol.* 9:42. doi: 10.3389/FPSYT.2018.00042/FULL

Kip, H., Kelders, S. M., Weerink, K., Kuiper, A., Brüninghoff, I., Bouman, Y. H. A., et al. (2019). Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. *Front. Psychol.* 10:406. doi: 10.3389/FPSYG.2019.00406/FULL

Klein Tuente, S., Bogaerts, S., Bulten, E., Keulen-de Vos, M., Vos, M., Bokern, H., et al. (2020). Clinical medicine virtual reality aggression prevention therapy (VRAPT) versus waiting list control for forensic psychiatric inpatients: a multicenter randomized controlled trial. *J. Clin. Med.* 9:2258. doi: 10.3390/jcm9072258

Klein Tuente, S., Bogaerts, S., van IJzendoorn, S., and Veling, W. (2018). Effect of virtual reality aggression prevention training for forensic psychiatric patients (VRAPT): study protocol of a multi-center RCT. *BMC Psychiatry* 18:251. doi: 10.1186/s12888-018-1830-8

Kobes, M. H. B. M., Nijman, H. H. L. I., and Bulten, E. B. H. (2012). Assessing aggressive behavior in forensic psychiatric patients: validity and clinical utility of combining two instruments. *Arch. Psychiatr. Nurs.* 26, 487–494. doi: 10.1016/J.APNU.2012.04.004

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., and Brinkman, W. P. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PLoS One* 9, 1–12. doi: 10.1371/journal.pone.0096144

Lipsey, M. W., and Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: a review of systematic reviews. *Ann. Rev. Law Soc. Sci.* 3, 297–320. doi: 10.1146/annurev.lawsocsci.3.081806.112833

McGuire, J. (2008). Review. A review of effective interventions for reducing aggression and violence. *Phil. Trans. R. Soc. B Biol. Sci.* 363, 2577–2597. doi: 10.1098/rstb.2008.0035

Mcguire, J. (2018). *Understanding prison violence: A rapid evidence assessment*.

Mcneeley, S. (2020). Situational risk factors for inmate-on-staff assault. Available at: www.doc.state.mn.us (Accessed August 17, 2023).

Mei, X., Li, J., Li, Z. S., Huang, S., Li, L. L., Huang, Y. H., et al. (2022). Psychometric evaluation of an adverse childhood experiences (ACEs) measurement tool: an equitable assessment or reinforcing biases? *Health Justice* 10, 34–11. doi: 10.1186/s40352-022-00198-2

Mooney, J. L., and Daffern, M. (2015). The relationship between aggressive behaviour in prison and violent offending following release. *Psychol. Crime Law* 21, 314–329. doi: 10.1080/1068316X.2014.989163

Muñoz García-Largo, L., Martí-Agustí, G., Martin-Fumadó, C., and Gómez-Durán, E. L. (2020). Intellectual disability rates among male prison inmates. *Int. J. Law Psychiatry* 70:101566. doi: 10.1016/j.ijlp.2020.101566

Murphy, M. G., Rakes, S., and Harris, R. M. (2020). The psychometric properties of the session rating scale: a narrative review. *J. Evid. Based Soc. Work* 17, 279–299. doi: 10.1080/26408066.2020.1729281

Papalia, N., Spivak, B., Daffern, M., and Ogloff, J. R. P. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clin. Psychol. Sci. Pract.* 26, 1–28. doi: 10.1111/cpsp.12282

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Leg. Criminol. Psychol.* 17, 1–17. doi: 10.1111/j.2044-8333.2011.02038.x

Pot-Kolder, R., Geraets, C. N. W., Veling, W., van Beilen, M., Staring, A. B. P., Gijsman, H. J., et al. (2018). Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. *Lancet Psychiatry* 5, 217–226. doi: 10.1016/S2215-0366(18)30053-1

Recidivism Among Federal Violent Offenders (2019). Washington.

Roberton, T., Daffern, M., and Bucks, R. S. (2014). Maladaptive emotion regulation and aggression in adult offenders. *Psychol. Crime Law* 20, 933–954. doi: 10.1080/1068316X.2014.893333

Rose, J. L., Dodd, L., and Rose, N. (2008). Individual cognitive behavioral intervention for anger. *J. Ment. Health Res. Intellect. Disabil.* 1, 97–108. doi: 10.1080/19315860801988368

Sarchiapone, M., Carli, V., Cuomo, C., Marchetti, M., and Roy, A. (2009). Association between childhood trauma and aggression in male prisoners. *Psychiatry Res.* 165, 187–192. doi: 10.1016/j.psychres.2008.04.026

Schippers, G. M., Broekman, T. G., Buchholz, A., Koeter, M. W. J., and Van Den Brink, W. (2010). Measurements in the addictions for triage and evaluation (MATE): an instrument based on the World Health Organization family of international classifications. *Addiction* 105, 862–871. doi: 10.1111/J.1360-0443.2009.02889.X

Schippers, G., Broekman, T., Merkx, M., and Buchholz, A. (2011). Treatment allocation guidelines on treatment outcome in routine outpatient addiction treatment using structured assessment of patient characteristics. *Suchttherapie* 12:S3_1. doi: 10.1055/s-0031-1284488

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: factor analytic insights. *Presence Teleop. Virt.* 10, 266–281. doi: 10.1162/105474601300343603

Smeijers, D., Bulten, E., Buitelaar, J., and Verkes, R. J. (2018). Treatment responsivity of aggressive forensic psychiatric outpatients. *Int. J. Offender Ther. Comp. Criminol.* 62, 3834–3852. doi: 10.1177/0306624X17747052

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., and Patton, J. H. (2009). Fifty years of the Barratt impulsiveness scale: an update and review. *Personal. Individ. Differ.* 47, 385–395. doi: 10.1016/j.paid.2009.04.008

Sygel, K., and Wallinius, M. (2021). Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. *Front. Psychol.* 12:673089. doi: 10.3389/fpsyt.2021.673089

van der Feltz-Cornelis, C. M., Potters, E. C., van Dam, A., Koorndijk, R. P. M., Elfeddali, I., and Van Eck Van Der Sluijs, J. F. (2019). Adverse childhood experiences (ACE) in outpatients with anxiety and depressive disorders and their association with psychiatric and somatic comorbidity and revictimization. Cross-sectional observational study. *J. Affect. Disord.* 246, 458–464. doi: 10.1016/J.JAD.2018.12.096

van Esch, A. Y. M., de Vries, J., and Masthoff, E. D. M. (2020). Screening for intellectual disability in Dutch psychiatrically disturbed detainees: assessing the psychometric properties of the screener for intelligence and learning disability (SCIL). *J. Appl. Res. Intellect. Disabil.* 33, 1418–1427. doi: 10.1111/jar.12769

Walk, D., Haviv, N., Hasisi, B., and Weisburd, D. (2021). The role of employment as a mediator in correctional Education's impact on recidivism: a quasi-experimental study of multiple programs. *J. Crim. Just.* 74:101815. doi: 10.1016/j.jcrimjus.2021.101815

Wildra, E. (2020). No escape: the trauma of witnessing violence in prison. Available at: https://www.prisonpolicy.org/blog/2020/12/02/witnessing-prison-violence/ [Accessed June 6, 2023].

Wistedt, B., Rasmussen, A., Pedersen, L., Malm, U., Träskman-Bendz, L., Wakelin, J., et al. (1990). The development of an observer-scale for measuring social dysfunction and aggression. *Pharmacopsychiatry* 23, 249–252. doi: 10.1055/S-2007-1014514

Xie, Q., Zhou, B., Bi, T., Yang, B., and Kou, H. (2023). Trait anger and aggression among male violent offenders: the mediating effect of sadistic impulse and the moderating effect of emotion regulation. *J. Psychol. Afr.* 33, 10–16. doi: 10.1080/14330237.2023.2182968

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 16 June 2023
DOI 10.3389/fpsyg.2023.1184016

OPEN ACCESS

EDITED BY
Joan E. Van Horn,
De Forensische Zorgspecialisten, Netherlands

REVIEWED BY
Ko-Chiu Wu,
National Taipei University of Technology, Taiwan
Loredana Ivan,
National School of Political Studies and Public Administration, Romania

*CORRESPONDENCE
Ann-Pierre Raiche
✉ ann-pierre.raiche@umontreal.ca

RECEIVED 10 March 2023
ACCEPTED 05 June 2023
PUBLISHED 16 June 2023

CITATION
Raiche A-P, Dauphinais L, Duval M, De Luca G, Rivest-Hénault D, Vaughan T, Proulx C and Guay J-P (2023) Factors influencing acceptance and trust of chatbots in juvenile offenders' risk assessment training. *Front. Psychol.* 14:1184016. doi: 10.3389/fpsyg.2023.1184016

Factors influencing acceptance and trust of chatbots in juvenile offenders' risk assessment training

Ann-Pierre Raiche[1]*, Léanne Dauphinais[1], Manon Duval[1], Gino De Luca[2], David Rivest-Hénault[2], Thomas Vaughan[2], Catherine Proulx[2] and Jean-Pierre Guay[1]

[1]School of Criminology, University of Montreal, Montreal, QC, Canada, [2]National Research Council Canada, Boucherville, QC, Canada

Introduction: Research has identified simulation-based training with chatbots and virtual avatars as an effective educational strategy in some domains, such as medicine and mental health disciplines. Several studies on interactive systems have also suggested that user experience is decisive for adoption. As interest increases, it becomes important to examine the factors influencing user acceptance and trust in simulation-based training systems, and to validate applicability to specific learning tasks. The aim of this research is twofold: (1) to examine the perceived acceptance and trust in a risk assessment training chatbot developed to help students assess risk and needs of juvenile offenders, and (2) to examine the factors influencing students' perceptions of acceptance and trust.

Methods: Participants were 112 criminology students in an undergraduate course in a Canadian university. Participants were directed to use a custom-designed chatbot with a virtual 3D avatar for juvenile offenders' risk assessment training, to complete online questionnaires and a risk assessment exercise.

Results: Results show satisfactory levels of acceptance and trust in the chatbot. Concerning acceptance, more than half appeared to be satisfied or very satisfied with the chatbot, while most participants appeared to be neutral or satisfied with the benevolence and credibility of the chatbot.

Discussion: Results suggest that acceptance and trust do not only depend on the design of the chatbot software, but also on the characteristics of the user, and most prominently on self-efficacy, state anxiety, learning styles and neuroticism personality traits. As trust and acceptance play a vital role in determining technology success, these results are encouraging.

KEYWORDS

virtual learning-based simulation, chatbot, education, training, risk assessment, acceptance, trust

Introduction

Education and simulation

The didactic lecture format is the dominant teaching method within most higher education courses across disciplines (Butler, 1992). This method is commonly used because of its economical and practical features, especially with many students and limited resources (Alaagib et al., 2019). Although didactic lecture is one of the most common teaching methods, it presents numerous limitations across disciplines, especially those requiring clinical learning and training skills. For example, the concepts taught during didactic lectures are difficult to translate into practice and opportunities to gain clinical experience with real patients are limited (Mazmanian and Davis, 2002; Rizzo and Talbot, 2016). Research has shown that to be more effective, those lectures must be used combined with other methods and techniques (Butler, 1992; Meyers and Jones, 1993).

Simulations-based learning is considered one of the most effective methods to improve the learning of complex skills across disciplines (Chernikova et al., 2020). Simulation is defined as "[. . .] a technique (not a technology) to replace and amplify real experiences with guided ones, often 'immersive' in nature, that evoke or replicate substantial aspects of the real world in a fully interactive fashion" (Lateef, 2010, p. 348). Simulations can range in complexity and presentation, for example peer-to-peer role play or using live actors to portray patients (Chernikova et al., 2020). Currently, training methods using simulated patients are considered the gold standard to develop interviewing, assessment, and diagnostic skills in nursing, medicine, and psychology (Mooradian, 2008; McGaghie et al., 2011).

