



Feasibility of Remote Performance Assessment Using the Free Research Executive Evaluation Test Battery in Adolescents

Isis Angelica Segura[†] and Sabine Pompéia^{*†}

Departamento de Psicobiologia, Universidade Federal de São Paulo, São Paulo, Brazil

OPEN ACCESS

Edited by:

Natasha Kirkham,
Birkbeck, University of London,
United Kingdom

Reviewed by:

Cynthia Whissell,
Laurentian University, Canada
Bhoomika Kar,
Allahabad University, India

*Correspondence:

Sabine Pompéia
spompéia@unifesp.br

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 10 June 2021

Accepted: 20 September 2021

Published: 14 October 2021

Citation:

Segura IA and Pompéia S (2021)
Feasibility of Remote Performance
Assessment Using the Free Research
Executive Evaluation Test Battery in
Adolescents.
Front. Psychol. 12:723063.
doi: 10.3389/fpsyg.2021.723063

Lockdowns and other preventive measures taken to curb the spread of diseases such as COVID-19 have restricted the use of face-to-face cognitive assessment. Remote testing may be an alternative, but it should first be shown to be comparable to in-person assessment before being used more widely, during and after the pandemic. Our aim was to evaluate the suitability of online, examiner-mediated administration of an open-access battery of executive function tests (the Free Research Executive Evaluation battery, or FREE) that can be adapted considering various characteristics of diverse populations and therefore used worldwide. A total of 96 9–15-year olds (42 girls) were tested, half of whom online through video calls mediated by an examiner. Their performance was compared to that of the other 48 individuals tested face-to-face, who were matched against the online-tested participants for age, pubertal status, sex, and parental schooling. The battery consists of two tests of the following executive domains: Updating (2-Back and Number Memory tests), Inhibition (Stroop Victoria and Stroop Happy-Sad), and Switching (Color Shape and Category Switch). Answers were vocal and self-paced, and the examiner recorded accuracy and time taken to complete in-person and online tasks. Only free software is needed for the assessment. Executive measures obtained from the tasks did not differ statistically between online and in-person tested participants and effects sizes of group effects were small, thus showing that the FREE test battery holds promise for online cognitive assessment, pending confirmation in different samples and further validation studies.

Keywords: adolescents, executive functions, COVID-19, online testing, updating, inhibition, shifting

INTRODUCTION

In the context of the COVID-19 pandemic's social distancing, researchers interested in cognition have looked at the feasibility of remote cognitive testing. Perhaps surprisingly, there is a substantial body of evidence from the last two decades showing that online cognitive assessment may be equivalent to lab-based, face-to-face testing (Krantz and Dalal, 2000; McGraw et al., 2000; Nosek et al., 2002; Temple et al., 2010; Soto et al., 2011; Germine et al., 2012; Cullum et al., 2014) in elderly (e.g., Geddes et al., 2020), adult (e.g., Kirkwood et al., 2000), and

pediatric populations (e.g., Hodge et al., 2019; Worhach et al., 2021). Although the validity and reliability of remote assessment has been questioned due to difficulties in controlling stimuli presentation and measuring response, both in terms of accuracy and reaction times (Germiné et al., 2012), remote testing has unprecedented advantages that make it worth pursuing. As long as testees have internet access, these advantages include: (1) less travel time and expense, as well as lower implementation costs (Reips, 2000); (2) the possibility of reaching more diverse and less accessible samples, such as those from remote countries or areas, and/or patients with clinical conditions such as reduced mobility and/or higher vulnerability to diseases like COVID-19; and (3) testing people in familiar settings (their homes) instead of unknown locations, which has been shown to improve performance on some types of tasks (Sucksmith et al., 2013).

These advantages extend to testing BAME (Black, Asian and other non-white minority ethnic backgrounds) and non-WEIRD (Western Educated Industrialized Rich and Democratic) populations with internet access. These minorities or minoritized individuals have been under-represented in the cognitive literature in general, despite being more representative of humanity as a whole (see Henrich, 2010; Rad et al., 2018). Cultural, socioeconomic and ethnic differences affect not only the cognitive processes one might expect (e.g., social cognition and moral judgement) but also abilities such as visual perception, memory, categorization, attention, and executive functions (EF) (Henrich, 2010; Kelkar, 2013; Hackman and Gallop, 2015). EF is an umbrella-term for top-down cognitive functions that regulate an individual's behavior and emotions in order to achieve goals that are in peoples' minds (in their working memory) (Baggetta and Alexander, 2016; Friedman and Miyake, 2017). These behavioral self-regulation abilities have been found to be affected by many factors related to developing nations' poor, minority and minoritized groups (Moffitt et al., 2011), which have been hit harder by the pandemic, likely a longer-lasting threat for them due to a host of environmental factors (see Silva and Ribeiro, 2021).

In this scenario, remote EF cognitive testing must ensure health and safety of testees and examiners and also allow administration of cognitive tests that are adaptable to different cultural and socioeconomic contexts so as to more reliably capture the cognitive constructs under investigation (Fernández and Abe, 2018). Bearing this in mind, we investigated the adequacy of remote EF assessment mediated by examiners using a test battery designed to be adaptable to different contexts and populations (FREE: Free Research Executive Function Evaluation; Zanini, 2021).

The FREE test battery includes tasks that measure three types of executive functions that are interrelated, yet separable, based on a theoretical framework called the Unity and Diversity of Executive Functions (see Friedman and Miyake, 2017). These types of EF are inhibition of automatic responses, shifting between tasks, and updating information held in working memory. Importantly, these tasks were adapted for affordable testing by researchers using basic equipment. Task presentation and scoring are not automated. Testees themselves regulate the speed at which they can do tasks and respond vocally

while the examiner measures accuracy and time taken to complete blocks of trials instead of each individual trial. This is important because many studies have shown that limiting exposure and response times and requiring key presses for verbal answers can negatively influence measurements of EF in samples that include participants with different characteristics, such as various ages, and who are from diverse backgrounds (see Zanini, 2021). This test battery may therefore be administered remotely and be moderated by an online examiner, using screen sharing services that may be downloaded and used by examiners and testees free of charge without any special hardware, software downloads or plug-ins.

Here, the performance of adolescents tested online was compared with that of adolescents tested face-to-face in their own schools. Investigating EF is especially important during this phase of life because these cognitive skills develop during this period becoming differentiable in three distinct domains (see Lee et al., 2013), so factors that affect the environment and health at this age can potentially impact the development of EF performance, which influences a wide range of outcomes such as physical and mental well-being, academic and financial success, criminal and addictive behavior (see Moffitt et al., 2011). Hence, EF assessment of populations that include this vulnerable age is important. Additionally, it should be considered that many other factors reduce the possibility that people will be available for face-to-face testing. These include not only pandemics but also having debilitating illnesses that limit locomotion or the immune system, living in isolated locations, or others variables associated with poverty (e.g., shortage of means to pay for transport to and from research laboratories, or not having guardians who are available to accompany minors in person, which is often necessary). All of these conditions can also potentially impair EF, especially during sensitive phases of development like adolescence (Moffitt et al., 2011), which may go unnoticed if remote testing is not possible. Hunersen et al. (2021) call for the need to increase remote data collection strategies for testing adolescents, especially those from low-income settings using free tools, as proposed here.

The test battery used here was built for research purposes and group comparisons, not diagnostic evaluation, so norms are not available. For this reason, we matched participants tested online to others tested in person according to factors that are known to potentially affect brain and cognitive development (parental schooling, age, pubertal status, and sex: Foulkes and Blakemore, 2018). With the present study, we aim to establish whether applying the FREE battery is feasible under supervised online testing, to describe how the online testing was implemented and to compare the pattern of effects in both in-person and face-to-face conditions. Although completely remote, the supervised online presentation preserve some important aspects of in-person assessment that might be sources of bias if absent from online tests. These procedures included: (1) testee-examiner interaction throughout testing to prevent distraction from task objectives and misinterpreted instructions (Feenstra et al., 2017); (2) same format of stimuli presentation (PDF viewing of instructions and stimuli) that are seen (shared

screen when online) concomitantly by the testees and examiner; (3) same type of response to the tests (vocal, in which the examiner writes down the responses); and (4) self-paced format (clicking a mouse or tapping a keyboard to progress to the next stimulus during online testing and swiping the screen for in-person testing); the examiner used a stopwatch to time how long testees took to complete each task and wrote down the answers. Because the procedures were essentially the same except for being administered in person or online, we hypothesized that performance would be equivalent.

MATERIALS AND METHODS

Participants

Our convenience sample consisted of 96 native Portuguese-speaking, typically developing adolescents aged 9–16, of whom 48 were tested online. These individuals were matched to 48 adolescents who were evaluated in person at their schools (see matched pairing details below). Exclusion criteria were: (1) having been held back for a year or more at school; (2) being a student with special needs, which may be associated with clinical or cognitive limitations; and (3) taking daily medication to exclude any presence of chronic clinical disorders that could affect cognitive and/or developmental outcomes.

Procedures

This study was approved by the local Ethics Committee (CAAE # 56284216.7.0000.5505). Prior to the testing sessions, participants' guardians provided signed informed consent. Informed participant assent was always confirmed before test administration. Participants answered a socio-demographic questionnaire, self-evaluated their pubertal status filling in the 5-item Pubertal Development Scale [PDS, adapted from Carskadon et al., 1993 into Portuguese by Pompéia, 2019] and were administered the FREE executive function tasks in four pseudorandom orders to avoid the effects of fatigue. Participants from the supervised online group (hereafter called as "online group") were recruited through contact from their schools and social media after the authorities introduced social distancing and closed schools due to the COVID-19 pandemic. They were tested between 3 and 9 months after lockdown. All online participants were individually paired/matched to in-person tested individuals recruited at their schools and assessed the year before the pandemic broke out (some of whom were part of a prior study: Zanini, 2021). Matching was based on sex, age, pubertal status, and parent's average years of schooling as a proxy for cognitive stimulation, or socioeconomic status (SES; Sirin, 2005).

Both groups (online and in-person) were tested individually with supervision in a single session. Participants from the in-person group were tested at their schools using touchscreen tablets holding PDF files containing stimuli and instructions. Differently, participants tested online were at home and assessed remotely through an internet connection using their own hardware (computer or laptop, with web camera and

basic free software such as Adobe Acrobat, PowerPoint, and a free Zoom video communication application). These individuals were instructed by the experimenter to share their screens (step-by-step written and oral instructions were provided for those unfamiliar with Zoom). Next, participants were helped to download tests in PDF format from their own e-mails or their guardian's. These files were not available until testing started, so that they could not preview the tasks. For both groups, the examiner was present during the whole test session, supplying instructions, answering questions and ensuring participants were doing the tests as expected (e.g., not being interrupted by their cellphones and such like). All participants were awarded a "science partner" certificate after taking part and those tested in-person were reimbursed for their travel expenses. The EF test battery took around 40 min to be completed, including instructions and rest breaks if the participants required them. Approximate time taken to complete each task was: 2 min for both the Inhibition tasks, 4 min to complete the Color Shape task and 6 min to complete Category Switch (Shifting tasks), around 5 min to complete 2-Back task and 8 min to complete the Number Memory task (Updating tasks). Other tasks were administered to the same samples and their results will be reported elsewhere.

Cognitive Measures

The FREE battery contains six tests adapted for use in diverse samples in terms of SES and cultural context. The theoretical basis for the battery, the rationale for choice of tasks, description of tasks, answer sheets and scoring method are detailed in Zanini (2021). A brief description of tasks and scoring procedure for each domain can be found in **Table 1** and **Figure 1**. Following prior studies (see Zanini, 2021), the Inhibition and Switching tasks included blocks used to control for vocal/psychomotor speed and a corresponding block with the same requirements plus executive demands, while the Updating tasks contained no such control, as is the norm in this field.

To carry out the executive function tasks, testees read the instructions or had them read to them if preferred. The examiner clarified any questions that came up. Answers were vocal and the tasks were self-paced (swiping to pass from instructions to stimuli and between stimuli for the in-person group and clicking a mouse or tapping a forward keyboard arrow in the online group). The examiner wrote down the vocal answers and timed testees' task completions using a handheld stopwatch, akin to classic tests used to assess intelligence, for instance, following the long tradition of paper-and-pencil testing (for a further discussion on the advantages of this, see Kessels, 2019; Zanini, 2021). Sessions were recorded with participants' and their guardians' consent and erased once adequate scoring was ensured.