Simulations can also be enhanced by technology. Cook et al. (2013) define technology-enhanced simulation as an "educational tool or device with which the learner physically interacts to mimic real life and in which they emphasize the necessity of interacting with authentic objects" (p. 876). Technology-enhanced simulation offers innovative solutions to address many limitations associated with the use of standard simulated patients (Washburn et al., 2016). Until recently, there were very few avenues available to organizations wishing to enhance the knowledge of students, but technological advances have enabled the development of innovative methods.

Artificial intelligence and education

There is a growing interest in the use of artificial intelligence in the field of education (Roos, 2018; Okonkwo and Ade-Ibijola, 2021). To support teaching and learning activities, chatbots powered by artificial intelligence are one of the most popular technology-enhanced simulation applications across fields of study such as nursing, medicine, psychology (Okonkwo and Ade-Ibijola, 2021). In the fields of clinical psychology, psychiatry, social work, criminology and particularly in learning tasks, various chatbots have been created in the last years. For example, in the medical field, Kenny et al. (2007) developed "Justin," a human virtual agent used to practice professional interviewing techniques as well as to improve recognition of signs and symptoms of behavioral disorders. More recently, Washburn et al. (2016, 2020), developed six different virtual personas designed to allow social work students to practice asking interview questions, creating a positive therapeutic alliance, and gathering clinical information to recognize mental health disorders.

In artificial intelligence research, terms like chatbot, conversational agent, embodied conversational agent, virtual agent, virtual assistant, and even avatar are used synonymously and interchangeably (von der Pütten et al., 2010). Although there are some subtle distinctions between these terms (see McTear, 2020 for more details), for the purpose of this study, the term chatbot is used, and refers to

. . . digital tools existing either as hardware (such as an Amazon Echo running the Alexa digital assistant software) or software (such as Google Assistant running on Android devices or Siri running on Apple devices) that use machine learning and artificial intelligence methods to mimic humanlike behaviors and provide a task-oriented framework with evolving dialogue able to participate in conversation (Vaidyam et al., 2019, p. 457).

There are two categories of chatbots, "simple chatbot" and "smart or advanced chatbot" (Veretskaya, 2017). Simple chatbots are rule-based chatbots, which means that they depend on prewritten keywords chosen by the developer. In other words, predetermined options restrict user interaction and there are very few opportunities for free responses from the user. For example, if a user enters a question without one of the prewritten keywords, the chatbot won't be able to understand the question and will respond a default message like "Sorry, I did not understand" (Veretskaya, 2017). Despite these restrictions, simple chatbots are widely used in several areas because they are easy to use and quick to implement (Schmitt, n.d.). Smart or advanced chatbots are artificial intelligence-based chatbots, which means they use Machine learning (ML) and Natural Language Processing (NLP). ML is a "branch of artificial intelligence and computer science which focuses on the use of data and algorithms to imitate the way that humans learn" (IBM Cloud Education, 2020a, para 1), while NLP refers to "the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can" (IBM Cloud Education, 2020b, para 1).

Virtual simulated-based learning using chatbot systems present several advantages over traditional learning methods. One of the benefits is the great versatility and adaptability of the virtual characters. Chatbots offer the possibility to create diverse personalities or case studies with different physical/sociodemographic characteristics such as hair color, skin color, gender, and age, but also different clinical needs such as mental health concerns, physical health concerns, criminal dynamics, etc. (Washburn et al., 2020). Another advantage is availability and accessibility. Chatbots can be installed on or accessed from personal computers and do not require a specific space or specialized equipment. Effectively, they can be used at any moment and at any place (Triola et al., 2006; Washburn and Zhou, 2018). They can also be used repeatedly and by multiple users at the same time, which can be particularly useful for large cohorts of students (Washburn and Zhou, 2018; Washburn et al., 2020). In addition, unlike traditional approaches using actors, systems using chatbots are not subject to the variability within actors or

the availability issues of actors, stakeholders, and organizations (Washburn et al., 2020). In the long run, the use of virtual patients may be more affordable than actor-based simulations as they can be used yearly and can be shared across departments or institutions (Washburn et al., 2020). Chatbot programs are not only a safe learning environment for students but also for patients or clients. They offer students a space to safely try new approaches and new techniques. As many professionals from different fields such as medicine or psychology work with vulnerable populations, it is important to offer students a place where they can make mistakes and try strategies without having a negative impact on their patients (Kenny et al., 2008; Washburn and Zhou, 2018; Coyne et al., 2021). Chatbots can also offer systematic feedback to the user. Some chatbot programs automatically save a text log of their interactions with their user, which can be used to review their performance, including successes and mistakes (Washburn et al., 2020).

Research on virtual simulated-based learning using chatbot systems identifies this method as an effective educational strategy (Chernikova et al., 2020). Research suggests that the skills learned by students using virtual patient simulations can be equivalent to the skills learned using standard simulations with actors (Cook et al., 2010) and that these skills are applicable in real-world situations involving patients (Triola et al., 2006; Washburn et al., 2016). Previous studies have focused on mechanisms that explain the effectiveness of this educational strategy. Those studies suggest that factors such as interactivity, ease of use, well-developed backstories, the realism of the clinical scenarios, and the availability of timely feedback increased usability and clinical skill acquisition (Cook et al., 2010; Bateman et al., 2012).

Factors influencing acceptance and trust

As interest in chatbots as an effective learning tool increase, it is important to examine the factors that influence user acceptance and trust to use them. In their systematic review, Ling et al. (2021) have identified five categories of factors that influence chatbot adoption, namely usage-related factors (such as perceived usefulness and ease of use), agent-related factors (such as visual appearance and gesturing), user-related factors (such as demographic information and technology experience), attitude and evaluation factors (such as attitudes and satisfaction), and other factors (such as social influence). This study focuses on the user-related factors because studies suggest that these factors can influence engagement, acceptance, and trust in technologies but that they have not been sufficiently studied (Philip et al., 2020).

Acceptance

Several factors were identified to impact acceptance of chatbot. As present by Ling et al. (2021), these factors included demographic factors (gender, age), users' expertise with technology and psychological factors.

Some studies indicate that there are some age-related differences in the usability and acceptance of a chatbot. Research in the field of technology acceptance indicates that perceived ease of use and perceived security of several technologies differ between older and younger adults (Grimes et al., 2010; Mitzner et al., 2010). Grimes et al. (2010) found out that older adults are less likely to be using technologies and less knowledgeable about security than younger adults (Grimes et al., 2010). Other research suggests that there is no difference between age groups and that the relation between age and technology acceptance is a complex one (Mitzner et al., 2010; McLean and Osei-Frimpong, 2019).

Research conducted more than a decade ago also suggested gender-related differences (Thompson and Lim, 1996; Milis et al., 2008; Padilla-Meléndez et al., 2008). There are some gender differences in perceptions of whether the technologies are easy to use. Thus, females tend to view technologies as being less easy to use compared to males (Thompson and Lim, 1996; van Braak, 2004; Milis et al., 2008). The results also show that males appear to have more previous experience with technologies than females (Thompson and Lim, 1996). Moreover, more recent research about technology acceptance indicated the opposite. Milis et al. (2008) suggest that females feel insecure when using a new virtual learning environment due to the novelty. However, they also indicate that females with attitudes more favorable toward thinking and learning are more likely to have a more favorable perception of usability. In opposite, males feel more secure, but they need an external motivation to engage in a virtual learning environment. In their study about the acceptability of an application for collecting symptom and quality-of-life information for patients, Wolpin et al. (2008) found that women found the program more acceptable than man. There is also inconsistency within research regarding the difference between males and females. Although some studies suggest that gender plays a significant role in determining the intention of accepting new technology, other studies found no differences between males and females (Suri and Sharma, 2013; McLean and Osei-Frimpong, 2019).

Beyond the degree of experience or familiarity with technology, research suggests that the user's immersive tendencies can influence chatbot acceptance. Previous research demonstrates that participants with highly immersive tendencies will feel more present in the virtual environment and enjoy the experience more than a participant who does not generally become immersed in activities (Witmer and Singer, 1998; Johns et al., 2000; Nunez, 2003).

In terms of personality traits, their effects on technology acceptance have rarely been studied. Available research shows that different personality traits impact acceptance (McKnight et al., 2002; Brown et al., 2004; Müller et al., 2019). Research demonstrates that curiosity (Brandtzaeg and Følstad, 2017), personal innovativeness (Frambach et al., 2000; Richad et al., 2019), and hypervigilance (Mäurer and Weihe, 2015) have a positive influence on their perception of acceptance and usefulness of chatbots. In addition, research suggests that openness to experience and extraversion are also positively related to the acceptance of new technology (Islam et al., 2017). Research also suggests that self-efficacy and anxiety can play a role in technology acceptance (Czaja et al., 2006). In their study, Czaja et al. (2006) found that computer self-efficacy was an important predictor of general use of technology and that people with lower self-efficacy are less likely to use technology in general. They also found that self-efficacy has an indirect effect on technology adoption through anxiety, such that people with lower self-efficacy would have higher anxiety.

In addition, psychological traits such as learning styles seem to play a role in explaining and understanding user reactions to

systems. Learning styles refer to the preferential way in which the individual absorbs, processes, and retains information and skills (Reid, 1995). Individual learning styles depend on cognitive, affective, environmental factors, and prior experience (Othman and Amiruddin, 2010). Studies on learning styles suggests that it is important to match the learning and teaching styles because it affects academic achievement and learner satisfaction (Felder and Silverman, 1988; Felder, 1993; Coffield et al., 2004). However, some others suggest that mismatch (i.e., using teaching style that are not suitable with learning style) might challenge students to adjust and learn in more integrated ways (Entwistle, 1988; Robotham, 1995; Vita, 2001). Despite some inconsistencies in the studies about the relationship between learning style and technology acceptance, the relationship between learning styles and perceived satisfaction is evident (Felder and Brent, 2005). Within the psychological domain, some authors claim that the learning style is one of the most important individual differences that affect learner performance and satisfaction, which also influences acceptance (Dunn and Dunn, 1974; Felder, 1988; Kolb and Kolb, 2005). According to these authors, learning styles can motivate students and thereby enhance sense of achievement and/or satisfaction.

Trust

Concerning trust, few studies have focused on factors identified to impact trust of chatbot. These studies also indicate that there are some age-related differences in the trust of a chatbot. Hoff and Bashir (2015) suggested that older people trust automated processes less than younger people. Følstad and Brandtzaeg (2020) also found out that older adults appreciated the pragmatic chatbot attributes (i.e., usefulness and usability) while younger participants appreciated the hedonic chatbot attributes (i.e., characteristics associated with the mental or emotional wellbeing of the user).

Acceptance, trust in chatbot and education

Studies on interactive systems emphasize on the fact that acceptance and trust play a vital role in determining technology success. User experience is decisive for the adoption and implementation of such systems, especially in education (Young et al., 2008; Hornbæk and Hertzum, 2017). When accepted and implemented correctly, chatbots can be a useful technology to facilitate learning within the educational context (Clarizia et al., 2018). Until now, very few studies have looked at the user-related factors that influence acceptance and trust of a chatbot in a training context. Indeed, except for demographic factors such as age and gender, knowledge is very limited.