For each task or task block (depending on the test), the Rate Correct Score (RCS; see Vandierendonck, 2017) was calculated by dividing the number of correct answers by the time (in seconds) taken to complete each block. This metric controls for speed-accuracy trade-offs, that is, the

TABLE 1 | Description of the self-paced executive function tasks per domain and their corresponding scores (based on Zanini, 2021).

Domain (Task)	Paradigm	Scores
Inhibition		
(Stroop Victoria)	Contains two blocks, each of which consists of 24 stimuli (color patches or words) displayed on a single screen. Participants name the ink color of patches (block 1) and words that are color names written in incongruous ink colors (block 2). Block 1 is the control block, measuring speed to name colors. Block 2 involves inhibition (naming ink colors of words instead of reading the color names)	Cost of inhibition: RCS in block 2 minus RCS in block 1
(Happy-Sad Stroop)	Contains two scored blocks, each of which consists of 20 facial emotions that are displayed on a single screen. In block 1, participants name the emotions (happy or sad). In block 2, they inhibit naming the emotion and must name the opposite one (happy as sad or <i>vice-versa</i>)	Cost of inhibition: RCS in block 2 minus RCS in block 1
Switching		
(Color-Shape)	Contains three blocks in which single-colored geometric pictures are presented on each screen. As participants pass from screen to screen, pictures must be classified by shape (squares/circles) (block 1: 20 trials or screens), by color (black/gray) (block 2: 20 trials) or alternating (switching) classifications (block 3: 40 trials), according to cues presented above the pictures	Shifting costs: RCS in block 3 minus the sum of RCS in blocks 1 and 2
(Category Switch)	Contains three blocks in which single pictures are presented on each screen. As participant pass from screen to screen, each pictures must be classified as living or non-living (block 1: 20 trials or screens), big or small (block 2: 20 trial) or alternating (switching) classifications without cues (living/non-living, then big/small and so forth) (block 3: 40 trials)	Shifting costs: RCS in block 3 minus the sum of RCS in blocks 1 and 2
Updating		
(2-Back)	Each screen contains 10 square outlines in fixed locations, one of which is filled in with black ink. As participants pass from screen to screen, they answer if the black square location they see is in the same or a different location as the black square two screens back. The total number of updating opportunities is 66	Total RCS (for accuracy, in this case only: hits minus false alarms) (no control block)
(Number Memory)	Each screen contains a single digit number (1 to 9). As participants pass from screen to screen, they report the last three digits (trios) seen, in the same order as they were presented, having to continuously update information held in working memory, discarding the first digit of the trio and adding the new digit that appears next. The total number of updating opportunities is 24	Total RCS (no control block)

RCS = Rate Correct Score obtained by accuracy (vocal responses) divided by time (s) to complete blocks/task timed by the experimenter. See **Figure 1** for a visual illustration of the task.

between-participant variability in deciding to do tasks slowly, which can increase accuracy, or quickly, with higher error rates.

Statistical Analyses

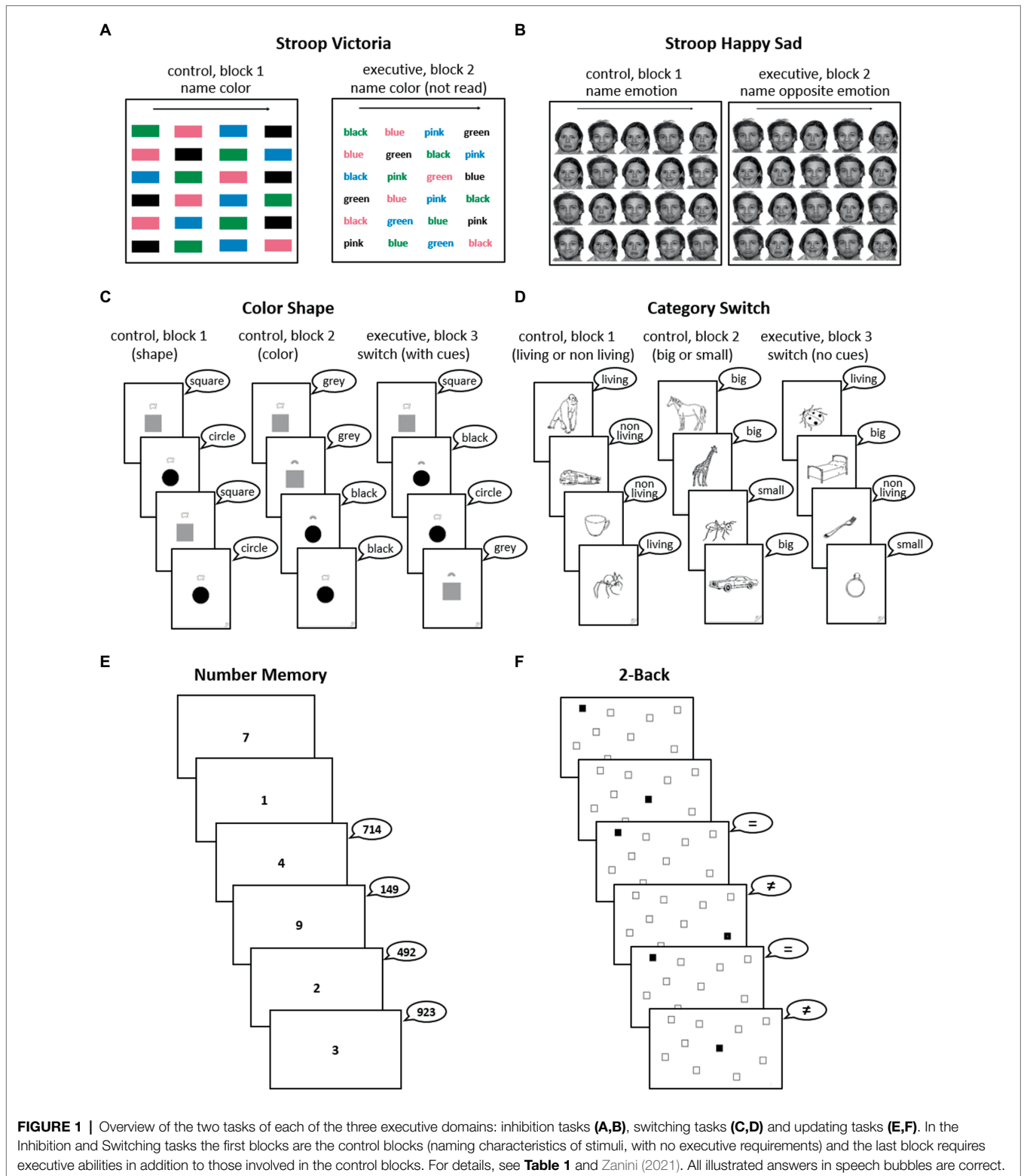
Descriptive and inferential analyses were conducted using IBM SPSS Statistics version 21 software. The scores entered as dependent variables in the statistical analyses of the Inhibition and Shifting tasks were the executive costs (RCS from the block with executive requirements minus the RCS from control blocks). These dependent variables for each task were used in separate univariate General Linear Models (GLMs) with the factor group (online vs. in-person). For the Updating¹ measures, total RCS was used as the dependent variables in similar GLMs because they do not include baseline/control blocks. To correct for speed (of vocal responses and/or passing from stimulus to stimulus), our analyses included another continuous predictor, the mean RCS of the control blocks of the Inhibition and Shifting tasks, which have no executive requirements and basically involve answering aloud about an attribute of the stimuli (retrieval of phonological representations from long-term