In health-related professions, the level of education and clinical competency is a key factor in improving client outcomes (Coyne et al., 2021). Professionals must be competent in interviewing techniques, symptom/ability assessment, diagnosis, motivational interviewing, and interpersonal communication. An effective interview structure needs to cover all areas of potential clinical concerns and no mistakes can be made (Fernández-Ballesteros et al., 2003). In the course of their work, professionals are asked to interact and make crucial decisions in sensitive contexts that may have an influence on both individuals being assessed and on society. In the forensic field, it is the responsibility of the professionals to assess the risk of violence. Risk assessment is a process involving the systematic collection of information from several sources (e.g., data collection from interviews, case files, family, parents, employers, or teachers) to determine whether someone is likely to use violence, against themselves or another person, in the near future. This evaluation is important since it allows professionals to establish a treatment plan adapted to the person's needs, treatment plan which aim to reduce the risk of violence and promote community reintegration (Guay et al., 2022). To do this evaluation, professionals use structured risk assessment instruments. For both adults and youth in Canada, these assessments are conducted systematically and influence the entire judicial process, particularly at the release level. It is crucial for public safety that professionals are competent because a bad decision can have serious impacts on public safety.

The current research

To our knowledge, there is a limited number of chatbots with virtual avatars available that are useful for training professionals working in the forensic field. This research aims to examine how are acceptance and trust perceived in a recently developed juvenile risk assessment training chatbot, and what are the user-related factors influencing this perception? In this order, the aim of this research is twofold:

1. Examine the perceived acceptance and trust in a risk assessment training chatbot developed to assess risk and needs of juvenile offenders.
2. Examine the factors influencing students' perception of acceptance and trust.

Materials and methods

Participants and recruitment

Participants were all criminology students at a Canadian university. More precisely, participants were mostly female, between 20 and 25 years old and in their second year of criminology program. Recruitment of participants took place from January 2022 to April 2022, in an undergraduate course on risk assessment. As part of the course and separately from this study, 112 students were asked to complete questionnaires and a scoring exercise based on an interview with a simulated offender (chatbot). At the end of the course, all students were verbally solicited by the professor. All students were informed that participation was independent of any class credit or grade, and consent was requested after the final grade was delivered to students. All interested participants gave their written informed consent before entering the study. Ethical approval from the University of Montreal (#CERSC-2022-024-D) and CÉR-Jeunes en difficulté (#MP-CER-JD-20-19) was obtained.

Data collection procedures

Participants were invited to complete different online questionnaires and complete the risk assessment exercise using the

chatbot. In addition, participants answered a series of open-ended questions about strengths, limitations, difficulties encountered, recommendations for improvement and benefits from the chatbot exercise. All data were collected with LimeSurvey (Limesurvey GmbH, 2003). The risk assessment tool used to complete the exercise with the chatbot is the Youth Level of Service/Case Management Inventory (YLS/CMI). The YLS/CMI is one of the most widely used structured risk and need assessment measures across many countries. The validity of the YLSC/CMI is supported by several peer-reviewed and published studies conducted with different research groups (Catchpole and Gretton, 2003; Schmidt et al., 2005; Onifade et al., 2008; Rennie and Dolan, 2010; McGrath and Thompson, 2012; Takahashi et al., 2013; Campbell et al., 2014; Chu et al., 2015). The YLS/CMI is a standardized instrument that estimates the level of risk of recidivism by assessing the number of static and dynamic recidivism risk factors present in the lives of young offenders aged 12–18 (Hoge and Andrews, 2011). The YLS/CMI assesses the presence or absence of 42 factors that have been grouped into eight domains empirically related to re-offending: Prior and Current Offenses, Family Circumstances/Parenting, Education/Employment, Peer Relations, Substance Abuse, Leisure/Recreation, Personality/Behavior and Attitudes/Orientation. The YLS/CMI is the preferred instrument in this study, as it is widely used in Quebec.

The chatbot

Conversation engine

The chatbot software used in this study has been developed in collaboration with the National Research Council of Canada. The software is based on Rasa, an open-source framework, which leverages ML for building AI assistants and chatbots (Bocklisch et al., 2017). Rasa is based on two principles, namely Natural Language Understanding (NLU) and Dialogue Management. NLU (named Rasa NLU) extracts intents and entities from the user's messages, while Dialogue Management (named Rasa Core) leverages stories and rules to determine what the bot will do or say based on the user's message and context of the conversation (Bocklisch et al., 2017).

The chatbot software runs on a standard desktop or personal laptop computer. Communication with the chatbot can be done through voice leveraging a speech-to-text service, and via a text-based interface if necessary. In other words, participants would speak to the chatbot, then the user would review the text generated by the speech-to-text service before submitting it to the engine. The chatbot would answer vocally and with text. A text box of the conversation between the participant and the chatbot would also be generated for later feedback.

Chatbot development

The platform was developed with Unity 3D, a game development platform used to create and operate interactive, real-time 3D content (Unity Technologies, 2021). Character models were created with a universal framework called MakeHuman (MakeHuman Community, 2016). MakeHuman is an open-source tool for making 3D characters. The software offers more than 3,000 parameters to create highly detailed and unique characters: hair, skin, measurements, tooth shape, posture, etc.

In this specific study, the chatbot portrays a young adult on probation following a teenage sentence, and the chatbot appeared in a setting that resembled a traditional professional's office. To provide a realistic experience to users, the scenario (youth response) is based on a real young adult followed in a youth center in Quebec. We conducted interviews and asked him to answer questions generated in a previous data collection. We asked the participant to respond as naturally as possible. The interviews were filmed, and his voice was recorded. **Figure 1** shows the chatbot program interface.

Statistical analysis

The collected data was analyzed using SPSS statistical software version 25 (IBM Corp, 2017). The general characteristics of the participants were analyzed using frequency, percentages, means (M), and standard deviations (SD). Student's *t*-tests were conducted to compare the means of acceptance and trust in two age and gender group. To investigate the relationship between acceptance, trust and the participant's characteristics, Pearson's correlation coefficients were used for continuous variables and means comparisons for categorical variables. Multiple regression analysis was conducted to analyze the factors influencing the subject's trust and acceptance of the chatbot.

Measures

Sociodemographic Questionnaire: Participants were asked to complete a standard sociodemographic questionnaire. Sociodemographic information collected included age, gender, ethnicity, relationship status, education, type of graduate program, years in the program, present occupation, and desired occupation.

AES

To measure the acceptance of the chatbot we used the validated French version of the Acceptability E-scale (AES) (Tariman et al., 2011; Micoulaud-Franchi et al., 2016). This scale is a 6-item scale designed to assess usability (i.e., the perceived ease of using the system or app) and satisfaction (i.e., the perceived enjoyment of the use and usefulness of the system or app). All items were measured on a 1 (very unsatisfied) to 5 (very satisfied) Likert-like scale. Total scores can range from 6 to 30, with a higher score indicating higher acceptance. The internal consistency of this scale ranges between 0.70 and 0.76 (Tariman et al., 2011; Micoulaud-Franchi et al., 2016) which is similar to the internal consistency of 0.79 found in the current study. An example of an item for usability is "How easy was this computer program for you to use?" and an example for satisfaction is "How much did you enjoy using this computer program?" The original version of the AES has been validated with an English-speaking adult population being treated for various forms of cancer (Tariman et al., 2011). The French version of the scale was validated with a sample of 178 French-speaking patients having psychiatric or sleep complaints (Micoulaud-Franchi et al., 2016).

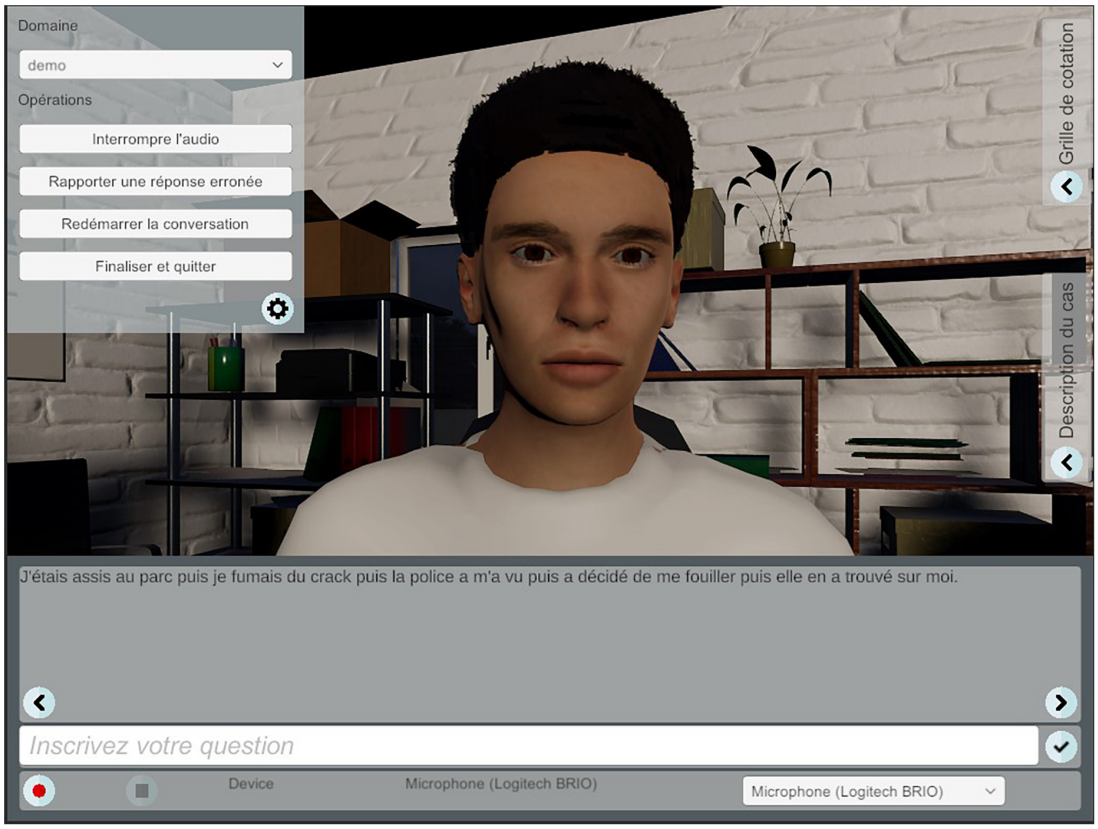


FIGURE 1
The chatbot program interface.

ETQ

To measure the students' perceived trust of the chatbot, we used the ECA Trust Questionnaire (ETQ) (Philip et al., 2020). This six-item French questionnaire is designed to assess users' trust in virtual agents based on two subdimensions: perceived credibility (ability and expertise of the virtual agent) and perceived benevolence (well-intentioned and accurately take the user's interests into account). All items were measured on a 1 (disagree strongly) to 5 (agree strongly) Likert-like scale. Total scores can range from 0 to 18, with a higher score indicating a more favorable attitude toward the agent. The internal consistency of this scale is 0.71 (Philip et al., 2020), while the internal consistency found in the present study is 0.30. An example of an item for perceived credibility is "Did you feel that the virtual agent was competent?" and an example for benevolence is "Did you feel that the interview with the virtual agent was pleasant?" The scale was validated with a sample of 318 patients suffering from various sleep disorders (Philip et al., 2020).