¹For the 2-Back task, because participants can adopt a strategy of guessing (without actually updating the content in their working memory) by responding that all spatial configurations are different from the one presented two trials back (only 36% of trials are the same), accuracy in this case was calculated as hits minus false alarms (Jaeggi et al., 2010). Guessing in the other tasks can be picked up by the examiner but this never occurred in our experience.

memory) and speed of progressing through all stimuli, which is also done in the Updating task (see Zanini, 2021).

Because there is no prior data on online administration of the FREE test battery in the literature, calculating sample sizes was not strictly possible. We therefore focused on determining effect sizes (unstandardized and standardized) of the effect of group, which are useful to indicate the magnitude of the reported effects in metrics that are comparable across the tests. Standardized effect sizes are also useful to communicate the practical significance of the results, for meta-analysis and, importantly, can be used in future studies to determine sample size if these investigations intend to compare participants tested online and in-person (see Lakens, 2013). Effect sizes were presented as partial eta squared provided by SPSS and Hedges g [using Excel spreadsheet],² with corresponding confidence intervals. Rules of thumb describe medium effect sizes as those, respectively, between 0.059–0.138 and 0.5–0.8; values equal to or higher than 0.138 and 0.8 are considered large effect sizes (see Richardson, 2011; Lakens, 2013). Because participants for the different groups were matched by sex, age, pubertal stage, and SES, these variables were not supposed to be different between groups, so we did not include them as covariates in the analyses. Data were inspected for outliers (values over three SD of the mean).

²<https://www.cem.org/effect-size-calculator>



RESULTS

After matching participants tested online and in person (see **Table 1**), the proportion of participants of each sex, their age,

and pubertal status and their parents' average years of schooling were not different across groups. However, for one pair of matched participants there was a difference of 7 years in the mean of parental schooling. We therefore sought to further

control possible differences in cognitive stimulation between groups. As the in-person sample was mostly from public schools, except one, while all but one member of the online group were from private schools, the type of school (public vs. private) was included as a covariant in the analyses. This was done because Brazilian private-school students often outperform those from public schools (which tend to have lower quality education) on executive function tasks (e.g., Guerra et al., 2021).

The following outliers were found per variable: one in the Happy-Sad Stroop and one in the 2-Back for the in-person group and one in the Number Memory task for the online group. Both exclusion of these values or replacement for the value of the mean plus three SD retrieved similar results, so we report the results including data of these outliers.

Regarding type of test administration, we found no significant group effects (online vs. in person) in any of the executive tasks except for a marginal effect (small effect size) in the 2-Back test [$F(1,92)=3.899$, $p=0.051$, $p_{\eta^2}=0.04$], with a non-significant tendency for lower scores in the in-person group. No effects of type of school in any of the tasks were observed (see **Table 2** and **Figure 2**), but there was an effect of the variable used to control for vocal/psychomotor speed in the 2-Back task [$F(1,92)=23.49$, $p<0.001$, $p_{\eta^2}=0.20$] and in the Number Memory task [$F(1,92)=55.09$, $p<0.001$, $p_{\eta^2}=0.37$], as expected, because the metrics used in the analysis of these tests do not have a baseline condition, unlike the inhibition and shifting tasks. The databank is available at https://osf.io/h5akr/?view_only=ea08777d698c46b4ae8110b9f8df8057t.

DISCUSSION

This study found no evidence that online, examiner-moderated use of the FREE test battery differs from in-person testing

(effect were not significant and of small effect sizes). This suggests that remote testing this way may be a comparable alternative when face-to-face assessment is not possible as found for other cognitive test batteries (Krantz and Dalal, 2000; McGraw et al., 2000; Nosek et al., 2002; Temple et al., 2010; Soto et al., 2011; Germine et al., 2012; Cullum et al., 2014). Hence, performing self-paced tasks and responding vocally, either personally or online, and using different hardware under these conditions did not affect results. This makes sense considering that the executive function variables were controlled for speed of vocal responses and passing from one stimulus to the next, irrespective of the conditions and equipment used by the participants: smaller screens and swiping to progress to the next stimuli during face-to-face assessment, or larger screens (laptops or personal computers) and mouse or key presses, remotely.

This absence of significant difference between in-person and remote testing may seem surprising considering that there can be minor delays and variation in precision when transmitting images, sounds and registering motor responses over the internet, although these seem to vary little across browsers, platforms, and operating systems (Anwyl-Irvine et al., 2020). The similar performances in the testing conditions used here may be explained by the FREE executive tests' design and scoring system, which is similar to classic paper and pencil tests, that have been used for decades and have considerable advantages over automated ones (Kessels, 2019). Time taken for each vocal answer is not the focus of interest in this test battery. Instead, it takes RCS into account, which is response accuracy divided by self-paced time taken for each task throughout a series of trials. Thus, total time to complete a task is much longer than reaction time per trial, which varies by milliseconds when computed in automatized tasks and could be affected by online transmission lags. Additionally, small variations in reaction

TABLE 2 | Descriptive statistics of demographics and Rate Correct Scores (RCS: accuracy divided by total time in s) of executive functions performance.