Mini-IPIP

To measure participants' personality traits we used the validated French version of the Mini-International Personality Item Pool (Mini-IPIP) (Donnellan et al., 2006; Laverdière et al., 2020). This 20-item scale is designed to evaluate personality traits according to the Big Five Model. Each of the five factors (neuroticism, extraversion, intellect, agreeableness, and conscientiousness) were assessed with four items, comprising a total scale that included 20 items. All items were measured on a 1 (very inaccurate) to 5 (very accurate) Likert-like scale. The internal consistency of this scale ranged between 0.64 and 0.81 (Donnellan et al., 2006; Laverdière et al., 2020) while the internal consistency found in the present study is 0.59. Example of item include "Have frequent mood swings" (neuroticism), "Talk to a lot of different people at parties" (extraversion), "Have a vivid imagination" (intellect), "Feel others' emotions" (agreeableness), and "Get chores done right away" (conscientiousness). The Mini-IPIP was validated with a sample of 139 French-Canadian psychology undergraduates (Laverdière et al., 2020).

Immersive tendencies questionnaire

To measure participants' immersive tendencies we used the validated French version of Immersive Tendencies Questionnaires (Witmer and Singer, 1998; Robillard et al., 2002). This 18-item scale is designed to assess the degree to which a participant may easily feel immersed and present in virtual environments. All items were measured on a 1 (never) to 7 (often) Likert-like scale. The scale is separated into four separate subscales related to four distinct tendencies of immersion: focus on current activities, involvement in activities, emotions, and tendency to play video games. Total scores can range from 18 to 126, with a higher score indicating

more immersive tendencies. The internal consistency of this scale is 0.78 (Robillard et al., 2002) while in the present study, the internal consistency found is 0.69. Example of items includes "Do you easily become deeply involved in movies or TV dramas?" (focus on current activities), "How frequently do you find yourself closely identifying with the characters in a story line?" (involvement in activities), "Do you ever have dreams that are so real that you feel disoriented when you awake?" (emotions) and "How often do you play arcade or video games?" (tendency to play video games). The French version of the scale was validated with a sample of 94 participants who were taking part in a virtual immersion activity (Robillard et al., 2002).

STAI-Y

To measure participants' anxiety we used the validated French version of the State-trait Anxiety Inventory (STAI-Y) (Spielberger, 1989; Gauthier and Bouchard, 1993). This 40-item scale is divided into two subscales, which measure state and trait anxiety. The state anxiety scale consists of 20 items (item 1 to item 20) that measure the respondent's feeling at that moment. The trait anxiety scale consists of 20 items (item 21 to item 40), and this scale measures how the respondent "generally" feels. Each item of the STAI-Y is rated on a scale of 1 (not at all) to 4 (very much so) in terms of intensity for state anxiety (not at all = 1, somewhat = 2, moderately so = 3, very much so = 4) and on a scale of 1 (almost never) to 4 (almost always) in terms of frequency for trait anxiety. Scores range from 20 to 80 per subscale, with a higher score indicating a higher degree of state and/or trait anxiety. The internal consistency of this scale ranges between 0.86 and 0.95 (Spielberger, 1989; Gauthier and Bouchard, 1993) and in the present study the internal consistency is 0.94. State anxiety items include "I am tense" while trait anxiety items include "I worry too much over something that really doesn't matter." The STAI-Y's English and Spanish version were validated with two samples: 38 Spanish-English teachers and teacher assistants and 31 English-education undergraduates from Puerto Rico (Spielberger, 1989). Its French version was validated with a sample of 83 psychology undergraduates from Laval University in Quebec (Gauthier and Bouchard, 1993).

LSQ-Fa

To measure participants' learning styles we used the abridged French version of the Learning Style Questionnaire (LSQ; Honey and Mumford, 1982; Fortin et al., 1997). This 48-item questionnaire is designed to assess preference for learning methods. Of the 48 items, there are 12 items for every learning style (active, reflector, theorist, and pragmatist). All items were measured on a 1 (totally disagree) to 7 (strongly agree) Likert-like scale. The total score for each learning style ranges between 12 and 84, with a higher score indicating a higher preference for the learning style. The internal consistency of this scale ranges between 0.86 and 0.95 (Fortin et al., 1997) and in the present study the internal consistency found is 0.85. Active style items include "I like to be the one who talks a lot," reflector style items include "I am careful not to jump to conclusions too quickly," theorist style items include "I like to be able to relate my actions to a general principle" and pragmatist style items include "In discussions, I like to get straight to the point." The French version of the LSQ has been validated with 205 university students in education (Fortin et al., 1997).

Self-efficacy questionnaire

Based on the available research (Schwarzer and Jerusalem, 1995; Delgadillo et al., 2014; Washburn et al., 2020), we developed a 12-item questionnaire to assess the sense of perceived self-efficacy within the use of the risk assessment tool. All items were measured on a 1 (disagree strongly) to 5 (agree strongly) Likert-like scale. Total scores can range from 12 to 60, with a higher score indicating

TABLE 1 Sociodemographic characteristics of participants.

Characteristics		
	(*N* = 112)	%
Gender		
Female	96	85.7%
Male	12	10.7%
Age		
Under 20	3	2.7%
20–25	85	75.9%
26–30	12	10.7%
31–35	4	3.6%
36–40	2	1.8%
Older than 41	2	1.8%
Highest level of education		
Diploma of vocational or college studies	62	55.4%
Certificate/Bachelor's degree	44	39.3%
Master's degree	1	0.9%
Actual level of education		
First year of bachelor's degree	4	3.6%
Second year of bachelor's degree	89	79.5%
Third year of bachelor's degree	10	8.9%
Master	1	0.9%
Discipline of actual education		
Criminology	104	92.9%
Independent studies	1	0.9%
Current occupation		
Full-time student	24	21.4%
Part-time student	2	1.8%
Full-time job	2	1.8%
Full-time student and full-time job	9	8.0%
Full-time student and part-time job	65	58.0%
Part-time student and full-time job	4	3.6%
Part-time student and part-time job	2	1.8%

10.3389/fpsyg.2023.1184016

TABLE 2 Distribution of satisfaction and usability subscales.

Items	Score					Descriptive statistics		
	Very unsatisfied	Unsatisfied	Neutral	Satisfied	Very satisfied	M (SD)	Mdn	Min-Max
Satisfaction								
How much did you enjoy using this chatbot?	1 (0.9%)	6 (5.4%)	17 (15.2%)	66 (58.9%)	22 (19.6%)	3.91 (0.80)	4	1–5
How useful was this chatbot to you in assessing the risk of recidivism?	0 (0%)	6 (5.4%)	20 (17.9%)	57 (50.9%)	29 (25.9%)	3.97 (0.81)	4	2–5
How would you rate your overall satisfaction with this chatbot?	2 (1.8%)	8 (7.1%)	30 (26.8%)	62 (55.4%)	10 (8.9%)	3.63 (0.82)	4	1–5
Usability								
How easy was this chatbot for you to use?	0 (0%)	12 (10.7%)	21 (18.8%)	68 (60.7%)	11 (9.8%)	3.70 (0.79)	4	2–5
How understandable were the answers provided by the chatbot?	1 (0.9%)	24 (21.4%)	23 (20.5%)	55 (49.1%)	9 (8%)	3.42 (0.95)	4	1–5
How acceptable is the time spent asking questions to this chatbot?	6 (5.4%)	26 (23.2%)	17 (15.2%)	36 (32.1%)	27 (24.1%)	3.46 (1.24)	4	1–5

TABLE 3 Distribution of benevolence and credibility subscales.

Items	Score					Descriptive statistics		
	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree	M (SD)	Mdn	Min-Max
Benevolence								
Did you feel that your questions were correctly understood by the chatbot?	6 (5.4%)	61 (54.5%)	18 (16.1%)	1 (0.9%)	26 (23.2%)	2.82 (1.30)	2	1–5
Did you feel that the answers provide by the chatbot were clear?	2 (1.8%)	21 (18.8%)	60 (53.6%)	6 (5.4%)	23 (20.5%)	3.24 (1.04)	3	1–5
Did you feel that the interview with the chatbot was pleasant?	2 (1.8%)	15 (13.4%)	54 (48.2%)	14 (12.5%)	27 (24.1%)	3.44 (1.05)	3	1–5
Credibility								
The chatbot should be integrated into training practices?	1 (0.9%)	4 (3.6%)	53 (47.3%)	39 (34.8%)	15 (13.4%)	3.56 (0.80)	3	1–5
The chatbot should obligatory be used in training?	5 (4.5%)	19 (17%)	33 (29.5%)	19 (17.0%)	36 (32.1%)	3.55 (1.23)	3	1–5
Did you feel that the chatbot was credible?	2 (1.8%)	9 (8%)	60 (53.6%)	18 (16.1%)	23 (20.5%)	3.46 (0.98)	3	1–5

more self-efficacy within the use of the risk assessment tool. To ensure the internal consistency of the scale, Cronbach's alpha coefficients were calculated. According to Cronbach's threshold, analyses showed good results ($\alpha = 0.85$). An example of items includes "In an interview, I know how to address the different themes included in the YLS/CMI."

Results

Characteristics of the participants

A total of 112 students were analyzed. Participant characteristics are summarized in Table 1. Results show that participants were mostly female (85.7%) between 20 and 25 years old (75.9%). The highest level of education was mostly a college-level diploma (55.4%) and, except for one, all of them were in a criminology program (92.9%), mostly in their second year (79.5%). Participant occupations were mostly full-time student and part-time job (58%).

Acceptance and trust perception with the chatbot

Acceptance

As shown in Table 2, results indicate that the overall system acceptance (satisfaction and usability subscales) was rated mostly positively by the participants, with more than half being "satisfied" or "very satisfied" with every item of the scale. Results show that median scores for all the items were 4 (satisfied), which means that half the scores are greater than or equal to "satisfied" and half are lower.

Concerning satisfaction, results indicate that most participants enjoyed using the chatbot, with 78.5% being "satisfied" or "very satisfied." Participants also found the chatbot useful for risk

assessment training, with 76.8% being either "satisfied" or "very satisfied" and 5.4% being "unsatisfied" and no one being "very unsatisfied." Overall, participants were mostly satisfied with the chatbot, with 64.3% being "satisfied" or "very satisfied." As for usability, results indicate that participants mostly found the chatbot easy to use, with over 70% being "satisfied" or "very satisfied" and 10.7% being "unsatisfied" and no one being "very unsatisfied." Results show that 57.1% of participants were "satisfied" or "very satisfied" with the answers provided by the chatbot during the exercise, while 22.3% were "unsatisfied" or "very unsatisfied." More than half of the participants also found that the time spent asking questions to the chatbot was acceptable, with 56.2% being "satisfied" or "very satisfied" and 28.6% were "unsatisfied" or "very unsatisfied."

According to comments made in the qualitative section of the questionnaire, the lower usability level in this study is likely due to technical issues that some participants experienced during the study. The first technical issue reported by participants is that the chatbot software was too resource intensive for their computer. For example, one participant stated that "The biggest difficulty I encountered was on the computer side. Indeed, after 5 min of use, my computer was overheating, so I had to quit and come back each time" [author's translation]. The second technical issue also reported by participants is that during the exercise they had to restart the conversation with the chatbot several times. One participant stated that:

> "After a few hours of consecutive use, the chatbot simply stopped answering my questions, even if I reset the conversation. So, I had to quit the application and restart it so that it would start answering again. It wasn't a big problem and didn't bother me much, but I just wanted to share it with you" [author's translation].

Trust

As shown in Table 3, results indicate that the overall system trust (benevolence and credibility subscales) was rated more positively than negatively by the participants. Except for the item "Did you feel that your questions were correctly understood by the chatbot," more than half responded that they were either neutral or agreed with all items. Results show that median scores for all the items, except for the one named above, were 3 (neutral), which means that half the scores are greater than or equal to "neutral" and half are lower.

Concerning benevolence, when asked if their questions were correctly understood by the chatbot, 59.9% of participants disagreed with this statement (disagree or strongly disagree). As for the answers provided by the chatbot, participants most often neither agreed nor disagreed (53.6%) with the clarity of the answers provided by the chatbot. Results also show that 48.2% of participants found the interview with the chatbot neither pleasant nor unpleasant, while 36.6% found that it was pleasant. As for credibility, almost half the participants agreed with the integration of the chatbot into training practices and with the mandatory integration at 48.2% and 49.1%, respectively. As for the credibility of the chatbot, 53.6% of participants neither agreed nor disagreed with it.

According to comments made in the qualitative section of the questionnaire, the lower trust levels in this study are likely due to logistic issues that participants experienced during the study. The first and main logistical issue reported by participants is that the chatbot did not understand several of their questions. One participant states that "The difficulty I encountered that stood out the most in my use of the chatbot was the fact that there were so many questions that led to an answer like 'I don't understand the question"' [author's translation]. Participants also indicate that because of this issue, the session was time-consuming. For example, one student said, "I felt like I spent more time trying to write questions that he understood rather than doing the scoring [of the YLS/CMI] itself" [author's translation]. Because of this, multiple participants also experienced frustration and anxiety. One participant said, "it can be frustrating to ask questions that you think are necessary for your rating and the chatbot just doesn't have the answer, no matter how you ask it" [author's translation]. Another participant stated that "[. . .] the fact that we were evaluated on the exercise made the whole thing very stressful and increased the frustration of the normal misunderstanding of the chatbot when faced with certain questions" [author's translation].

Factors associated with acceptance and trust

Age and gender

As presented in Tables 4, 5, results showed no significant relationship between the satisfaction, usability, benevolence, credibility subscales and age or gender.

Acceptance

As shown in Table 6, a series of Pearson's correlations were conducted to determine if there was any significant relationship between acceptance and diverse factors identified in research. Regarding satisfaction subscale of the AES, results showed no significant relationship with any variables. As for the usability

TABLE 4 Differences in acceptance and trust subscales scores between younger and older adult.

	Under 30 years old		Over 30 years old		*t*	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Acceptance–Satisfaction	3.814	0.682	3.833	0.690	−0.740	0.941	0.027
Acceptance–Usability	3.501	0.706	3.791	0.754	−1.111	0.269	0.397
Trust–Benevolence	3.168	0.639	3.166	0.816	0.007	0.994	0.002
Trust–Credibility	3.532	0.622	3.083	0.849	1.907	0.059	0.603

subscale of the AES, results showed moderate positive correlations with the theorist learning style of the LSQ-Fa ($r = 0.25$; $p < 0.05$) and the reflector learning style of the LSQ-Fa ($r = 0.23$; $p = 0.05$). Results also show a moderate negative correlation with the neuroticism dimension of the Mini-IPIP ($r = -0.22$; $p < 0.05$). Results show high positive correlations with the self-efficacy scale ($r = 0.32$; $p < 0.01$). Figure 2 presents the relationship between acceptance and these variables.

Trust

As shown in Table 6, a series of Pearson's correlations were also conducted to determine if there was any significant relationship between trust and diverse factors identified in research. Concerning the benevolence subscale of the ETQ, results showed no significant relationship with any variables. Regarding the credibility subscale of the ETQ, results showed a moderate positive correlation between the state anxiety dimension of the STAI-Y ($r = 0.22$; $p < 0.05$). Figure 2 also presents the relationship between trust and these variables.

As shown in Table 7, multivariate analyses were conducted. A significant regression equation was found only for the usability subscale [$F(16.24) = 1.951$, $p < 0.005$], with an R^2 of 0.33. Results show significant relationships between usability and state anxiety ($b = 0.43$; $p < 0.05$), self-efficacy ($b = 0.51$; $p < 0.01$), and the theorist learning style ($b = 0.45$; $p < 0.05$). Models predicting satisfaction, benevolence and credibility subscales were not significant (see Supplementary Tables 1–3). Figure 3 presents the relationship between usability and these variables.

Discussion

The objective of this study was twofold: (1) to examine the perceived acceptance and trust in a risk assessment training chatbot developed to assess risk and needs of juvenile offenders, and (2) to

TABLE 5 Differences in acceptance and trust subscales scores between female and male.

	Female		Male		*t*	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Acceptance–Satisfaction	3.854	0.652	3.484	0.848	−1.723	0.088	0.489
Acceptance–Usability	3.559	0.684	3.212	0.885	−1.543	0.126	0.438
Trust–Benevolence	3.194	0.629	2.939	0.800	−1.236	0.219	0.354
Trust–Credibility	3.524	0.631	3.272	0.771	−1.222	0.224	0.357

TABLE 6 Correlations between acceptance, trust, and independent variables.

	Acceptance–Satisfaction	Acceptance–Usability	Trust–Benevolence	Trust–Credibility
Mini-International Personality Item Pool (Mini-IPIP)				
Neuroticism	−0.00	−0.22*	−0.05	0.14
Extraversion	0.05	−0.01	0.11	0.08
Intellect	−0.03	−0.04	−0.08	−0.12
Agreeableness	0.07	0.02	−0.06	−0.09
Conscientiousness	0.18	0.10	0.01	0.06
Immersive Tendencies Questionnaire				
Focus on current activities	0.05	−0.07	−0.03	−0.05
Involvement in activities	0.06	−0.06	−0.10	−0.14
Emotions	0.08	0.19	−0.04	−0.01
Tendency to play video games	−0.05	−0.08	−0.11	−0.01
State-trait Anxiety Inventory (STAI-Y)				
Trait anxiety	−0.10	−0.18	0.00	0.15
State anxiety	0.03	−0.04	0.09	0.22*
Learning Style Questionnaire-Fa (LSQ-Fa)				
Active	0.10	−0.02	0.03	−0.02
Reflector	−0.04	0.23*	0.06	0.07
Theorist	0.09	0.25*	0.03	0.14
Pragmatist	0.07	0.14	0.07	−0.02
Self-efficacy questionnaire	0.19	0.32**	−0.02	−0.06

*$p < 0.05$; **$p < 0.01$.

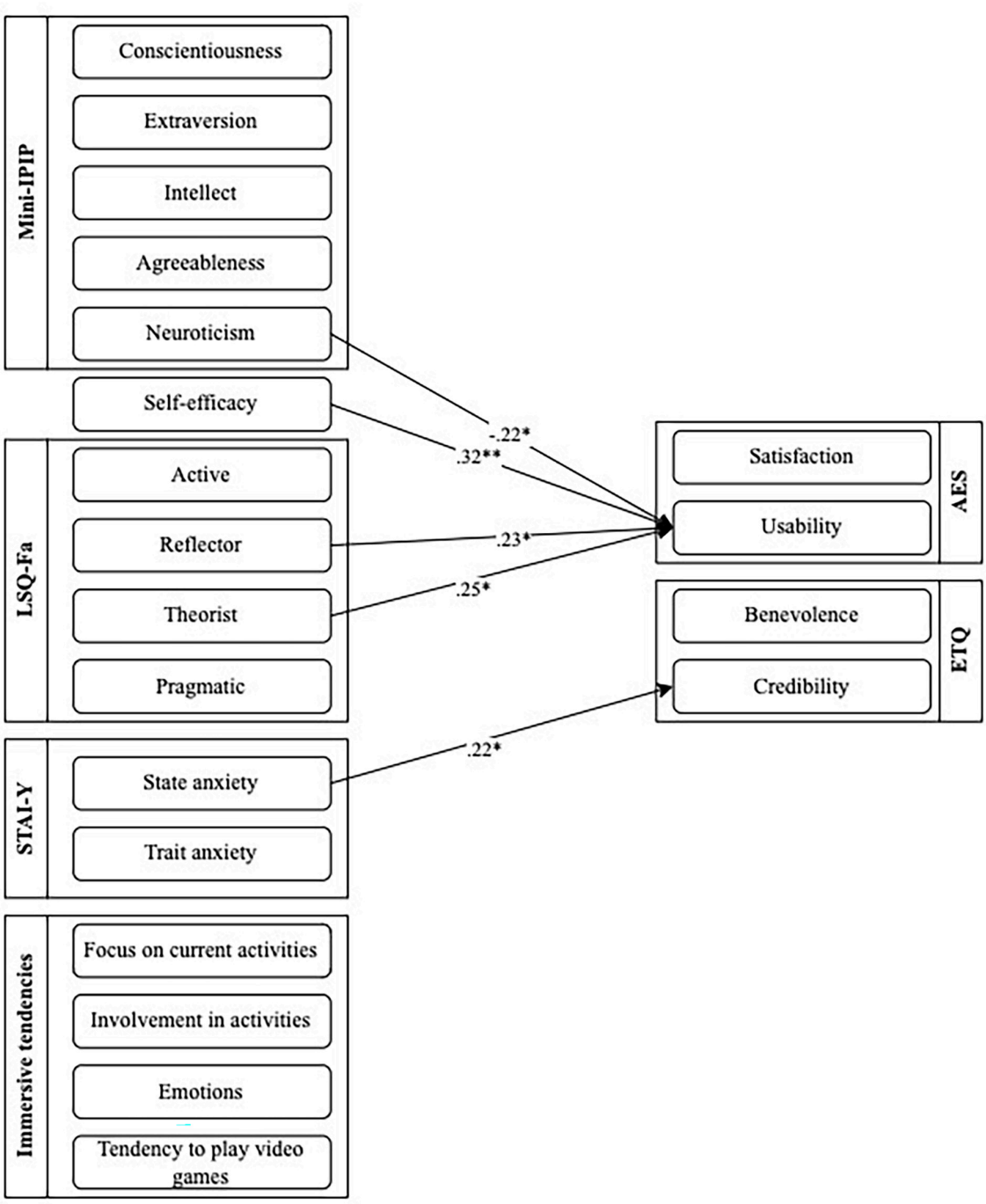


FIGURE 2
Correlation coefficients between user-related factors and acceptance and trust of the chatbot.

examine the factors influencing students' perception of acceptance and trust. These findings are very encouraging and suggest that the chatbot could be an effective educational method.