Variables	In person (n=48; 21 girls)	Online (n=48; 21 girls)	Hedges g (95% CI)	Group effects			Type of School effects		
	Mean (\pm SD)	Mean (\pm SD)		F	p	p_{η^2}	F	p	p_{η^2}
Demographics									
Age (years)	12.29 (1.97)	12.17 (1.96)							
PDS (score)	2.37 (0.71)	2.27 (0.71)							
Guardian's schooling (\bar{x} yrs.)	14.78 (2.27)	14.86 (2.58)							
Executive functions									
Inhibition									
Stroop Victoria (inhibition cost)	-0.66 (0.28)	-0.54 (0.27)	-0.43 (-0.84/-0.03)	2.239	0.14	0.02	0.016	0.9	<0.01
Stroop Happy-sad (inhibition cost)	-0.44 (0.29)	-0.45 (0.27)	0.04 (-0.36/0.44)	0.878	0.88	0.01	0.728	0.40	0.01
Switching									
Color-Shape (switching cost)	-0.45 (0.18)	-0.47 (0.17)	0.11 (-0.29/0.51)	0.028	0.87	0.001	0.038	0.85	<0.01
Category Switch (switching cost)	-0.33 (0.15)	-0.38 (0.14)	0.34 (-0.06/0.74)	1.525	0.22	0.02	0.009	0.93	<0.01
Updating (corrected for baseline speed)*									
2-Back	0.27 (0.17)	0.37 (0.15)	-0.62 (-1.03/-0.21)	3.899	0.051	0.04	0.147	0.70	<0.01
Number Memory	0.18 (0.05)	0.17 (0.05)	0.20 (-0.20/0.60)	0.884	0.35	0.01	0.000	0.99	<0.01

PDS=Pubertal Development Scale; *baseline speed effect for the 2-Back $F(1,92)=23.49$, $p<0.001$, $p_{\eta^2}=0.20$; Number Memory $F(1,92)=55.09$, $p<0.001$, $p_{\eta^2}=0.37$.

times of examiners, who mark the beginning and end of each block of trials, are comparable in the control and executive conditions/blocks, online and in person, and become irrelevant when compared to total time taken to complete each block/task. Although this presents some extra workload for examiners, it allows testing under a wider variety of conditions (Kessels, 2019). These tasks can therefore be used by examiners who do not have access to special hardware and software that automatically time responses for each trial and in private practices or poorly equipped laboratories, given that most of the world's neuropsychologists cannot afford these gadgets and applications. The present study shows that this advantage extends to online testing, as long as it is mediated by an examiner. In effect, inconsistent results have been found when comparing performance in tasks that are self-administered (e.g., in hospital vs. online settings: Feenstra et al., 2017), possibly because cognitive testing often needs supervision. Mediated testing allows examiners to make sure that testees understand tasks, pay attention when doing them and do not engage in the use of strategies that may distort performance.

We did find, however, a marginal group effect in the 2-Back task, which should be addressed. We envisage two possible explanations for this: (1) it could have been a spurious effect; and (2) although we controlled for type of school, it is possible that this statistical adjustment did not correct for a putative advantage of the online group, which had access to better schooling, having all been from private schools except for one individual. Indeed, Guerra et al. (2021) have recently shown that children from private schools in Brazil had higher spans for spatial locations (however, of a small effect size), but that spatial updating task performance, adjusting for span, did not differ between public and privately schooled individuals. Consequently, possible slightly higher spans due to higher cognitive stimulation could have contributed to the 2-Back results in the online group, which was not assessed here because of the nature of our 2-Back task, which differs from the n-back test of Guerra et al. (2021). In line with this idea, our findings show that the marginal non-significant effect in the 2-Back task does not seem to be specific to updating-related executive functioning, because group effects in the other task of this domain (Number Memory) were nowhere near significant, nor close to a medium effect size. It is also of note that, despite the lack of norms, the marginally significant difference found between scores from the in-person and online groups in the 2-Back task seem to have been due to a lower performance of the in-person group (mean \pm SD 0.27 ± 0.17) because the online group (0.37 ± 0.15) presented results in this task that are very similar (0.35 ± 0.13) to those found by Zanini (2021), in which all participants were tested in person. Indeed, overall, performance of the online sample was very similar to that obtained in a comparable population using the same battery in person: means obtained here were within mean ± 0.5 SD of Zanini (2021), which correspond to low effect size differences (Lakens, 2013).

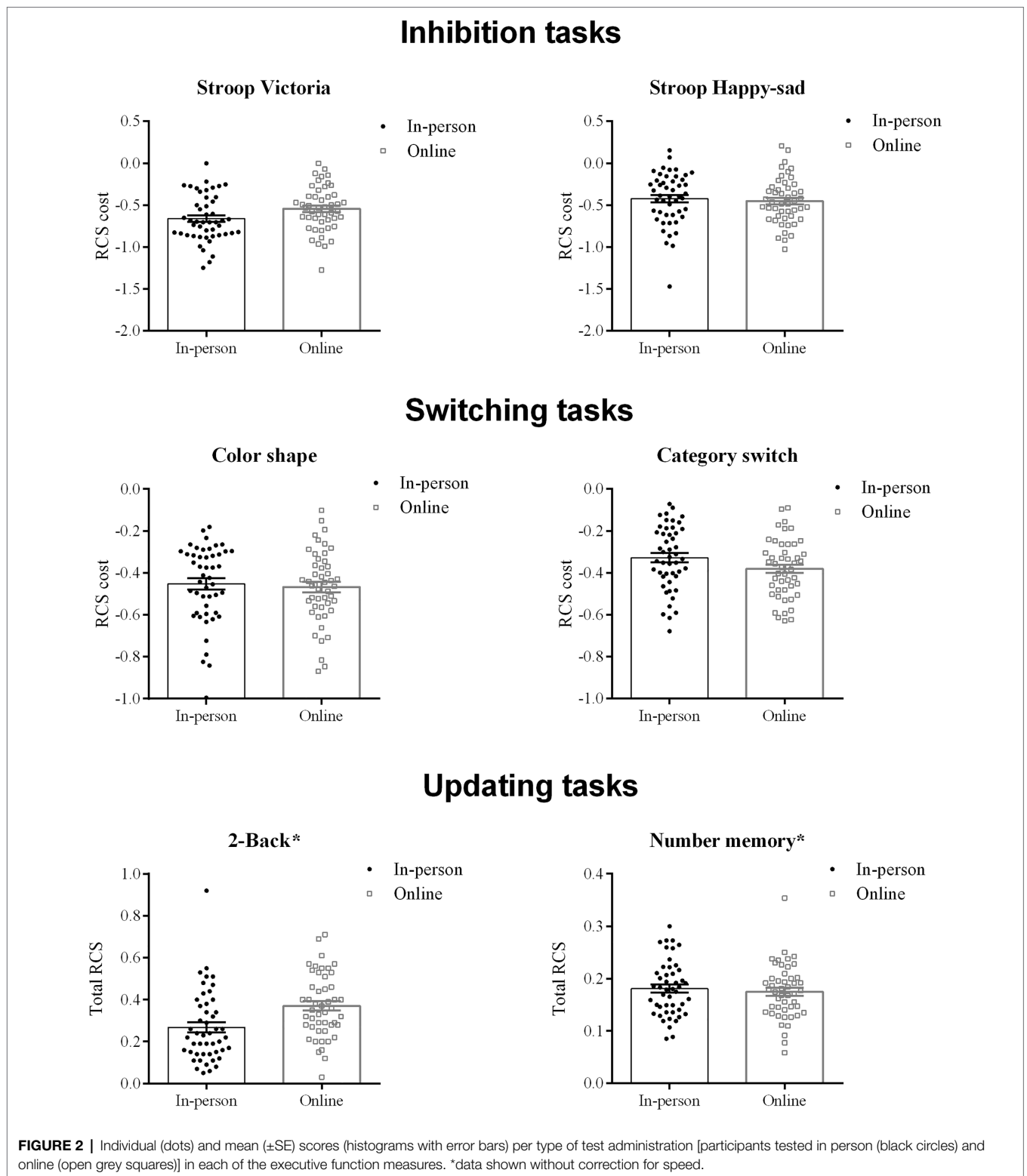
The FREE test battery was designed with two tasks of three executive domains, so that consistency of effects between each pair of tests of each type can be used to ascertain the influence

of various factors such as mode of testing (online vs. in-person), sex, SES, and so forth. Having two tasks of each domain also allows latent factors to be estimated, considering many multifactor configurations found for the Unity and Diversity Model of Executive Functions (see Karr, 2018). Unlike raw scores used here, latent factors capture the common variance in performance across different tasks, free of measurement errors (Brown et al., 2015). Therefore, specific cognitive requirements of a task that are not shared with the corresponding test of the same domain (such as spatial span in the 2-Back but not the Number Memory task) should not contribute to the latent factor. Our sample was too small to explore the latent nature and best model configuration of the executive functions' Unity and Diversity model, and to obtain evidence of invariance (Meade and Bauer, 2007) across mode of testing (online vs. in-person), so this must be undertaken in future studies. Another advantage of having a couple of tests per domain is that researchers who intend to use only one task of each EF type can pick the one which they deem more adequate for their purposes, although the ideal is to obtain latent scores. This, however, is only possible with large samples.

Considering the present scenario, evidence is emerging that COVID-19 and similar infections may lead to cognitive problems (cognitive COVID) beyond the acute stage (Ritchie and Chan, 2021), so repeatedly assessing infected patients may become essential to understand possible long-term cognitive impact. This includes children and adolescents, some of whom present long-term COVID effects (see Hertting, 2021) and will need to be followed up. As we have shown here, this may be done remotely with supervision, enabling a greater number of patients to be assessed if they have internet access, some familiarity with digital tools, a computer or device that runs basic software, a web camera, and a reasonably sized screen. All tests used may be easily adapted for diverse populations and are affordable (see Zanini, 2021) for poorly funded researchers.

Additionally, a point to bear in mind is that online testing may pose technical issues such as unreliable connections or slow speeds. Testees may also not be comfortable with the technology involved or not have the hardware and an internet connection, which might be the case for those from low SES. Furthermore, examiners may fail to notice difficulties that are not clear through vocal communication when participants are unable or unwilling to turn on their cameras (common in youngsters: Castelli and Sarvary, 2021). Online testing also poses some ethical problems that must be minimized such as violating privacy. Nonetheless, despite these shortcomings, online testing probably reaches much larger numbers at low cost and will therefore probably become more prevalent in the post-pandemic period.

The main limitation of the present study is that participants in the in-person group were tested before the pandemic and those tested online were assessed during the pandemic, so this could have affected the results. Ideally, both groups should have been evaluated in parallel, but due to the pandemic this was not ethically acceptable. Nonetheless, the negative acute effects of the lockdown, which seem to be more severe (Creswell et al., 2021), were avoided, as we only tested participants from



3 months after the beginning of the social distancing measures. Additionally, potential participants who were reported by guardians as not being healthy were not tested in either group. If the COVID-19 crisis had affected participants in ways that

interfered with their EF, beyond those that were controlled for by matching participants (which are the known factors that influence these cognitive abilities: Foulkes and Blakemore, 2018), it would be expected that the online group perform

worse, which was not observed. Various factors that could have changed due to the pandemic, such as increases in mental health problems, sedentarism, body weight, alterations in sleep patterns, and so forth are not usually considered at all in prior studies that investigated the same EF model as used here (e.g., Friedman et al., 2011, who used the same sample of twins in many publications). Hence, it would be unclear how or whether they could have affected results. Another limitation of our study includes its sample size. Because our aim was to test feasibility of supervised remote online testing using the FREE battery, we did not have a sample large enough to allow us to run confirmatory factor analysis of the Unity and Diversity model of executive functions (see Friedman and Miyake, 2017; Karr, 2018) and to perform invariance testing and other types of validation techniques to ensure that web-based assessment was tapping the same constructs as those measured face-to-face (see Germine et al., 2012). Matching participants not only by parental schooling but also the type of school would have also been ideal. We attempted to control for the latter statistically, but it might have not been an effective control as almost all participants in the online group were from private schools, which more readily agreed to help volunteers participants during school closure. We had little success accessing families through public schools because their staff were much more severely overburdened due to the pandemic and extremely low governmental funding to aid the transition to online teaching. Finally, unlike other similar studies that investigated the adequacy of remote cognitive assessment by controlling for individual differences in performance using within-participant designs (e.g., Cullum et al., 2014; Feenstra et al., 2017; Backx et al., 2020), we did not test the same participant face-to-face and online because test–retest reliability of executive functions is known not to be high (Karlsen et al., 2020) because testees develop strategies that minimize executive functioning. On the upside, this experiment used a sample from a developing nation, which is still rare in the international literature (Rad et al., 2018), especially regarding the adequacy of remote cognitive assessment.