Acceptance and trust perception

Taken together the results of the present study show satisfactory level of acceptance and trust in chatbot. Overall, except for few technical and logistical limitations, the chatbot was functional and well-appreciated by participants. These findings are consistent with other studies that show satisfactory levels of acceptance involving chatbots (Philip et al., 2020; Richardson et al., 2021). In prior works, most respondents appeared to find chatbots acceptable and usable, but they also mentioned some technical issues that affected their experience. As pointed out by Richardson et al. (2021), even if most participants found the virtual software usable, all participants suggested some form of technological amendment to improve the user experience. As for trust, few studies have looked at the notion of trust, and available studies have focused on the notion of trust

10.3389/fpsyg.2023.1184016

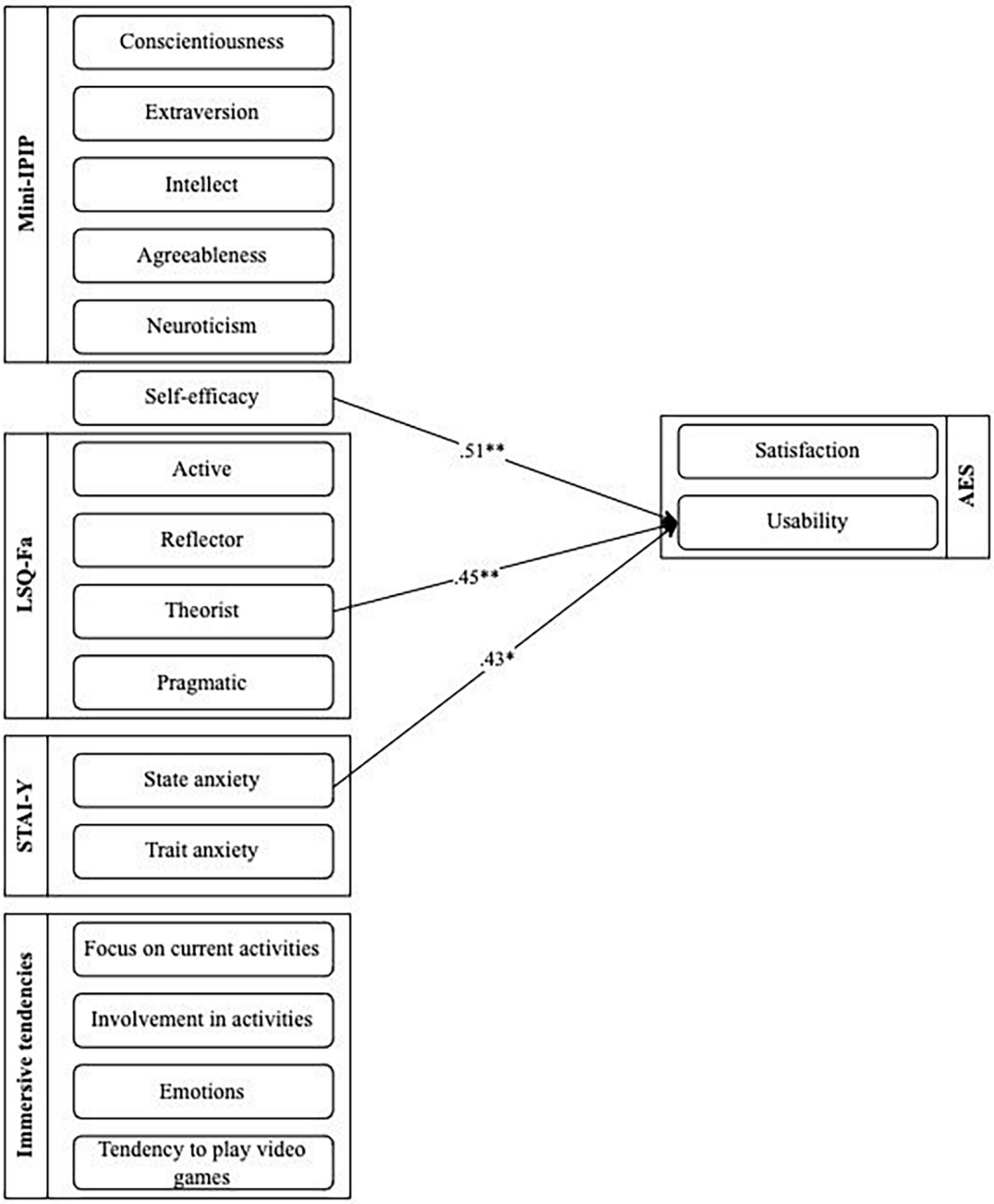


FIGURE 3
Linear regression coefficients between user-related and factors and perception of usability.

in a customer service context or in clinical interviews with patients (Philip et al., 2020).

Factors associated with acceptance and trust

Several factors associated with acceptance and trust have been identified in this study, but mostly with the usability dimension. Results show that learning style has an influence on chatbot acceptance. More precisely, results show that the theorist learning style has an influence on usability. Reflective and analytical people who like to understand the theories behind actions seem to find the chatbot easier to use. As theorists enjoy following models and reading up on facts to better engage in the learning process, those people may have taken the time to clearly understand the functioning of the chatbot and its features.

Results also demonstrate that self-efficacy has an influence on chatbot usability. People who have a higher sense of efficacy within the use of the risk assessment tool seem to find the chatbot easier to use. This result is consistent with previous research which suggests that self-efficacy has a potential impact on acceptance and trust

(Agarwal et al., 2000; Czaja et al., 2006). Participants with higher belief in their own capacity to achieve the risk assessment's exercise found the chatbot easier to use. As risk assessment is a complex task with important consequence on individual liberties and public security, future professionals must feel competent in their risk assessment. There is also some evidence supporting that computer self-efficacy (one's belief about his ability to perform a specific task using a computer) (Compeau and Higgins, 1995) has been shown to be a strong determinant of perceived ease of use before firsthand experience (Agarwal et al., 2000; Venkatesh, 2000). It will be interesting to integrate this specific type of self-efficacy in future work.

Results point out that state anxiety has an influence on chatbot usability. People who present situational anxiety like unpleasant feelings of tension and apprehension seem to find the chatbot easier to use. This result is consistent with the studies on trust and on computer anxiety. Indeed, as suggested by Müller et al. (2019), people who experience more anxiety tend to present a higher ability to trust chatbots. As credibility also measures the relevance of a chatbot in risk assessment training in this study, this result may suggest that the chatbot can be an effective training method for anxious students. Anxious people can find the chatbot credible but also easier to use because the chatbot allows students to put into practice the theory learned in a much more concrete way, which can reduce the level of anxiety during exercise. This allows students to practice their interview skills without the time constraints and pressure/stress of the task in the practical setting. In addition, chatbot programs offer students a space to safely try new approaches and new techniques without any consequences for real clients. We found consistent, but not statistically significant, differences between males and females for acceptance, usability, benevolence, and credibility. The results show slight differences between men and women, i.e., women seem to find the chatbot more acceptable and reliable than men. The results of this study are negatively impacted by the small sample size because small sample sizes significantly decrease statistical power and the flexibility of detecting any type of effect size (Heidel, 2016). In addition, there are few males in the sample since there were less men than women in the mandatory course on risk assessment. Future research should include a heterogeneous sample. The findings in the current study are consistent with recent research on technology acceptance.

We also found no significant differences between age for acceptance, usability, benevolence, and credibility. Some research highlighted differences between older and younger adults, whereas others did not (Grimes et al., 2010; Mitzner et al., 2010; McLean and Osei-Frimpong, 2019; Følstad and Brandtzaeg, 2020). Our result could be explained by age-homogeneous composition of the sample. Since most participants were between 20 and 25 years old, it is more difficult to compare groups.

TABLE 7 Linear regressions of factors associated with usability with the chatbot.

Independent variables	Coefficient	S.E.	Beta	*T*
Mini-International Personality Item Pool (Mini-IPIP)				
Neuroticism	−0.11	0.13	−0.14	−0.85
Extraversion	−0.05	0.13	−0.07	−0.42
Intellect	−0.02	0.13	−0.02	−0.18
Agreeableness	0.27	0.15	0.22	1.78
Conscientiousness	−0.06	0.11	−0.07	−0.52
Immersive Tendencies Questionnaire				
Focus on current activities	−0.03	0.09	−0.05	−0.38
Involvement in activities	−0.09	0.10	−0.11	−0.90
Emotions	0.13	0.10	0.17	1.30
Tendency to play video games	0.03	0.06	0.05	0.43
State-trait Anxiety Inventory (STAI-Y)				
Trait anxiety	−0.37	0.29	−0.25	−1.28
State anxiety	0.43*	0.20	0.34	2.12
Learning Style Questionnaire-Fa (LSQ-Fa)				
Active	−0.02	0.16	−0.02	−0.12
Reflector	−0.15	0.15	−0.15	−0.99
Theorist	0.45**	0.17	0.42	2.71
Pragmatist	−0.01	0.16	−0.01	−0.07
Self-efficacy questionnaire	0.51**	0.17	0.36	2.98

*$p < 0.05$; **$p < 0.01$.

Strengths and limitations

This study has several strengths and implications. First, to our knowledge, there are no chatbots with virtual avatar available that are useful for training professionals who assess the risk of recidivism of offenders. This is the first study to examine the perceived acceptance and trust in a risk assessment training chatbot. Considering the consequences that such assessments have on individual liberties and public security, developing effective and realistic training methods is warranted. Second, this study has also highlighted some factors that are associated with the acceptance and trust of a chatbot, such as self-efficacy, learning style and anxiety. This study provides a better understanding of the factors that facilitate user acceptance and trust of a chatbot, and a solution to the modifications needed for successful adoption. Future studies should examine those factors because such investigations may provide more comprehensive information regarding how to successfully integrate a chatbot into training programs.

This study also has several limitations. The first one is that the chatbot exercise was conducted as a mandatory exercise in a risk assessment course. Since this was a practical examination, it is possible that students answered and reacted differently. For example, they may have experienced more stress knowing that they were going to be graded following the exercise. The second limit is that the sample was homogeneous (i.e., age and gender), which makes it difficult to compare the groups. In addition, the generalizability of this study is limited by the lack of diversity in the sample. It would be interesting to examine chatbot's acceptance and trust with working professionals, who do not feel pressure to succeed and who represent a more heterogeneous group. The number and length of online questionnaires is the

third limit of this study. Prior work has found the length of questionnaires to affect response rate (Sahlqvist et al., 2011). The response rate may therefore have been affected by the number and length of questionnaires in this study. Finally, the reliability of the ECA Trust Questionnaire is also a limit, as it shows an extremely low internal consistency. Such a difference between the original reliability and our findings can be explained by the different population composing the different samples. The ECA Trust Questionnaire may not be an appropriate scale to use with students or professionals working in the risk assessment field. If this study was to be replicated, another measurement scale should be developed to evaluate trust.

Conclusion

The objective of the current study was to examine the perceived acceptance and trust in a risk assessment training chatbot developed to assess risk and needs of juvenile offenders and the factors influencing acceptance and trust. Results show a high level of acceptance in the chatbot. Participants were satisfied with their experience with the chatbot. Most users found the chatbot easy to use, even if they noted some technical issues, such as resource intensive software and conversation problems. As for trust in the chatbot, results show a satisfactory level. Participants found that the chatbot was benevolent, but numerous participants reported that the chatbot did not understand nor answered several of their questions. As for credibility, participants found the chatbot credible. They mentioned being in favor of integration into practice, but perhaps not as a mandatory evaluation.

Furthermore, results also suggest that acceptance and trust do not only depend on the design of the chatbot software, but may also vary depending on the characteristics of the user. Results suggested that self-efficacy, state anxiety and learning styles have an influence on the acceptance and trust of a chatbot, and especially on usability. Analytical individuals and anxious individuals seem to find the chatbot easier to use. Those who found the chatbot easier to use had higher belief in their own capacity to achieve the risk assessment's exercise.

As trust and acceptance play a vital role in determining technology success, these results are encouraging. Future studies are required to explore how several factors influence acceptance and trust in a risk assessment training chatbot. However, as reported by participants, some improvements need to be done prior to that. Since there are no such chatbots available for training professionals working in the fields of clinical psychology, psychiatry social work and criminology, these results are important as they tell us about the limitations of chatbots and the modifications that are needed.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Comité d'éthique de la recherche–Société et culture at the University of Montreal (#CERSC-2022-024-D) and Comité d'éthique de la recherche–Jeunes en difficulté (#MP-CER-JD-20-19). The patients/participants provided their written informed consent to participate in this study.

Author contributions

A-PR and J-PG contributed to the conception and design of the study. A-PR organized the database, performed the statistical analysis, and wrote the manuscript. All authors contributed to the manuscript revision and read and approved the submitted version.

Funding

This work was supported by the Social Sciences and Humanities Research Council (grant number 430-2019-00408).

Acknowledgments

We thank Violette Prignac and Arianne Chartrand for the help provided for this study. We also thank Anne Crocker and Denis Lafortune for the support provided for this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1184016/full#supplementary-material

References

Agarwal, R., Sambamurthy, V., and Stair, R. M. (2000). Research report: The evolving relationship between general and specific computer self-efficacy—An empirical assessment. *Inf. Syst. Res.* 11, 418–430. doi: 10.1287/isre.11.4.418.11876

Alaagib, N. A., Musa, O. A., and Saeed, A. M. (2019). Comparison of the effectiveness of lectures based on problems and traditional lectures in physiology teaching in Sudan. *BMC Med. Educ.* 19:365. doi: 10.1186/s12909-019-1799-0

Bateman, J., Allen, M. E., Kidd, J., Parsons, N., and Davies, D. (2012). Virtual patients design and its effect on clinical reasoning and student experience: A protocol for a randomised factorial multi-centre study. *BMC Med. Educ.* 12:62. doi: 10.1186/1472-6920-12-62

Bocklisch, T., Faulkner, J., Pawlowski, N., and Nichol, A. (2017). "Rasa: Open source language understanding and dialogue management," in *Proceedings of the 31st convention of the NIPS–conversational AI worskhop*, Long Beach, CA.

Brandtzaeg, P. B., and Følstad, A. (2017). "Why people use chatbots," in *Paper presented at the international conference on internet science*, Thessaloniki.

Brown, H. G., Poole, M. S., and Rodgers, T. L. (2004). Interpersonal traits, complementarity, and trust in virtual collaboration. *J. Manag. Inf. Syst.* 20, 115–138. doi: 10.1080/07421222.2004.11045785

Butler, J. A. (1992). Use of teaching methods within the lecture format. *Med. Teach.* 14, 11–25. doi: 10.3109/01421599209044010

Campbell, C., Onifade, E., Barnes, A., Peterson, J., Anderson, V., Davidson, W., et al. (2014). Screening offenders: The exploration of a youth level of service/case management inventory (YLS/CMI) brief screener. *J. Offender Rehabil.* 53, 19–34. doi: 10.1080/10509674.2013.861314

Catchpole, R. E. H., and Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Crim. Justice Behav.* 30, 688-708. doi: 10.1177/0093854803256455

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., and Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Rev. Educ. Res.* 90, 499–541. doi: 10.3102/0034654320933544

Chu, C. M., Lee, Y., Zeng, G., Yim, G., Tan, C. Y., Ang, Y., et al. (2015). Assessing youth offenders in a non-Western context: The predictive validity of the YLS/CMI ratings. *Psychol. Assess.* 27, 1013–1021. doi: 10.1037/a0038670

Clarizia, F., Colace, F., Lombardi, M., Pascale, F., and Santaniello, D. (2018). "Chatbot: An education support system for student," in *Lecture notes in computer science: Cyberspace safety and security*, Vol. 11161, eds A. Castiglione, F. Pop, M. Ficco, and F. Palmieri (Cham: Springer), 291–302. doi: 10.1007/978-3-030-01689-0_23

Coffield, F., Moseley, D., Hall, E., and Ecclestone, K. (2004). *Should we be using learning styles? What research has to say to practice*. London: Learning and Skills Research Centre.

Compeau, D. R., and Higgins, C. A. (1995). Application of social cognitive theory to training for computer skills. *Inf. Syst. Res.* 6, 118–143. doi: 10.1287/isre.6.2.118

Cook, D., Brydges, R., Zendejas, B., Hamstra, S., and Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. *Acad. Med.* 88, 872–883. doi: 10.1097/ACM.0b013e31828ffdcf

Cook, D., Erwin, P., and Triola, M. (2010). Computerized virtual patients in health professions education: A systematic review and meta-analysis. *Acad. Med. J. Assoc. Am. Med. Coll.* 85, 1589–1602. doi: 10.1097/ACM.0b013e3181edfe13

Coyne, E., Calleja, P., Forster, E., and Lin, F. (2021). A review of virtual-simulation for assessing healthcare students' clinical competency. *Nurse Educ. Today* 96:104623. doi: 10.1016/j.nedt.2020.104623

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., et al. (2006). Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychol. Aging* 21, 333–352. doi: 10.1037/0882-7974.21.2.333

Delgadillo, J., Moreea, O., Outhwaite-Luke, H., Dace, T., Nicholls, B., Ramseyer, G., et al. (2014). Confidence in the face of risk: The risk assessment and management self-efficacy study (RAMSES). *Psychiatr. Bull.* 38, 58–65. doi: 10.1192/pb.bp.112.040394

Donnellan, M. B., Oswald, F. L., Baird, B. M., and Lucas, R. E. (2006). The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of personality. *Psychol. Assess.* 18, 192–203. doi: 10.1037/1040-3590.18.2.19

Dunn, R., and Dunn, K. (1974). Learning style as a criterion for placement in alternative programs. *Phi Delta Kappan* 56, 275–278.

Entwistle, N. J. (1988). *Styles of learning and teaching: An integrated outline of educational psychology for students, teachers and lecturers*. London: David Fulton Publishers, doi: 10.4324/9781315067506

Felder, R. M. (1988). Learning and teaching styles in engineering education. *J. Eng. Educ.* 78, 674–681.

Felder, R. M. (1993). Reaching the second tier: Learning and teaching styles in college science education. *Coll. Sci. Teach.* 23, 286–290.

Felder, R. M., and Brent, R. (2005). Understanding student differences. *J. Eng. Educ.* 94, 57–72. doi: 10.1002/j.2168-9830.2005.tb00829.x

Felder, R. M., and Silverman, L. V. (1988). Learning and teaching styles in engineering education. *Eng. Educ.* 78, 674–681.

Fernández-Ballesteros, R., Bartram, D., Caprara, G. V., Hambleton, R. K., Hornke, L. F., ter Laak, J. J. F., et al. (2003). *Encyclopedia of psychological assessment*. Thousand Oaks, CA: SAGE Publications Inc.

Følstad, A., and Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Qual. User Exp.* 5:3. doi: 10.1007/s41233-020-00033-2

Fortin, G., Chevrier, J., and Amyot, É. (1997). Adaptation française du Learning Styles Questionnaire de Honey et Mumford. *Mes. Éval. Éduc.* 19, 95–118. doi: 10.7202/1091396ar

Frambach, R. T., Schillewaert, N., Ahearne, M. J., and Moenaert, R. K. (2000). The acceptance of information technology in the sales force. *J. Mark.* 34, 323–336.

Gauthier, J., and Bouchard, S. (1993). A French-Canadian adaptation of the revised version of Spielberger's State-Trait Anxiety Inventory. *Can. J. Behav. Sci. Rev. Can. Sci. Comportement* 25, 559–578. doi: 10.1037/h0078881

Grimes, G., Hough, M., Mazur, E., and Signorella, M. (2010). Older adults' knowledge of internet hazards. *Educ. Gerontol.* 36, 173–192.

Guay, J. P., Da Silva Guerreiro, J., and Crocker, A. G. (2022). Les méthodes et enjeux relatifs à l'évaluation du risque de la violence hétérodirigée [methods and current issues related to risk assessment of other-directed violence]. *Santé Ment. Québec* 47, 63–85. doi: 10.7202/1094145ar

Heidel, R. E. (2016). Causality in statistical power: Isomorphic properties of measurement, research design, effect size, and sample size. *Scientifica* 2016:8920418. doi: 10.1155/2016/8920418

Hoff, K. A., and Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Hum. Factors* 57, 407–434. doi: 10.1177/0018720814547570

Hoge, R. D., and Andrews, D. A. (2011). *Youth level of service/case management inventory 2.0 (YLS/CMI 2.0): User's manual*. Toronto, ON: Multi-Health Systems.

Honey, P., and Mumford, A. (1982). *The manual of learning styles*. Maidenhead: Peter Honey.

Hornbæk, K., and Hertzum, M. (2017). Technology acceptance and user experience: A review of the experiential component in HCI. *ACM Trans. Comput. Hum. Interact.* 24, 1–30. doi: 10.1145/3127358

IBM Cloud Education (2020a). *Machine learning*. IBM. Available online at: https://www.ibm.com/cloud/learn/machine-learning (accessed July 6, 2022).

IBM Cloud Education (2020b). *Natural language processing (NLP)*. IBM. Available online at: https://www.ibm.com/cloud/learn/natural-language-processing (accessed October 24, 2022).

IBM Corp (2017). *IBM SPSS statistics for windows, Version 25.0*. Armonk, NY: IBM Corp.

Islam, J., Rahman, Z., and Hollebeek, L. (2017). Personality factors as predictors of online consumer engagment: An empirical investigation. *Mark. Intell. Plan.* 35, 510–528. doi: 10.1108/MIP-10-2016-0193

Johns, C., Nuñez, D., Daya, M., Sellars, D., Casanueva, J., and Blake, E. (2000). "The interaction between individuals' immersive tendencies and the sensation of presence in a virtual environment," in *Virtual environments 2000*, eds J. Mulder and R. van Liere (Vienna: Springer), 65–74. doi: 10.1007/978-3-7091-6785-4_8

Kenny, P., Parsons, T. D., Gratch, J., Leuski, A., and Rizzo, A. A. (2007). "Virtual patients for clinical therapist skills training," in *Lecture notes in computer science. Intelligent virtual agents*, Vol. 4722, eds C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé (New York, NY: Springer), 197–210. doi: 10.1007/978-3-540-74997-4_19

Kenny, P., Parsons, T. D., Gratch, J., and Rizzo, A. A. (2008). "Evaluation of Justina: A virtual patient with PTSD," in *Intelligent virtual agents*, eds H. Prendinger, J. Lester, and M. Ishizuka (Berlin: Springer), 394–408. doi: 10.1007/978-3-540-85483-8_40

Kolb, A. Y., and Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Acad. Manag. Learn. Educ.* 4, 193–212. doi: 10.5465/amle.2005.17268566

Lateef, F. (2010). Simulation-based learning: Just like the real thing. *J. Emerg. Trauma Shock* 3, 348–352. doi: 10.4103/0974-2700.70743

Laverdière, O., Gamache, D., Morin, A. J. S., and Diguer, L. (2020). French adaptation of the Mini-IPIP: A short measure of the Big Five. *Eur. Rev. Appl. Psychol.* 70:100512. doi: 10.1016/j.erap.2019.100512

Limesurvey GmbH (2003). *LimeSurvey: An open source survey tool [Software]*. LimeSurvey GmbH. Available online at: http://www.limesurvey.org (accessed January 4, 2022).

Ling, E. C., Tussyadiah, I., Tuomi, A., Stienmetz, J., and Ioannou, A. (2021). Factors influencing users' adoption and use of conversational agents: A systematic review. *Psychol. Mark.* 38, 1031–1051. doi: 10.1002/mar.21491

MakeHuman Community (2016). *Documentation: What is MakeHuman?*. Available online at: http://www.makehumancommunity.org/wiki/Documentation:What_is_MakeHuman%3F (accessed July 7, 2022).

Mäurer, D., and Weihe, K. (2015). "Benjamin Franklin's decision method is acceptable and helpful with a conversational agent," in *Smart innovation, systems and technologies: Intelligent interactive multimedia systems and services*, Vol. 40, eds E. Damiani, R. Howlett, L. Jain, L. Gallo, and G. D. Pietro (Cham: Springer), 109–120. doi: 10.1007/978-3-319-19830-9_10

Mazmanian, P. E., and Davis, D. A. (2002). Continuing medical education and the physician as a learner. *JAMA* 288, 1057–1060. doi: 10.1001/jama.288.9.1057

McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H., and Wayne, D. B. (2011). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad. Med.* 86, 706–711. doi: 10.1097/ACM.0b013e318217e119

McGrath, A., and Thompson, A. P. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Crim. Justice Behav.* 39, 250–263.