Overall, although our findings cannot be generalized to samples from other cultures and age groups, we have shown that online testing using the FREE test battery is a potentially viable means of remotely assessing EF. Because this test battery is open access, adaptable to populations with different characteristics and remote testing was done using only free software, our results provide initial evidence for a much-needed remote way to assess adolescents at low costs, including those who are more vulnerable to factors that negatively affect developmental (Hunersen et al., 2021; Moffitt et al., 2011), including BAME and non-WEIRD populations, who are under-represented in the cognitive literature (see Henrich, 2010;

Rad et al., 2018). We conclude that online testing the way it was administered here is feasible way of collecting data on EF, making this a potential alternative when face-to-face testing is not possible. Until more controlled experiments are conducted, it is advisable to either test all participants online or in person and not mix these conditions, which was not assessed here.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found at: https://osf.io/h5akr/?view_only=ea08777d698c46b4ae8110b9f8df8057t.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comitê de Ética em Pesquisa of the Universidade Federal de São Paulo. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SP and IS: designing the study, planning of the study, data collection, data analysis, and interpretation of the data. SP: public responsibility for the content of the article. All authors contributed to the article and approved the submitted version.

FUNDING

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Processes: due to fellowships to author IS 2019/19709–6 and SP 2016/14750–0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (finance code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – #301899/2019–3 due to fellowships to author SP), and Associação Fundo de Incentivo à Pesquisa (AFIP).

ACKNOWLEDGMENTS

We thank Rafaella Sales de Freitas, Luanna Inácio, Eveli Truksinas, Robson Kojima, Beatriz Zappellini, Daniel Utsumi, Mariana Libano, Bianca Rodrigues and Diogo Marques for help with data collection.

REFERENCES

- Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., and Evershed, J. K. (2020). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behav. Res. Methods* 53, 1407–1425. doi: 10.3758/s13428-020-01501-5
- Backx, R., Skirrow, C., Dente, P., Barnett, J. H., and Cormack, F. K. (2020). Comparing web-based and lab-based cognitive assessment using the Cambridge neuropsychological test automated battery: a within-subjects counterbalanced study. *J. Med. Internet Res.* 22:e16792. doi: 10.2196/16792
- Baggetta, P., and Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind Brain Educ.* 10, 10–33. doi: 10.1111/mbe.12100

- Brown, G. T. L., Harris, L. R., O'Quin, C., and Lane, K. E. (2015). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *Int. J. Res. Method. Educ.* 40, 66–90. doi: 10.1080/1743727X.2015.1070823
- Carskadon, M. A., Vieira, C., and Acebo, C. (1993). Association between puberty and delayed phase preference. *Sleep* 16, 258–262. doi: 10.1093/sleep/16.3.258
- Castelli, F. R., and Sarvary, M. A. (2021). Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Acad. Pract. Ecol. Evol.* 2020, 3565–3576. doi: 10.1002/ece3.7123
- Creswell, C., Shum, A., Pearcey, S., Skripkauskaitė, S., Patalay, P., and Waite, P. (2021). Young people's mental health during the COVID-19 pandemic. *Lan. Child Adolesc. Health* 5, 535–537. doi: 10.1016/S2352-4642(21)00177-2
- Cullum, C., Munro, L., Hynan, S., Grosch, M., Parikh, M., and Weiner, M. F. (2014). Teleneuropsychology: evidence for video teleconference-based neuropsychological assessment. *J. Int. Neuropsychol. Soc.* 20, 1028–1033. doi: 10.1017/S1355617714000873
- Feenstra, H. E. M., Murte, J. M. J., Vermeulen, I. E., Kieffer, J. M., and Schagen, S. B. (2017). Reliability and validity of a self-administered tool for online neuropsychological testing: The Amsterdam cognition scan. *J. Clin. Exp. Neuropsychol.* 40, 253–273. doi: 10.1080/13803395.2017.1339017
- Fernández, A. L., and Abe, J. (2018). Bias in cross-cultural neuropsychological testing: problems and possible solutions. *Culture. Brain* 6, 1–35. doi: 10.1007/s40167-017-0050-2
- Foulkes, L., and Blakemore, S. J. (2018). Studying individual differences in human adolescent brain development. *Nat. Neurosci.* 21, 315–323. doi: 10.1038/s41593-018-0078-4
- Friedman, N. P., and Miyake, A. (2017). Unity and Diversity of executive functions: individual differences as a window on cognitive structure. *Cortex* 6, 186–204. doi: 10.1016/j.cortex.2016.04.023
- Friedman, N. P., Miyake, A., Robinson, J. L., and Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: a behavioral genetic analysis. *Dev. Psychol.* 47:1410. doi: 10.1037/a0023750
- Geddes, M. R., O'Connell, M. E., Fisk, J. D., Gauthier, S., Camicioli, R., and Ismail, Z. (2020). Remote cognitive and Behavioral assessment: report of the Alzheimer Society of Canada task force on dementia care best practices for COVID-19. *Alzheimer's. Dementia: Diagnosis. Assess. Dis. Monit.* 12, 1–11. doi: 10.1002/dad2.12111
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., and Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from web and lab in cognitive/perceptual experiments. *Psychon. Bull. Rev.* 19, 847–857. doi: 10.3758/s13423-012-0296-9
- Guerra, A., Hazin, I., Guerra, Y., Roulin, J. L., Le Gall, D., and Roy, A. (2021). Developmental profile of executive functioning in school-age children From Northeast Brazil. *Front. Psychol.* 11:596075. doi: 10.3389/fpsyg.2020.596075
- Hackman, D. A., Gallop, R., Evans, G. W., and Farah, M. J. (2015). Socioeconomic status and executive function: developmental trajectories and mediation. *Dev. Sci.* 5, 686–702. doi: 10.1111/desc.12246
- Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–135. doi: 10.1017/S0140525X0999152X
- Hertting, O. (2021). More research is needed on the long-term effects of COVID-19 on children and adolescents. *Acta Paediatrica. Int. J. Paediatrics.* 110, 744–745. doi: 10.1111/apa.15731
- Hodge, M. A., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., et al. (2019). Agreement between telehealth and face-to-face assessment of intellectual ability in children with specific learning disorder. *J. Telemed. Telecare* 25, 431–437. doi: 10.1177/1357633X18776095
- Hunersen, K., Ramaiya, A., Yu, C., Green, J., Pinandari, A. W., and Blum, R. (2021). Considerations for remote data collection among adolescents During the COVID-19 pandemic. *J. Adolesc. Health* 68, 439–440. doi: 10.1016/j.jadohealth.2020.11.020
- Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Yi Fen, S., Jonides, J., and Perrig, W. J. (2010). The relationship between N-Back performance and matrix reasoning-implications for training and transfer. *Intelligence* 38, 625–635. doi: 10.1016/j.intell.2010.09.001
- Karlsen, R. H., Karr, J. E., Saksvik, S. B., Lundervold, A. J., Hjemdal, O., Olsen, A., et al. (2020). Examining 3-month test-retest reliability and reliable change using the Cambridge neuropsychological test automated battery. *Appl. Neuropsychol. Adult.* 21, 1–9. doi: 10.1080/23279095.2020.1722126
- Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., Karr, J. E., et al. (2018). The Unity and Diversity of executive functions: a systematic review and re-analysis of latent variable studies The Unity and Diversity of executive functions: a systematic review and re-analysis of latent variable studies. *Psychol. Bull.* 144:1147. doi: 10.1037/bul0000160
- Kelkar, A. S., Hough, M. S., and Fang, X. (2013). Do we think alike? A cross-cultural study of executive functioning. *Culture. Brain.* 1, 118–137. doi: 10.1007/s40167-013-0010-4
- Kessels, R. P. (2019). Improving precision in neuropsychological assessment: bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *Clin. Neuropsychol.* 33, 357–368. doi: 10.1080/13854046.2018.1518489
- Kirkwood, K. T., Peck, D. F., and Bennie, L. (2000). The consistency of neuropsychological assessments performed via telecommunication and face to face. *J. Telemed. Telecare* 6, 147–151. doi: 10.1258/1357633001935239
- Krantz, J. H., and Dalal, R. (2000). "Validity of web-based psychological research," in *Psychological Experiments on the Internet*. ed. M. H. Birnbaum (California: Academic Press), 35–60.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863
- Lee, K., Bull, R., and Ho, R. M. (2013). Developmental changes in executive functioning. *Child Dev.* 84, 1933–1953. doi: 10.1111/cdev.12096
- McGraw, K. O., Tew, M. D., and Williams, J. E. (2000). The integrity of web-delivered experiments: can you trust the data? *Psychol. Sci.* 11, 502–506. doi: 10.1111/1467-9280.00296
- Meade, A. W., and Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Struct. Equ. Model. Multidiscip. J.* 14, 611–635. doi: 10.1080/10705510701575461
- 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
- Nosek, B. A., Mahzarin, R., and Greenwald, A. G. (2002). E-research: ethics, security, design, and control in psychological research on the internet. *J. Soc. Issues* 58, 161–176. doi: 10.1111/1540-4560.00254
- Pompéia, S., de Almeida Valverde Zanini, G., Inacio, L. M. C., da Silva, F. C., de Souza Vitale, M. S., Niskier, S. R., et al. (2019). Adapted version of the pubertal development scale for use in Brazil. *Rev. Saude Publica* 53:56. doi: 10.11606/s1518-8787.2019053000915
- Rad, M. S., Martingano, A. J., and Ginges, J. (2018). Toward a psychology of homo sapiens: making psychological science more representative of the human population. *Proc. Natl. Acad. Sci.* 115, 11401–11405. doi: 10.1073/pnas.1721165115
- Reips, U.-D. (2000). The web experiment: advantages, disadvantages, and solutions. *Psychol. Exp. Internet.* 1995, 89–117.
- Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educ. Res. Rev.* 6, 135–147. doi: 10.1016/j.edurev.2010.12.001
- Ritchie, K., and Chan, D. (2021). The emergence of cognitive COVID. *World Psychiatry* 20, 52–53. doi: 10.1002/wps.20837
- Silva, J., and Ribeiro, M. (2021). Short report social inequalities and the pandemic of COVID-19: the case of Rio de Janeiro. *J. Epidemiol. Community Health* 75, 1–5. doi: 10.1136/jech-2020-214724
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Rev. Educ. Res.* 75, 417–453. doi: 10.3102/00346543075003417
- Soto, C. J., John, O. P., Gosling, S. D., and Potter, J. (2011). Age differences in personality traits From 10 to 65: big five domains and facets in a large cross-sectional sample. *J. Pers. Soc. Psychol.* 100, 330–348. doi: 10.1037/a0021717
- Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., and Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia* 51, 98–105. doi: 10.1016/j.neuropsychologia.2012.11.013

- Temple, V., Drummond, C., Valiquette, S., and Jozsvai, E. (2010). A comparison of intellectual assessments over video conferencing and in-person for individuals with ID: preliminary data. *J. Intellect. Disabil. Res.* 54, 573–577. doi: 10.1111/j.1365-2788.2010.01282.x
- Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: a rejoinder on the binning procedure. *Behav. Res. Methods* 49, 653–673. doi: 10.3758/s13428-016-0721-5
- Worhach, J., Boduch, M., Zhang, B., and Maski, K. (2021). Remote assessment of cognition in kids and adolescents with daytime sleepiness: a pilot study of feasibility and reliability. *MedRxiv: Preprint. Server. Health. Sci.* doi: 10.1101/2021.03.24.21254190
- Zanini, G. A. V., Miranda, M. C., Cogo-moreira, H., Nouri, A., Fernández, A. L., and Pompéia, S. (2021). An adaptable, open-access test battery to study the fractionation of executive-functions in diverse populations. *Front. Psychol.* 12:627219. doi: 10.3389/fpsyg.2021.627219

Conflict of Interest: The authors declare no 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.

Copyright © 2021 Segura and Pompéia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.