McKnight, D., Choudhury, V., and Kacmar, C. (2002). Developing and validating trust measures for e-Commerce: An integrative typology. *Inf. Syst. Res.* 13, 334–359. doi: 10.1287/isre.13.3.334.81

McLean, G., and Osei-Frimpong, K. (2019). Chat now…Examining the variables influencing the use of online live chat. *Technol. Forecast. Soc. Change* 146, 55–67. doi: 10.1016/j.techfore.2019.05.017

McTear, M. (2020). *Conversational AI: Dialogue systems, conversational agents, and chatbots*. Williston, VT: Springer. doi: 10.2200/S01060ED1V01Y202010HLT048

Meyers, C., and Jones, T. B. (1993). *Promoting active learning: Strategies for the college classroom*. San Francisco, CA: Jossey-Bass Inc.

Micoulaud-Franchi, J. A., Sauteraud, A., Olive, J., Sagaspe, P., Bioulac, S., and Philip, P. (2016). Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. *Psychiatry Res.* 237, 196–200. doi: 10.1016/j.psychres.2016.01.043

Milis, K., Wessa, P., Poelmans, S., Doom, C., and Bloemen, E. (2008). *The impact of gender on the acceptance of virtual learning environments*. Available online at: https://www.researchgate.net/profile/Patrick_Wessa/publication/228641817_The_impact_of_gender_on_the_acceptance_of_virtual_learning_environments/links/00b495227866c2f4e1000000.pdf

Mitzner, T. L., Boron, J. B., Fausset, C. B., Adams, A. E., Charness, N., Czaja, S. J., et al. (2010). Older adults talk technology: Technology usage and attitudes. *Comput. Hum. Behav.* 26, 1710–1721. doi: 10.1016/j.chb.2010.06.020

Mooradian, J. K. (2008). Using simulated sessions to enhance clinical social work education. *J. Soc. Work Educ.* 44, 21–36. doi: 10.5175/JSWE.2008.200700026

Müller, L., Mattke, J., Maier, C., Weitzel, T., and Graser, H. (2019). "Chatbot acceptance: A Latent Profile Analysis on individuals' trust in conversational agents," in *Paper presented at the computers and people research conference*, Nashville, TN.

Nunez, D. (2003). *A connectionist explanation of presence in virtual environments*. Master's thesis. Cape Town: University of Cape Town.

Okonkwo, C. W., and Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Comput. Educ. Artif. Intell.* 2:100033. doi: 10.1016/j.caeai.2021.100033

Onifade, E., Davidson, W., Campbell, C., Turke, G., Malinowski, J., and Turner, K. (2008). Predicting recidivism in probationers with the youth level of service case management inventory (YLS/CMI). *Crim. Justice Behav.* 35, 474–483.

Othman, N., and Amiruddin, M. H. (2010). Different perspectives of learning styles from VARK model. *Procedia Soc. Behav. Sci.* 7, 652–660. doi: 10.1016/j.sbspro.2010.10.088

Padilla-Meléndez, A., Garrido-Moreno, A., and Del Aguila-Obra, A. R. (2008). Factors affecting e-collaboration technology use among management students. *Comput. Educ.* 51, 609–623. doi: 10.1016/j.compedu.2007.06.013

Philip, P., Dupuy, L., Auriacombe, M., Serre, F., de Sevin, E., Sauteraud, A., et al. (2020). Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. *NPJ Digit. Med.* 3:2. doi: 10.1038/s41746-019-0213-y

Reid, J. M. (1995). *Learning styles in the ESL/EFL classroom*. Boston, MA: Heinle & Heinle Publishers.

Rennie, C., and Dolan, M. (2010). Predictive validity of the youth level of service/case management inventory in custody sample in England. *J. Forensic Psychiatry Psychol.* 21, 407–425.

Richad, R., Vivensius, G., Sfenrianto, S., and Kaburuan, E. (2019). Analysis of factors influencing millennial's technology acceptance of chatbot in the banking industry in Indonesia. *Int. J. Civil Eng. Technol.* 10, 107–118. doi: 10.34218/IJM.10.3.2019.011

Richardson, C. L., Chapman, S., and White, S. (2021). Establishing the acceptability and usability of an animated virtual patient simulation. *Explor. Res. Clin. Soc. Pharm.* 4:100069. doi: 10.1016/j.rcsop.2021.100069

Rizzo, A. A., and Talbot, T. (2016). "Virtual reality standardized patients for clinical training," in *The digital patient: Advancing healthcare, research, and education*, eds C. D. Combs, J. A. Sokolowski, and C. M. Banks (New York, NY: Wiley), 257–272. doi: 10.1002/9781118952788.ch1

Robillard, G., Bouchard, S., Renaud, P., and Cournoyer, L. G. (2002). "Validation canadienne-française de deux mesures importantes en réalité virtuelle : l'Immersive Tendencies Questionnaire et le Presence Questionnaire," in *Paper presented at the société québécoise pour la recherche en psychologie (SQRP)*, Trois-Rivières, QC.

Robotham, D. (1995). Self-directed learning the ultimate learning style? *J. Eur. Ind. Train.* 19, 3–7. doi: 10.1108/03090599510092918

Roos, S. (2018). *Chatbots in education: A passing trend or a valuable pedagogical tool?* Doctoral dissertation. Uppsala: Uppsala University.

Sahlqvist, S., Song, Y., Bull, F., Adams, E., Preston, J., Ogilvie, D., et al. (2011). Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: Randomised controlled trial. *BMC Med. Res. Methodol.* 11:62. doi: 10.1186/1471-2288-11-62

Schmidt, F., Hoge, R. D., and Gomes, L. (2005). Reliability and validity analyses of the Youth Level of Service/Case Management Inventory. *Crim. Justice Behav.* 32:329344. doi: 10.1177/0093854804274373

Schmitt, S. (n.d.). *AI Chatbot: Ready for a conversational experience?* Liestal: Oneinside. Available online at: https://one-inside.com/ai-chatbot-conversational-experience/ (accessed July 6, 2022).

Schwarzer, R., and Jerusalem, M. (1995). "Generalized self-efficacy Scale," in *Measures in health psychology: A user's portfolio. Causal and control beliefs*, eds J. Weinman, S. Wright, and M. Johnston (Windsor: NFER-NELSON), 35–37.

Spielberger, C. D. (1989). *State-trait anxiety inventory: Bibliography*, 2nd, Edn. Palo Alto, CA: Consulting Psychologists Press.

Suri, G., and Sharma, S. (2013). "The impact of gender on attitude towards computer technology and e-learning: An exploratory study of Punjab University, India," in *Paper presented at the education, management and technology*, Mohali.

Takahashi, M., Mori, T., and Kroner, D. G. (2013). A cross-validation of the Youth Level of Service/Case Management Inventory (YLS/CMI) among Japanese juvenile offenders. *Law Hum. Behav.* 37, 389–400. doi: 10.1037/lhb0000029

Tariman, J. D., Berry, D. L., Halpenny, B., Wolpin, S., and Schepp, K. (2011). Validation and testing of the Acceptability E-scale for web-based patient-reported outcomes in cancer care. *Appl. Nurs. Res.* 24, 53–58. doi: 10.1016/j.apnr.2009.04.003

Thompson, S. H., and Lim, V. K. G. (1996). Factors influencing personal computer usage: The gender gap. *Women Manag. Rev.* 11, 18–26. doi: 10.1108/09649429610148746

Triola, M., Feldman, H., Kalet, A. L., Zabar, S., Kachur, E. K., Gillespie, C., et al. (2006). A randomized trial of teaching clinical skills using virtual and live standardized patients. *J. Gen. Internal Med.* 21, 424–429. doi: 10.1111/j.1525-1497.2006.00421.x

Unity Technologies (2021). *Unity [computer software]*. San Francisco, CA: Unity Technologies. Available online at: https://unity.com/

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., and Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *Can. J. Psychiatry* 64, 456–464. doi: 10.1177/0706743719828977

van Braak, J. P. (2004). Domains and determinants of university students' self-perceived computer competence. *Comput. Educ.* 43, 299–312. doi: 10.1016/j.compedu.2003.09.006

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology Acceptance model. *Inf. Syst. Res.* 11, 342–365. doi: 10.1287/isre.11.4.342.11872

Veretskaya, O. (2017). *What is a chatbot and how to use it for your business*. ANADEA. Available online at: https://anadea.info/blog/what-is-a-chatbot-and-how-to-use-it-for-business

Vita, G. D. (2001). Learning styles, culture and inclusive instruction in the multicultural classroom: A business and management perspective. *Innov. Educ. Teach. Int.* 38, 165–174. doi: 10.1080/14703290110035437

von der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S.-H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Comput. Hum. Behav.* 26, 1641–1650. doi: 10.1016/j.chb.2010.06.012

Washburn, M., and Zhou, S. (2018). Teaching note—Technology-enhanced clinical simulations: Tools for practicing clinical skills in online social work programs. *J. Soc. Work Educ.* 54, 554–560. doi: 10.1080/10437797.2017.1404519

Washburn, M., Bordnick, P., and Rizzo, A. S. (2016). A pilot feasibility study of virtual patient simulation to enhance social work students' brief mental health assessment skills. *Soc. Work Health Care* 55, 675–693. doi: 10.1080/00981389.2016.1210715

Washburn, M., Parrish, D. E., and Bordnick, P. S. (2020). Virtual patient simulations for brief assessment of mental health disorders in integrated care settings. *Soc. Work Ment. Health* 18, 121–148. doi: 10.1080/15332985.2017.1336743

Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence Teleoperators Virtual Environ.* 7, 225–240. doi: 10.1162/105474698565686

Wolpin, S., Berry, D., Austin-Seymour, M., Bush, N., Fann, J. R., Halpenny, B., et al. (2008). Acceptability of an electronic self-report assessment program for patients with cancer. *Comput. Inf. Nurs.* 26, 332–338. doi: 10.1097/01.NCN.0000336464.79692.6a

Young, J. E., Hawkins, R., Sharlin, E., and Igarashi, T. (2008). Toward acceptable domestic robots: Applying insights from social psychology. *Int. J. Soc. Robot.* 1:95. doi: 10.1007/s12369-008-0006-y

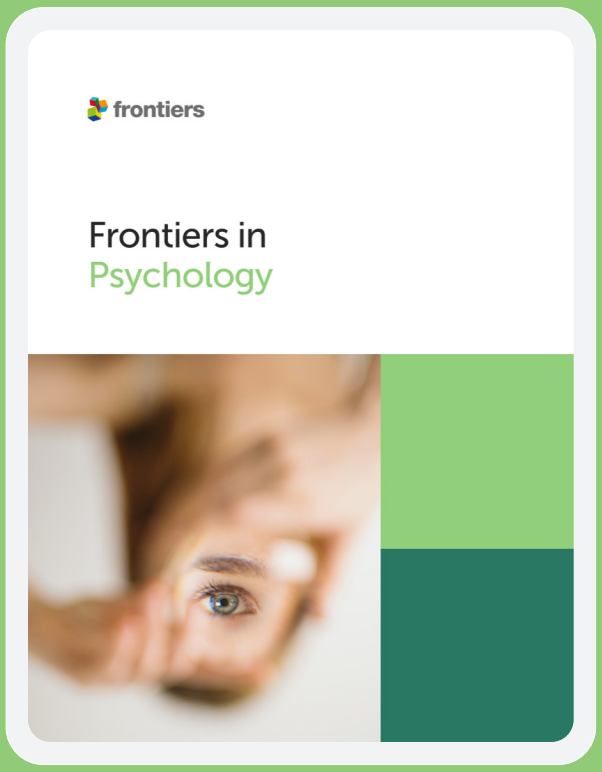
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact