*CORRESPONDENCE
Julian Fick
✉ j.fick@tu-braunschweig.de

How to talk about science in ways that are comprehensible and interesting? Evaluation of an evidence-based science communication training program for graduate students

Julian Fick[1,2]*, Friederike Hendriks[1,2] and Barbara Thies[2]

[1]Institute for Communication Science, Technische Universität Braunschweig, Braunschweig, Germany, [2]Institute of Educational Psychology, Technische Universität Braunschweig, Braunschweig, Germany

Introduction: In order to graduate, students in higher education need to acquire a diverse set of competencies, among them mastering communicative competencies and skills. However, training in science communication usually is not embedded in formal education programs and existing training programs are often insufficiently or not at all evaluated. We developed an evidence-based science communication training aiming to foster students' abilities in communicating about research comprehensively and in a way that better involves interlocutors. We evaluated the training's effectiveness in increasing students' attitudes, self-efficacy beliefs, knowledge and actual science communication performance.

Methods: We implemented a pre-post summative evaluation design. The sample consisted of 44 master's students from different disciplinary backgrounds.

Results: The participants were highly satisfied with the training. Their self-efficacy beliefs, self-rated abilities, knowledge, and favorable attitudes regarding strategies to enhance involvement increased. Strategy use in communication performance was measured by analyzing transcribed recordings of short video presentations. While use of comprehensibility-strategies remained on a high level, involvement-strategy use increased.

Discussion: We provide evidence for the effectiveness of our training program. Furthermore, we suggest an operationalization of basic science communication competencies and discuss challenges regarding the behavioral assessment of actual science communication performance, as well as the limitation of not having had a control group.

KEYWORDS

communication skills, evaluation, higher education, science communication, training program

1 Introduction

Successfully completing a university study program requires students to acquire a diverse set of skills. While this obviously includes subject-specific content and method knowledge, also covering science communication is increasingly demanded (Brownell et al., 2013). These science communication skills are not only demanded as learning outcomes in undergraduate and graduate study programs but also in the subsequent training of (future) scientists (European Commission, 2022). The importance of science communication is derived from the societal function of science in general, that is, to provide understanding of the world and orientation for decision making, especially when faced with high amounts of fake news.

Besides their importance, science communication skills are currently neither adequately covered in higher education (Brownell et al., 2013) nor in researcher education programs (European Commission, 2020). Furthermore, existing science communication programs have been criticized because they were badly or not at all evaluated (Baram-Tsabari and Lewenstein, 2017) and they were not described transparently enough to decide about the evidential basis of their methods and contents (Hagger et al., 2020). Therefore, the aim of the present study is to report the evaluation results of a science communication training program while both, the program itself and its evaluation were based on scientific evidence and theory.

1.1 Science communication as external communication by scientists

With *science communication* we refer to communication about a scientific topic between two individuals with different knowledge and understanding (one of them being a topic novice or expert in the scientific issue at hand; Bromme et al., 2001), and which is not mediated through third parties like press offices or mass media. That is, we address "external science communication" (Hanauska, 2020), in contrast to "internal science communication," which takes place within a scientific community (e.g., in scientific journals and at conferences). In such external communication settings, science communicators must not only be able to talk about their own research, but also to make it accessible to the interlocutor, who holds only limited understanding and interest.

As the specialization and proliferation of science increases, achieving expertise is only possible in few specific domains, for any one person. In consequence, a bounded understanding of science is a necessary precondition for learning not only during but also beyond formal science education (Bromme and Goldman, 2014). Despite their bounded understanding of science, laypeople do engage with scientific knowledge in several ways, as it is at the core of many decisions of daily or social relevance. This is especially prevalent in "socio-scientific issues," that is "complex societal issues with conceptual, procedural, and/or technological associations with science" (Sadler et al., 2016, p. 1622), such as climate change, genetically modified food, artificial intelligence, and public health. For example, deciding whether to vaccinate against seasonal influenza requires a basic knowledgeability about vaccine effectiveness, but also an attitude about public health measures, again based on an understanding of heard immunity. Often, basic science knowledge and understanding, and the related formation of attitudes rest on the reception of science communication, for example, watching an expert interview on the evening news. In such

ways, science communication shares critical features with science education, where contents are reconstructed to be comprehensible, but also to connect content to students' prior experiences, knowledge, abilities, and skills (Duit et al., 2012).

Hence, we determine two core aims of science communication in interpersonal settings, as previously defined for the scope of this paper: (a) achieving basic comprehension, and (b) evoking topic interest in target persons. Both aspects are quality criteria of science communication that are foundational in many different communicative occasions (Pilt and Himma-Kadakas, 2023), can be conceived as essential learning objectives for occasional communicators (Lewenstein and Baram-Tsabari, 2022), and as such, are already relevant for master's students as future graduates of a scientific study program.

1.2 Training as an educational and psychological intervention method

In general, training programs can be understood as specific interventions characterized by a high level of structure (often implemented through modularization and manualization), the repetitive exercise of specific tasks, and the aim to increase rather practical skills (Fries and Souvignier, 2015). From the perspective of educational psychology, these interventions are normatively required to be both, evidence-based and properly evaluated for their effectiveness. To be evidence-based, interventions should be based on theories (e.g., social cognitive theory, theory of planned behavior) as a rationale to decide which instructional methods (e.g., discussions, peer feedback) should be used and furthermore as a guide to describe how exactly behavior changes (Hagger et al., 2020). The claim for being evidence-based also applies to strategies taught within the training. To give an example, when teaching comprehensibility strategies, only those strategies should be covered that have already been shown empirically to be effective in increasing comprehensibility. Regarding the second requirement of being properly evaluated, we propose that the evaluation should be closely linked with the intervention's learning goals and its instructional methods. This linkage between intended learning outcomes, instructional methods, and evaluation is called constructive alignment (Biggs, 1996).

1.3 Evaluation of science communication training programs

After reviewing existing training programs on science communication, Baram-Tsabari and Lewenstein (2017) concluded that they often are either insufficiently evaluated or not evaluated at all. Insufficient evaluation entails assessments solely based on anecdotes (e.g., Holliman and Warren, 2017), satisfaction ratings (e.g., Bobroff and Bouquet, 2016; Seakins and Fitzsimmons, 2020) or self-report questionnaires (e.g., Ahmed et al., 2021; Akin et al., 2021; Cirino et al., 2017; Clark et al., 2016; Druschke et al., 2022; Schiebel et al., 2021; Smith and McPherson, 2020; Montgomery et al., 2022; Pfeiffer et al., 2022). While satisfaction ratings and self-report questionnaires are a legitimate and especially economic method of data collection, they face some instrumentation-related sources of bias, including social desirability bias, response bias, and response-shift bias (e.g., Barel-Ben David and Baram-Tsabari, 2019). In particular, a recent meta-analysis

reported student evaluations of teaching to be unrelated to actual student learning (Uttl et al., 2017) and satisfaction ratings seem to be neither necessary nor sufficient for successful learning (Clark, 2020). These challenges can be addressed by extending self-report and satisfaction measures with more objective assessments such as audience ratings, performance tests, or behavioral observations.

Furthermore, the few studies that do include more objective evaluation instruments report inconclusive results. While Clarkson et al. (2018) and Rodgers et al. (2018) state that their training had a significant effect on external reviewers' judgments of quality, Rubega et al. (2021) did not find meaningful improvement compared to a control group. One reason for inconclusive findings might be that most of these studies were conducted with a small number of participants bearing the risk of being underpowered, potentially resulting in low reproducibility and overestimation of effect sizes (Button et al., 2013). Another reason for inconclusive results may lie in the specific mechanisms underlying the intervention. To give an example, the training might actually lead to an improvement in knowledge about strategies but this change might not automatically manifest in actual behavior, a phenomenon called inert knowledge that is closely linked to the instructional methods applied in the intervention (Renkl et al., 1996). Lastly, the lack of effectiveness of the single strategies covered within the intervention might also be a reason for inconclusive evaluation results. Only few interventions were explicitly based on strategies that have empirically been shown to be effective.

1.4 (Science) communication competencies

Spitzberg (2015) differentiated two perspectives of communication competence: While the *ability perspective* conceptualizes competence as "potential to perform certain […] sequences of overt behaviors" (p. 560), the *impression perspective* conceptualizes it as subjective value judgments on "whether or not an ability has been manifested in a way that results in positive judgments of a person's communicative ability" (p. 561). The former is congruent with educational and psychological definitions of competence. According to Weinert (2001, p. 62), the term competence "refers to the necessary prerequisites available to an individual or a group of individuals for successfully meeting complex demands." These requirements involve cognitive components alongside aspects of attitude and motivation, which are reflected in knowledge, attitudes, and self-efficacy beliefs. In these conceptualizations, competence can be differentiated from performance, which is the mastery of demands in a specific situation (Klieme and Hartig, 2008). Put differently, having certain pieces of knowledge, attitudes, and self-efficacy beliefs might be necessary, but not sufficient to master the demands in a specific situation that requires the complex interplay of these aspects, and thus it might be useful to differentiate the competence as the potential to act in a certain way from the realization of that potential.

While from this perspective, competence is a characteristic of a person, it is rather seen as an "evaluative inference" (Spitzberg, 2009) from the impression perspective. As Spitzberg (2009) argues based on a systemic approach, "in social context, actual ability is less important than what people think about the ability" (p. 73). The reason for this lies in the "intrinsic complexity and unpredictability of any given interaction" (p. 73). More specifically, communication is characterized by equifinality and multifinality which means that different actions can lead to the same goal and the same action can lead to different communication goals

(Spitzberg, 2013). As a result, whether a communicative action can be judged to be competent is dependent on the perceiver. In turn, each perceiver may provide an individual perspective (influenced, for example, by personality traits or spontaneous desires) and "each of these perspectives permit, and may systematically imply, divergent types of competence evaluations" (Spitzberg, 2009, p. 73).

However, there seem to be some patterns between communication strategies and their effects on recipients that consistently emerge empirically under certain temporal and local conditions. For example, jargon usage has repeatedly been shown to reduce the impression of comprehensibility while jargon avoidance increases it (Shulman et al., 2020, 2021; Dayton and Dragojevic, 2024). Therefore, although this conception assumes competence to be a subjective judgment and therefore rejects the idea that strategy usage perfectly predicts its communicative effects, it is compatible with the idea that certain strategies increase the probability of certain outcomes (Spitzberg, 2009).

From our perspective, both approaches must be combined to grasp the term "science communication competence" appropriately. Following the ability approach "in this paper", we define science communication competence as the necessary prerequisites (i.e., declarative and procedural knowledge, attitudes and self-efficacy beliefs) available to an individual for successfully meeting the complex demands of science communication. Science communication performance, on the other hand, is the realization of this competence in a specific situation. To simultaneously acknowledge the impression approach, we only covered those communication strategies that have empirically shown to be effective in reaching a desired communicative goal or impression, and the implementation of these strategies is assumed to increase the probability of the desired impression, not to guarantee it.

1.5 The development and contents of our training program

To address the reviewed shortcomings in the development and evaluation of science communication interventions, we developed an evidence-based science communication training program. Based on the constructive alignment approach, we aligned the intervention's contents, its instructional methods and its evaluation on each other. For all three of these components, we tried to rely as closely as possible on empirical evidence as well as theories from (educational) psychology and the interdisciplinary field of science communication. The process of the training development is described in Fick et al. (2025).

Within this intervention, we aimed to teach basic practical science communication skills that master's students can apply when talking about their research to laypeople, be it in private occasions, in their studies or workplace. Therefore, the intervention consisted of four modules. The first covered science communication basics to build up a psychological model of science communication. Additionally, we discussed quality criteria for science communication, the need to adapt the own message to the specific audience and medium as well as the way in which the involved persons' personality characteristics influence the communication and its success. The second and third modules were about comprehensibility and involvement strategies, respectively. In both cases, specific strategies were introduced using authentic science communication examples and participants were instructed to implement these strategies on

their own research. As participants were required to have already completed a bachelor's degree program, most of the students worked on communicating the contents of their bachelor's thesis. From our perspective, it was essential for the participants to communicate about a topic on which they have a certain degree of expertise as they should be confronted with the demand to make their highly specialized knowledge accessible to a lay audience in terms of comprehensibility and involvement. Besides reducing jargon, and increasing syntactical simplicity as well as overall structure, we taught the comprehensibility strategy to increase vividness by giving concrete examples or using comparisons. This strategy was based on Dual-Coding Theory (Paivio, 1971), Cognitive Theory of Multimedia Learning (Mayer and Moreno, 2002) and empirical evidence from this context (e.g., Sadoski et al., 2000). The involvement strategies were based on the situated expectancy-value theory (Eccles and Wigfield, 2024) and focused on increasing the intrinsic and utility value of the communication. Increasing intrinsic value means making the communication as joyful and effortless as possible; increasing utility value means stressing the relevance of the topic for the recipient. While all modules had practical parts, the fourth and final module solely focused on actively applying all of the learned strategies using peer and video feedback methods in small groups. A detailed overview of the program and the taught strategies and their scientific basis is provided in Supplementary Tables S1, S2.

The instructional methods of the intervention were based on social cognitive theory (Bandura, 1989), mainly by providing mastery and vicarious experiences, as well as effective educational techniques (e.g., direct instruction, feedback from trainer and peers), and educational materials like slides and worksheets.

1.6 Hypotheses

The present study investigates whether our evidence-based training program is effective in raising the proposed competencies. We first assumed the participants to be highly satisfied with our training program (H1). We covered satisfaction ratings to receive feedback on different instructional features (e.g., structure, materials) and for formative evaluation purposes.

Further hypotheses reflect the intervention's summative evaluation and are based on our competence model introduced above: We expected participants' self-efficacy beliefs regarding communicating science (H2), their self-rated knowledge (H3), their self-rated practical skills (H4), their actual knowledge (H5), and their favorable attitudes regarding science communication (H6), to be higher after the intervention (T2) than before (T1). In addition, we hypothesized the intervention to be effective in increasing participant's science communication performance (H7).

2 Materials and methods

The present study was preregistered[1]. Changes between preregistration and actual analyses are reported in Supplementary material, code and data can be found on the open science framework[2].

2.1 Sample and sampling procedures

The study was advertised using typical channels of the university including mailing lists, social media accounts and the university's digital learning platform. To be eligible, participants had to currently be enrolled in a master's program and be fluent in German. The intended sample size of $n = 37$ was the result of power calculations with the smallest obtained effect size in a pilot study ($r = 0.43$). The intervention and therefore the data collection took place from May 2023 to January 2024 in university buildings.

All in all, 51 master's students took part in the intervention study. Some of them were able to earn a credit point for their studies through participation in the intervention. However, this was independent of their choice to simultaneously participate in the study. Five students dropped out of the study because of personal time constraints ($n = 3$), and because they wanted to only participate in the program ($n = 2$). Additionally, two participants had to be excluded because of missing data. Of the remaining 44 participants, 22 reported to be female, 21 male and 1 non-binary. They took part in the study on seven different days, had a mean age of 25.50 years ($SD = 2.75$), and were currently enrolled in their third master's semester on average ($SD = 1.65$). Most of them studied engineering subjects (52%), followed by natural sciences (32%), social sciences (9%), management (5%), and teacher training (2%).

Additional five participants had to be excluded from the performance data analyses, because they explicitly agreed to only participate in the paper-pencil part of the study ($n = 4$), or because their audio was not comprehensible and thus could not be transcribed in a useful way ($n = 1$). Therefore, the final sample size is 44 for the questionnaire data and 39 for the video data.

2.2 Procedure

We implemented a pre-post summative evaluation design. Therefore, participants had to complete one assessment before and one after the intervention. Both assessments consisted of a questionnaire asking for attitudes and self-efficacy beliefs regarding science communication, a self-rating of (science) communication abilities and a knowledge test. Only the prior-questionnaire covered socio-demographic information and only the posterior questionnaire included satisfaction ratings. Science communication performance was measured pre and post intervention. In this assessment, participants went to a separate room to give two-minute-presentations about the contents of a research project they already had conducted. These presentations were video recorded by an assistant who upheld a standardized protocol and read instructions to participants.

2.3 Intervention

The intervention has already been described in the introduction. It was held by the first author. To assure treatment fidelity, the training program was manualized and parts of it were video recorded.

2.4 Measures

We provide a complete list of items as well as item and scale characteristics in Supplementary material. The COTAN review system

for evaluating test quality (Evers et al., 2015) was used to interpret reliabilities.

2.4.1 Satisfaction with the training program

Satisfaction with the intervention was assessed using an adapted version of the *Trier Inventory of Teaching Quality* (TRIL; Fondel et al., 2015) on a Likert scale ranging from *totally disagree* (1) to *totally agree* (6). The scale covered four subscales: (a) satisfaction with structure and didactics (six items), (b) satisfaction with how stimulating the intervention was (six items), (c) satisfaction with the practical relevance (two items), and (d) satisfaction with the social climate (two items). The subscales achieved sufficient to good reliabilities ($\alpha_{structure} = 0.83$, $\alpha_{stimulation} = 0.88$, $\alpha_{relevance} = 0.69$, $\alpha_{climate} = 0.62$).

The TRIL was supplemented by two items to assess the overall satisfaction with the training program (taken from Thielsch and Hirschfeld, 2010). The first one asked the participants whether they would recommend the intervention to others (answer options: *yes*, *no*) and the second asked the students to rate the intervention on a common German grading scale ranging from *insufficient* (0) to *excellent* (15).

2.4.2 Self-efficacy beliefs in doing science communication

Two different measures were used to assess the participants' self-efficacy beliefs. First, we adapted the three items of the *Short Scale for Measuring General Self-efficacy Beliefs* (ASKU; Beierlein et al., 2013) to the context of science communication. While reliability was sufficient at T1 ($\alpha_{T1} = 0.62$), it was good at T2 ($\alpha_{T2} = 0.72$). Second, to assess self-efficacy in a constructively aligned way, we self-developed 16 items, eight of which focused on comprehensibility and eight on involvement. In case of comprehensibility, reliability was good at T1 ($\alpha_{T1} = 0.76$) and T2 ($\alpha_{T2} = 0.88$) and in case of involvement it was good at T1 ($\alpha_{T1} = 0.72$) and T2 ($\alpha_{T2} = 0.88$). All self-efficacy belief items were assessed using a Likert scale ranging from *totally disagree* (1) to *totally agree* (5).

2.4.3 Self-rated knowledge and practical skills

We asked participants to self-rate their skills in and their knowledge on communication in general and science communication in particular. The resulting four single item measures (adapted from Rodgers et al., 2020) were rated on a scale ranging from *no skill/ knowledge at all* (0) to *very high skill/knowledge* (10).

2.4.4 Actual knowledge test

We self-developed seven knowledge items with a total of 28 evaluative units that tested the declarative knowledge of the participants on central contents of the intervention. For example, participants were asked to name quality criteria of science communication. Two independent coders used a coding scheme to grade each of the evaluative units as either *incorrect* (0) or *correct* (1) and the units were summed up to form the knowledge index. The coders reached a moderate ($\kappa = 40$) to very high ($\kappa = 0.91$) agreement (Landis and Koch, 1977, see Supplementary material for details).

2.4.5 Attitudes toward science communication

Attitudes regarding science communication were assessed at two different levels. The first level covered whether participants think that scientists should communicate their research to society. Although we originally developed four items to cover this societal level, two of them were excluded from analysis based on the results

of a confirmatory factor analysis. While the reliability of the remaining two items was insufficient at T1 ($\alpha_{T1} = 0.56$), it was good at T2 ($\alpha_{T2} = 0.84$). The second level covered whether participants think that it is important for themselves to communicate in a comprehensible (a) and involving (b) way, and both of these scales were constructively aligned to the intervention contents. To give an example, we asked participants whether it is important for them to communicate about their research in an understandable way. Four items were used to assess each of both constructively aligned aspects and in either case, reliability was sufficient at T1 ($\alpha_{comp} = 0.66$, $\alpha_{invol} = 0.65$) and good at T2 ($\alpha_{comp} = 0.81$, $\alpha_{invol} = 0.87$). All attitude items were assessed on a Likert scale ranging from *totally disagree* (1) to *totally agree* (7).

2.4.6 Science communication performance

To increase ecological validity, we assessed actual science communication performance in a scenario-based way. Participants were instructed by assistants to imagine meeting an acquaintance who asks about the contents of a completed research project. Most participants chose their bachelor's thesis. They should assume this acquaintance to neither have an academic degree nor any prior knowledge on the topic. They were then given a maximum of 2 minutes to present their thesis. These presentations were video recorded, and transcribed using the rules proposed by Dresing and Pehl (2018). Subsequently the usage of strategies taught within the intervention was coded and the correlation between raters ranged from medium ($r = 0.56$) to very high ($r = 0.93$). Details about the instruction and coding are provided in Supplementary material. As an exception, we did not code the syntactical complexity but calculated the mean sentence length and used it as an indicator. While, on average, the transcripts had 240 words ($SD = 68$) before the intervention, they had 271 words ($SD = 68$) afterwards.

2.5 Analytic strategy

The analyses were done using R (R Core Team, 2021). If not stated otherwise, we used histograms and the Shapiro–Wilk test to determine whether the distribution of differences between T1 and T2 significantly deviated from a normal distribution. If this was the case, Wilcoxon signed-rank tests with continuity correction were performed. Otherwise, we used paired-sample t-tests. As we had directional hypotheses, one-sided tests were applied.

3 Results

The statistical results are reported in Table 1 and narratively described in the following.

3.1 H1: participants are satisfied with the training program

We could confirm the hypothesis that participants score significantly higher than the scale mean (i.e., 3.50) on the TRIL and each of its subscales (H1a). Similarly, the program received an average grade of 13.18, which was significantly above the assumed threshold

TABLE 1 Overview of descriptive and inferential statistic results ordered by hypothesis.

Hypothesis	Variable	M (SD) pre	M (SD) post	Test statistic
H1a	Satisfaction (TRIL) regarding…			
	… structure		5.46 (0.53)	$V = 989, p < 0.001$
	… level of excitation		5.26 (0.62)	$V = 989, p < 0.001$
	… relevance		5.41 (0.73)	$V = 986, p < 0.001$
	… social climate		5.72 (0.44)	$V = 990, p < 0.001$
H1c	Satisfaction (grade)		13.18 (1.69)	$V = 974, p < 0.001$
H2a	Self-efficacy beliefs (self-developed) regarding using …			
	… comprehensibility strategies	3.29 (0.54)	3.79 (0.60)	$t(43) = 6.11, p < 0.001, d = 0.88$
	… involvement strategies	3.16 (0.48)	3.60 (0.57)	$t(43) = 7.54, p < 0.001, d = 0.81$
H2b	Self-efficacy beliefs (ASKU)	2.80 (0.53)	3.52 (0.58)	$V = 16, p < 0.001, r = 0.84$
H3a	Self-rated knowledge about…			
	… communication	5.18 (1.66)	6.89 (1.45)	$V = 19.5, p < 0.001, r = 0.82$
H3b	… science communication	4.25 (1.95)	6.82 (1.32)	$V = 5.50, p < 0.001, r = 0.87$
H4a	Self-rated skills about…			
	… communication	5.55 (1.65)	7.05 (1.36)	$V = 18, p < 0.001, r = 0.84$
H4b	… science communication	4.64 (1.75)	6.91 (1.38)	$V = 6, p < 0.001, r = 0.87$
H5	Knowledge	0.18 (0.15)	0.57 (0.19)	$t(43) = 13.96, p < 0.001, d = 2.29$
H6a	Attitudes regarding using…			
	… comprehensibility strategies	6.23 (0.67)	6.27 (0.72)	$V = 186, p = 0.249, r = 0.13$
	… involvement strategies	5.50 (0.83)	5.79 (0.89)	$V = 240.5, p = 0.030, r = 0.31$
H6b	Attitudes on societal level	5.80 (0.87)	5.78 (1.00)	$V = 177.5, p = 0.526, r = -0.01$

of 10. Finally, all but one participant (98%) would recommend the intervention to others, thus H1b, which stated that at least 80% of the participants would do so, was confirmed as well.

3.2 H2 – H6: science communication competencies increase due to the training program

As expected, self-efficacy beliefs (H2; see Figure 1), self-rated knowledge about communication in general and science communication in specific (H3, see Figure 2, top), self-rated skills about communication in general and about science communication in specific (H4, see Figure 2, bottom), actual knowledge (H5, see Figure 3), and attitudes regarding involvement strategy (H6a) use increased from T1 to T2. Against our expectations, attitudes regarding the usage of comprehensibility strategies (H6a) did not differ significantly from before to after the intervention. Similarly, participants' attitudes regarding science communication on a societal level (H6b, see Figure 4) did not change due to the intervention.

3.3 H7: science communication performance increases due to the training program

We expected participants to more often implement comprehensibility and involvement strategies after the intervention

than before. In particular, we hypothesized participants to increase vividness of their presentation (H7a), to increase structure (H7b), to reduce jargon (H7c) and to simplify the syntactical structure (H7d) in case of comprehensibility and to increase the use of intrinsic (H7e) and utility value (H7f) strategies in case of involvement. Using the one sided McNemar's Chi-squared test with continuity correction, we were able to confirm the hypothesis regarding vividness ($\chi^2(1) = 5.88$, $p = 0.008$), but not regarding jargon reduction ($t(38) = 1.67, p = 0.051$, $d = 0.28$) and structuring. In case of the latter, all participants already applied a structuring strategy at T1 and continued to do so at T2. Concerning involvement strategy usage, we expected participants to implement more strategies that increase the intrinsic value (H7e) and the perceived utility (H7f) of their presentation after the intervention than before. Both, intrinsic value strategies ($\chi^2(1) = 4.90, p = 0.013$) and utility value strategies ($\chi^2(1) = 3.27, p = 0.035$) significantly increased due to the training program. Against expectation, the syntactical complexity of the presentations increased instead of decreased due to the intervention as indicated by the rise of the average sentence length from 19.10 ($SD = 6.24$) before to 21.02 ($SD = 4.38$) after the intervention ($V = 225, p = 0.990, r = 0.37$).

Furthermore, we exploratively analyzed whether not only the usage itself, but also the frequency of usage increased due to the intervention. Similarly, one-sided Wilcoxon signed-rank tests revealed a statistically significant higher frequency of vividness strategy ($V = 101.5, p = 0.005$, $r = 0.49$), intrinsic value strategy ($V = 5, p = 0.007, r = 0.79$) and utility value strategy usage ($V = 81, p < 0.001, r = 0.64$) at T2 compared to T1. Again, there was no significant difference in the usage of structuring strategies ($V = 37.5, p = 0.151, r = 0.29$).
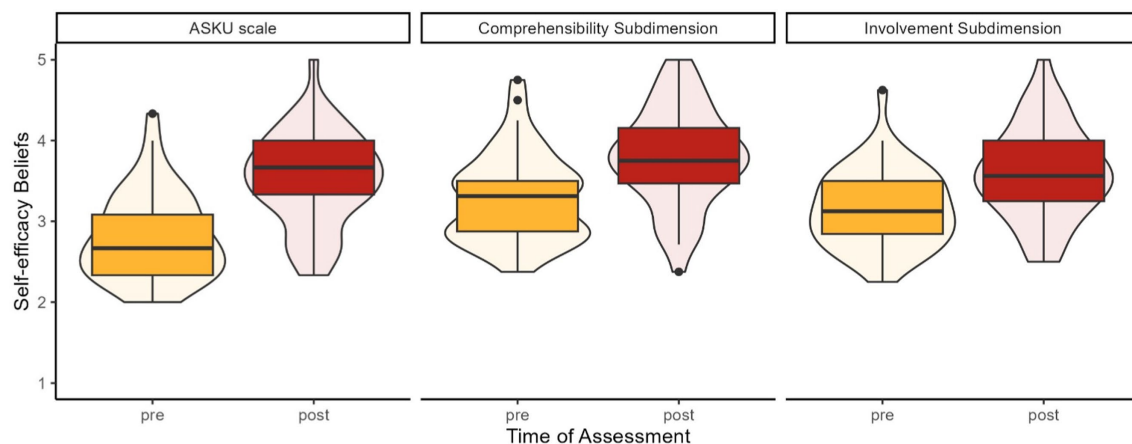
FIGURE 1
Ratings of self-efficacy beliefs before and after the intervention separated by the adapted ASKU scale and the self-developed scales representing self-efficacy regarding implementing comprehensibility and involvement strategies. Self-efficacy beliefs were assessed using a 5-point scale ranging from *totally disagree* (1) to *totally agree* (5).

4 Discussion

We conducted a pre-post summative evaluation study of an evidence-based science communication training program directed at master's students. The intervention can be considered as evidence-based in two ways. First, the contents and especially the communication strategies that we taught were derived from scientific theories and empirical evidence. Additionally, we used social cognitive theory as a theoretical approach to behavior change and to justify the usage of instructional methods. Second, conducting the evaluation study, we gathered evidence regarding the effectiveness of our intervention. In sum, we conclude that our intervention increased self-efficacy beliefs, self-rated abilities and actual knowledge when assessed directly after the intervention, but that its effect on favorable attitudes and actual performance was ambiguous. Our evaluation was based on a model of science communication competence that integrates (a) classical conceptions of competence as a potential (manifested in knowledge, attitudes and self-efficacy beliefs), (b) the realization of this competence in performance, and (c) the importance to consider the impression that this performance has on others.

4.1 Inconclusive results regarding the intervention's effect on attitudes

We could not confirm the hypotheses that attitudes regarding comprehensibility strategy usage and attitudes on societal level would increase due to the intervention. Having retrieved an average approval rating of 6.23 on a seven-point Likert scale, we explain the latter finding mainly due to the already very high favorable attitudes at T1. This in turn could be an indicator for sampling or selection bias which might raise the generalizability of our findings into question, especially for students that are more skeptical regarding science communication and less motivated to learn about it.

The second hypothesis that we could not confirm referred to more general, societal level attitudes on whether scientists should communicate and whether society would benefit from that. From our

perspective, the main explanation for this unexpected finding is that we did not address this topic explicitly within the training program. Besides the satisfaction ratings, this was the only scale that covered aspects that were not constructively aligned with the intervention's objectives, contents and methods. The reason for this was that we primarily strived to increase the quality of participants' science communication and not their quantity. Therefore, we did not explicitly discuss benefits of science communication on different levels, for example, for the individual scientists, their institutions, or for society in general. However, we assessed them because attitudes crucially contribute to the prediction of behavior (Ajzen, 1985). As we wanted to increase the quality of science communication or more specifically the quantity with which certain strategies are used, specific attitudes that refer to this usage seemed more relevant to address than society-level attitudes that explain and predict the frequency of science communication in general.

4.2 Mixed results regarding science communication performance

Regarding the performance in science communication, we confirmed an effect on the use of vividness (H7a), intrinsic value (H7e), and the utility value strategies (H7f). However, we could not confirm the hypotheses that participants more often use structuring, jargon avoidance, and syntactical simplification strategies after the intervention than before. While the former result can easily be explained by the fact that all participants already used structuring strategies at T1, and continued to do so at T2, the latter results are more challenging to explain. One reason is that through the course of the intervention, participants spent quite some time thinking about their thesis and what they wanted to say about it. As a result, the average number of words increased from 240 to 271. In addition, "the message production process is itself complex and multileveled" (Berger, 2003, p. 268) represented at semantic, syntactic, phonological and morphological levels (Dell, 1986) and involves several steps of formulation to come from conceptualization to articulation (Levelt
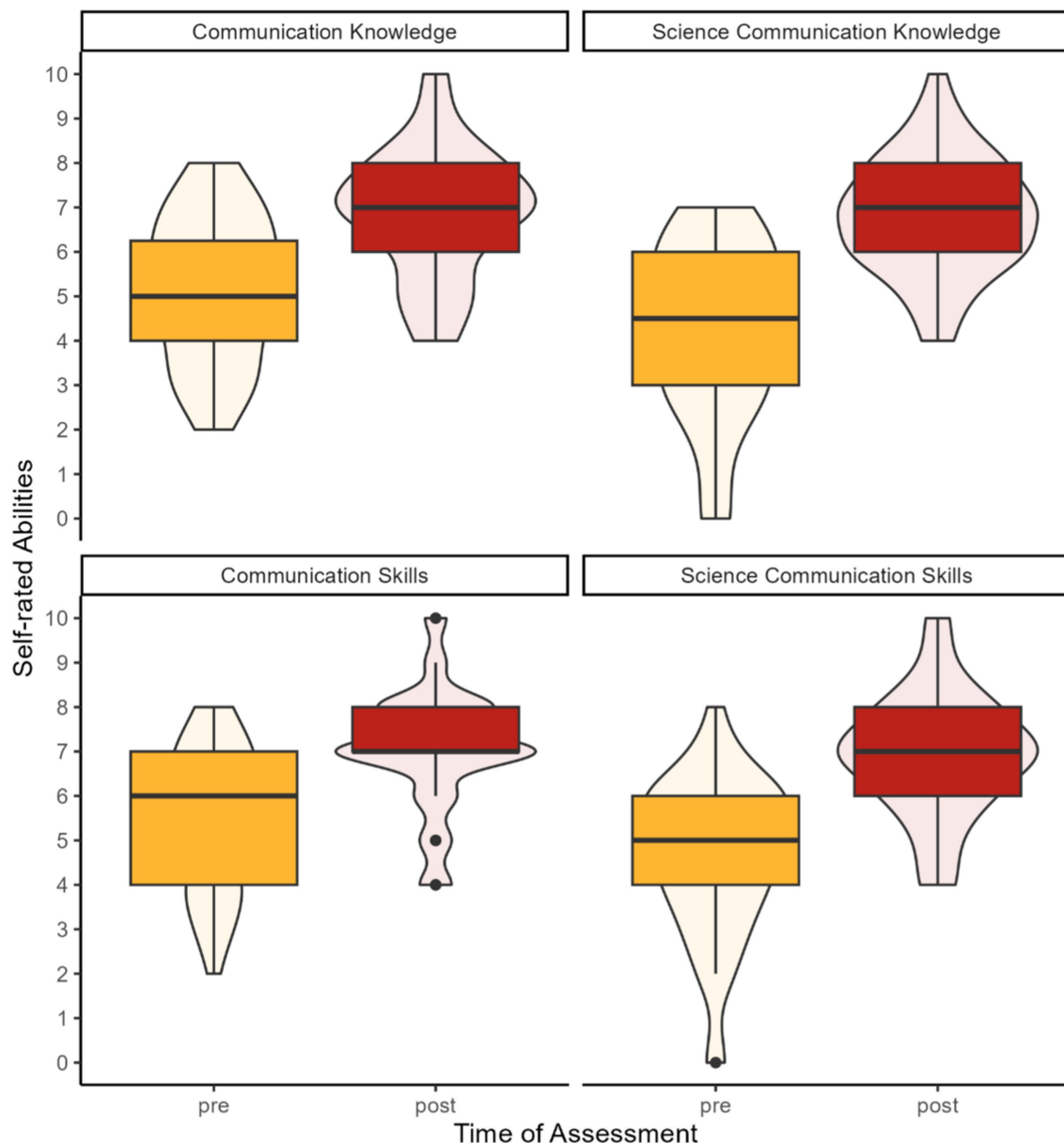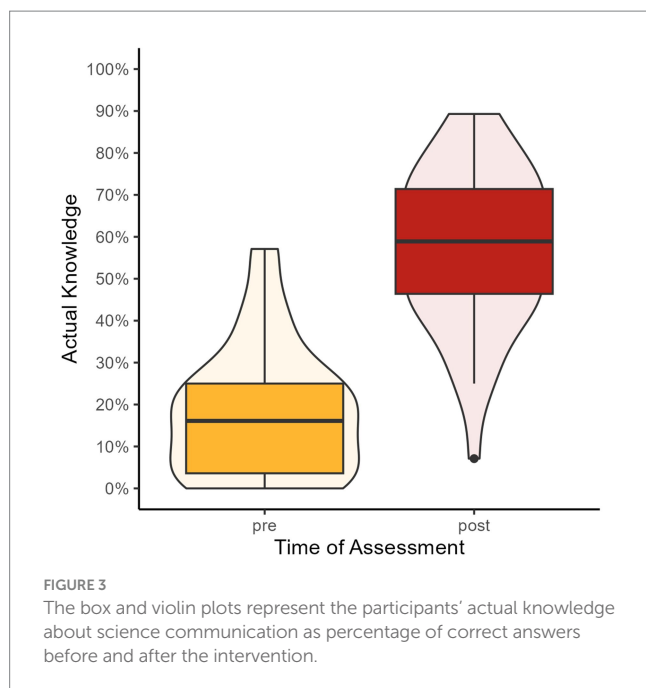
FIGURE 2
Box and violin plots showing self-rated abilities separated by the time of assessment. Self-rated knowledge on communication (left) and science communication (right) are presented at the top; practical skills in communication (left) and science communication (right) are presented at the bottom. Self-rated abilities were assessed using an 11-point scale ranging from *no skill/knowledge at all* (0) to *very high skill/knowledge* (10).

et al., 1999). As a result, producing oral language is a highly automatic process and changing practical communication behavior on the word or syntactic level might be a challenging proposition needing more than 1 day of training.

4.3 Limitations regarding the behavioral performance assessment

We assessed behavioral performance to figure out whether participants implemented the taught strategies when asked to communicate about science. Doing so, we faced some challenges that are reviewed in the following. First, coding text sequences within the

transcripts of two-minute presentations by two independent raters was quite effortful. We suggest to investigate whether some of these codings can be substituted by algorithms, such as methods to identify jargon (Liu and Lei, 2019). While we mainly used a judgment-based method that requires explicit decisions about whether a word is considered jargon, corpus-based methods like the De-Jargonizer (Rakedzon et al., 2017) use information about word frequencies to make this decision. This, however, is a rather technical operationalization that might ignore theoretical discussions about jargon and technical language (Roth, 2005). Second, we coded whether participants more often used strategies after the intervention and whether the number of usages of each strategy increased. The rationale behind this was that these strategies were derived from the

FIGURE 3
The box and violin plots represent the participants' actual knowledge about science communication as percentage of correct answers before and after the intervention.

literature because of their effectiveness in increasing either comprehensibility or involvement and that we assumed that simply using these strategies would have the same effect. However, we did not actually test these assumed effects on recipients that probably depend on the quality of strategy implementation (Spitzberg, 2015). For example, some participants applied the strategy of explaining jargon, but used other jargon words to do so and thus might not have achieved a substantial benefit. In this respect, our coding procedure did not account for the fact that some jargon words (e.g., "molecules") might be less problematic than others (e.g., "defensins"). The quality of strategy implementation, however, was not captured by our coding. To do so, rubrics could be used in future studies that either assess the performance as a holistic product or rate different aspects of the performance in an analytical way (Hunter et al., 1996). While ideas about the contents and criteria of such rubrics already exist (Sevian and Gonsalves, 2008; Rubega et al., 2021), the reliability and validity of these rubrics has to be examined before implementing them. This is of particular importance as criteria and ratings may differ by context (e.g., characteristics of the specific format and audience). Alternatively, the videos or transcripts could be shown to an audience which in turn rates how they perceive the products. This would provide evidence that the participants did not only learned to use certain strategies, but that they are actually capable of implementing them in an effective way and that they lead to the desired outcome (e.g., using comprehensibility strategies leads to an actual increase in perceived comprehensibility in the audience). The reasons for not including this impression perspective on competence is discussed in the next section.

4.4 Communicative ability between competence, performance and impression

Our broad definition of competence did include three components, competence in a narrow sense as the potential to act in

a skillful way, performance as the realization of competence, and impression as the quality perception of that performance. In our study, we empirically investigated the first two aspects, namely competence and performance.

As noted above, including the impression perspective would have been interesting to get information about the quality of the performances, However, we decided against including it for several reasons. First, there is not one impression or quality judgment but several, and "each of these perspectives permit, and may systematically imply, divergent types of competence evaluations" (Spitzberg, 2009, p. 73). For example, it is plausible to assume that comprehensibility judgments are dependent on general education level and specific prior knowledge on the topic. While the phrase "statistical model" would be easy to comprehend for a university professor in mathematics, it might be challenging to a person who has not studied mathematics.

In our performance assessment, we narrowed the intended audience down through giving the information that an old acquaintance should be addressed (introduced as not having a university degree but a vocational training in business). In principle, it would be possible to survey a sample that matches these characteristics on their perceptions of the recorded videos. This, however, would be cost and time intense, and would have provided further hurdles. Participants would have had to consent in the use of their training videos, further challenging their experience of a positive, non-competitive, and protected training experience.

A way to avoid this problem would be to fall back on existing data on what persons with certain socio-demographic characteristics probably know about science. However, the state of citizens' understanding of scientific facts, theories or methods has only been researched superficially so far (e.g., European Commission, 2021). Furthermore, even more systematic research findings would only give a rough indicator for inferring what communication partners probably know. Instead, it might be more fruitful to adapt the communication to the specific audience by using dialogic comprehensibility strategies like asking whether the other person knows a certain term or phenomenon (Clark and Brennan, 1991).

Third, there are also some validity issues regarding using impression ratings for evaluation. This concern mainly is about how well members of the audience really are in making quality judgments. For example, one of the most viewed TED talks (more than 74 million views by June 2025) was given by Amy Cuddy, who provided vivid examples and compelling narratives. However, the topic, power posing, is considered highly questionable within the scientific community (Simmons and Simonsohn, 2017). As this prominent example shows, quality judgments of science communication may be invalid, because scientific literacy and topic expertise is needed to judge some quality dimensions like accuracy or scientific integrity.

A similar issue concerns the slim line between being comprehensible and oversimplifying scientific contents. For example, the metaphor "brains are computers" is often used to explain how brains function. However, using this metaphor also may foster misconceptions, in this case a deterministic view of the brain, ignoring emotional or volitional processes (for a discussion, see Richards and Lillicrap, 2022). Thus, because of its persuasiveness, this metaphor might lead to an illusion of understanding in lay participants. Finally, evidence suggests that the assessment of competence through impression ratings might be biased because construct-irrelevant factors like voice pitch (Rodero, 2022) are able to affect responses.
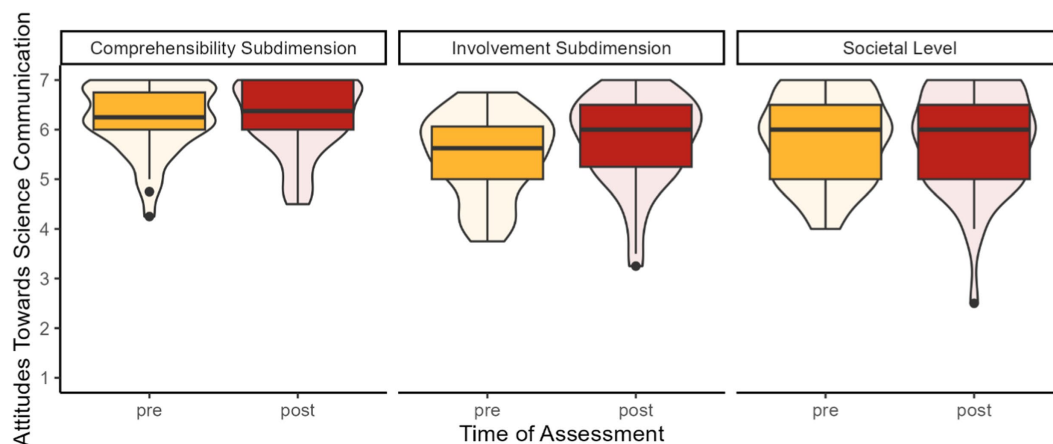
FIGURE 4
The boxplots show the three attitude scales, the self-developed scales of favorable attitudes toward using comprehensibility strategies (left) and involvement strategies (center) as well as attitudes regarding scientists' societal responsibility to communicate (right). Attitudes were assessed using a 7-point scale ranging from *totally disagree* (1) to *totally agree* (7).

4.5 Science communication competence aspects not covered

In regard to diversity trainings, there is a common differentiation between awareness- and skill-building trainings (Ferdman and Brody, 1996). While the former mainly focuses on fostering practical skills, and requires a value judgment about what behavior is qualitatively good (e.g., it is better to avoid jargon than to use it) the latter focuses on reflection processes. Having taught participants how to communicate in a comprehensible and involving way, our training had a clear focus on (basic) skills. As social cognitive theory suggests, affecting lasting change on practical skills is very time intense. Therefore, we strictly limited the goals covered in our one-day training program. As a result, we did not practice real dialogical science communication skills or skills needed to enable layperson's participation in science. These skills, however, seem important in order for science communication to move from a deficit toward an inclusive orientation (Vickery et al., 2023). Other relevant, more format-specific or advanced skills like recording a podcast, creating visualizations, or evaluating the own science communication were also not included.

Baram-Tsabari and Lewenstein (2017) reviewed existing science communication programs and listed potential learning objectives organized into six strands: affective, content, methods, reflective, participatory, as well as identity, and later elaborated on this competency catalogue by acknowledging the need to tailor it to different actors of science communication that have different demands regarding competencies (Lewenstein and Baram-Tsabari, 2022). Our training focused on building positive attitudes and building knowledge and skills about science communication (objectives from the affective, content and method strand), rather than the reflection, participation, and identity strand. The latter concern reflection on the aims and consequences of communication, active participation in communication activities, and identifying as communicating scientist, which we believe to increasingly become relevant in scientists' careers but are secondary for graduate students as we target in our training. Furthermore, we decided against

including other skills possibly relevant to science communication, such as project management (Baram-Tsabari and Lewenstein, 2017), skills for handling digital media (Fähnrich et al., 2021), or general presentation skills (Dudo et al., 2021) because we understand them as external prerequisites rather than core competencies of science communication.

4.6 A critical view on the evidential basis of our intervention

Our aim was not to just develop a new training program but for this training program to be based on scientific evidence. In our opinion, for an intervention to be evidence-based, its contents, instructional methods and evaluation procedures must be based on scientific theories and empirical evidence while new evidence for its effectiveness must be gathered by doing evaluations. Taking the example of the intervention's contents, it was quite difficult to distill those communication strategies that were actually based on evidence. Many studies show different, in part contradictory, results. Resolving these conflicts is challenging because of different operationalization of strategies and outcome measures, samples with different backgrounds, and often only sparse information about the exact methods (König et al., 2024). More systematic research is needed to identify what works for whom and under which conditions and whether some strategies may be advantageous for reaching one effect but disadvantageous regarding other outcomes. In doing so, it is not only crucial to gather empirical evidence, but to also develop theories that are able to explain them. This is important, because at some points, evidence and practice diverge. For example, many interventions cover strategies like storytelling and using narratives (e.g., Rodgers et al., 2018), although in their systematic review, König et al. (2024) conclude, that current evidence on this topic is quite inconsistent.

Conducting our summative evaluation, we collected evidence for the effectiveness of our intervention. However, we feel obliged to be precise about which evidence we produced – and which we did

not. Regarding the internal and statistical conclusion validity of our study (Shadish et al., 2002), which are important for assuming causality, we were able to rule out or minimize important threats like ambiguous temporal precedence, selection, and low statistical power. Other threats like history are implausible to affect the assumed causality in our case. As we had no control group, however, we were not able to rule out a third group of potential threats to internal validity. First, having used the same testing procedure twice (pre and post), the improvements may in part be due to repetition of the test rather than the intervention itself. Second, we could not account for short-term changes in participants that occurred independently of the intervention (i.e., maturation), such as arising bodily experiences like hunger, fatigue, and tiredness (Flannelly et al., 2018). Third, although we do not have concrete indications, attrition bias may have had an effect as well. Including a control group with random assignment in future studies is necessary to control for these threats to internal validity. Furthermore, we only claim that our intervention is effective compared to having had no treatment at all. Including a control group in future studies also enables us to investigate whether the intervention is more effective than a placebo intervention or even other science communication training programs.

The external and construct validity of our study (Shadish et al., 2002), relevant for generalizing the findings, may be threatened by trainer and sample effects. While we tried to minimize the former, that is, to maximize treatment fidelity (Swanson et al., 2013), through extensive preparation of the trainer as well as manualization and video recording of the intervention, we did not empirically investigate it. Similarly, there may have been effects of the sample composition or of participants' inter-individual differences. However, our sample was too small to explore such effects systematically. To give an example, we had the impression that the intervention was more effective when the sample composition was quite heterogeneous, that is, included participants with different disciplinary backgrounds. To increase our understanding of the underlying mechanisms, future studies should investigate the relationship between characteristics of trainees, trainer and instructional elements more closely and even manipulate them experimentally. Furthermore, as participation was voluntary, the participating students probably were highly motivated to learn more about science communication. Thus, obtained effect sizes may be smaller when the program is implemented in study programs more formally, so that students are obliged to participate. Lastly, as we did not include follow-up evaluations, we cannot make any claims about the stability of the achieved effects over time.

5 Conclusion

Overall, we gathered evidence for the effectiveness of our training program to foster basic science communication competence, more specifically self-efficacy, knowledge, favorable attitudes, and science communication performance. Furthermore, we introduced a concept of science communication competence and discussed challenges with operationalizing it. We addressed a need for better evaluated training programs (Baram-Tsabari and Lewenstein, 2017) by implementing a stringent evaluation based on a constructive alignment approach and reporting its results in a comprehensive and transparent way.

With developing, conducting, and evaluating our training program, we also wanted to contribute to the education of students in higher education and future scientists. Successfully completing a study program means to deeply dive into a scientific subject and its methods. We think it is also important to use the acquired competencies to solve complex, interdisciplinary problems, advice laypeople or educate novices. In all these cases, science communication can help being successful. As our intervention covered basic practical competencies, interventions targeting scientists in later career stages should also include reflection and identity as learning objectives (Lewenstein and Baram-Tsabari, 2022). As time in higher and researcher education is limited, it should be spent with attending training programs that have shown to be effective and that are themselves based on the best available evidence.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Open Science Framework: https://osf.io/4ypj3/.

Ethics statement

Ethical approval was not required for the studies involving humans because this study was conducted in line with the ethical guidelines of the American Psychological Association (APA; www.apa.org/ethics/code) as well as the German Psychological Association (DGPs; https://www.dgps.de/die-dgps/aufgaben-und-ziele/berufsethische-richtlinien/), and the General Data Protection Regulation (GDPR; http://data.europa.eu/eli/reg/2016/679/oj) of the European Union. We did not ask for formal ethical approval because the following conditions were met: (1) We did not collect data on vulnerable groups. (2) We did not deceive participants (no experimental conditions) or withheld information from them. (3) We did not have a control group, so all participants benefited from the intervention. (4) We actively provided the opportunity to participate in the intervention without simultaneously participating in the study. (5) We comprehensively informed participants about the study, the data collection and data protection measures and asked them to provide explicit informed consent before data were collected. 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

Funding

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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/feduc.2025.1558203/full#supplementary-material

References

Ahmed, S. Z., Hjorth, A., Rafner, J. F., Weidner, C. A., Kragh, G., Jensen, J. H. M., et al (2021). A training programme for early-stage researchers that focuses on developing personal science outreach portfolios. Available online at: https://arxiv.org/abs/2103.03109 (Accessed June 18, 2025).

Ajzen, I. (1985). "From intentions to actions: a theory of planned behavior" in Action control: from cognition to behavior. eds. J. Kuhl and J. Beckmann (Berlin, Heidelberg: Springer), 11–39.

Akin, H., Rodgers, S., and Schultz, J. (2021). Science communication training as information seeking and processing: a theoretical approach to training early-career scientists. J. Sci. Commun. 20:A06. doi: 10.22323/2.20050206

Bandura, A. (1989). Human agency in social cognitive theory. Am. Psychol. 44, 1175–1184. doi: 10.1037/0003-066X.44.9.1175

Baram-Tsabari, A., and Lewenstein, B. V. (2017). Science communication training: what are we trying to teach? Int. J. Sci. Educ. B 7, 285–300. doi: 10.1080/21548455.2017.1303756

Barel-Ben David, Y., and Baram-Tsabari, A. (2019). "Evaluating science communication training: going beyond self-reports" in Theory and best practices in science communication training. ed. T. P. Newman (New York: Routledge), 122–138.

Beierlein, C., Kemper, C. J., Kovaleva, A., and Rammstedt, B. (2013). Short scale for measuring general self-efficacy beliefs (ASKU). Methoden Daten Anal. 7, 251–278. doi: 10.12758/MDA.2013.014

Berger, C. R. (2003). "Message production skill in social interaction" in Handbook of communication and social interaction skills. eds. J. O. Greene and B. R. Burleson (New York: Routledge), 275–308.

Biggs, J. (1996). Enhancing teaching through constructive alignment. High. Educ. 32, 347–364. doi: 10.1007/BF00138871

Bobroff, J., and Bouquet, F. (2016). A project-based course about outreach in a physics curriculum. Eur. J. Phys. 37:045704. doi: 10.1088/0143-0807/37/4/045704

Bromme, R., and Goldman, S. R. (2014). The public's bounded understanding of science. Educ. Psychol. 49, 59–69. doi: 10.1080/00461520.2014.921572

Bromme, R., Rambow, R., and Nückles, M. (2001). Expertise and estimating what other people know: the influence of professional experience and type of knowledge. J. Exp. Psychol. Appl. 7, 317–330. doi: 10.1037/1076-898X.7.4.317

Brownell, S. E., Price, J. V., and Steinman, L. (2013). Science communication to the general public: why we need to teach undergraduate and graduate students this skill as part of their formal scientific training. J. Undergrad. Neurosci. Educ. 12, E6–E10.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376. doi: 10.1038/nrn3475

Cirino, L. A., Emberts, Z., Joseph, P. N., Allen, P. E., Lopatto, D., and Miller, C. W. (2017). Broadening the voice of science: promoting scientific communication in the undergraduate classroom. Ecol. Evol. 7, 10124–10130. doi: 10.1002/ece3.3501

Clark, R. C. (2020). Evidence-based training methods: a guide for training professionals. Alexandria, VA: ATD Press.

Clark, H. H., and Brennan, S. E. (1991). "Grounding in communication" in Perspectives on socially shared cognition. eds. L. B. Resnick, J. M. Levine and S. D. Teasley (Washington: American Psychological Association), 127–149.

Clark, G., Russell, J., Enyeart, P., Gracia, B., Wessel, A., Jarmoskaite, I., et al. (2016). Science educational outreach programs that benefit students and scientists. PLoS Biol. 14:e1002368. doi: 10.1371/journal.pbio.1002368

Clarkson, M. D., Houghton, J., Chen, W., and Rohde, J. (2018). Speaking about science: a student-led training program improves graduate students' skills in public communication. J. Sci. Commun. 17:A05. doi: 10.22323/2.17020205

Dayton, Z. A., and Dragojevic, M. (2024). Effects of jargon and source accent on receptivity to science communication. J. Lang. Soc. Psychol. 43, 104–117. doi: 10.1177/0261927X231191787

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychol. Rev. 93, 283–321. doi: 10.1037/0033-295X.93.3.283

Dresing, T., and Pehl, T. (2018). Praxisbuch interview, transkription & analyse: anleitungen und regelsysteme für qualitativ forschende. Marburg: Eigenverlag.

Druschke, C. G., Karraker, N., McWilliams, S. R., Scott, A., Morton-Aiken, J., Reynolds, N., et al. (2022). A low-investment, high-impact approach for training stronger and more confident graduate student science writers. Conserv. Sci. Pract. 4:e573. doi: 10.1111/csp2.573

Dudo, A., Besley, J. C., and Yuan, S. (2021). Science communication training in North America: preparing whom to do what with what effect? Sci. Commun. 43, 33–63. doi: 10.1177/1075547020960138

Duit, R., Gropengießer, H., Kattmann, U., Komorek, M., and Parchmann, I. (2012). "The model of educational reconstruction – a framework for improving teaching and learning science" in Science education research and practice in Europe. eds. D. Jorde and J. Dillon (Rotterdam: Sense Publishers), 13–37.

Eccles, J. S., and Wigfield, A. (2024). The development, testing, and refinement of Eccles, Wigfield, and colleagues' situated expectancy-value model of achievement performance and choice. Educ. Psychol. Rev. 36:51. doi: 10.1007/s10648-024-09888-9

European Commission. (2020). MORE4: support data collection and analysis concerning mobility patterns and career paths of researchers: survey on researchers in European higher education institutions. Available online at: https://data.europa.eu/doi/10.2777/132356 (Accessed July 3, 2024).

European Commission (2021). European citizens' knowledge and attitudes towards science and technology. Available online at: https://portugalglobal.pt/en/news/2025/february/eu-new-eurobarometer-survey/ (Accessed November 4, 2021).

European Commission (2022). The European competence framework for researchers. Brussels: European Commission.

Evers, A., Lucassen, W., Meijer, R., and Sijtsma, K.. (2015). Cotan review system for evaluating test quality. Available online at: https://nip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf. (Accessed January 25, 2023).

Fähnrich, B., Wilkinson, C., Weitkamp, E., Heintz, L., Ridgway, A., and Milani, E. (2021). Rethinking science communication education and training: towards a competence model for science communication. *Front. Commun.* 6:795198. doi: 10.3389/fcomm.2021.795198

Fick, J., Hendriks, F., Kumpmann, N., and Thies, B. (2025). Teaching Science Communication to Master's Students in STEM. *European Journal of Psychology Open*, 2673-8627/a000073. doi: 10.1024/2673-8627/a000073

Ferdman, B. M., and Brody, S. E. (1996). Models of diversity training. New York: Sage, 282–303.

Flannelly, K. J., Flannelly, L. T., and Jankowski, K. R. B. (2018). Threats to the internal validity of experimental and quasi-experimental research in healthcare. *J. Health Care Chaplain.* 24, 107–130. doi: 10.1080/08854726.2017.1421019

Fondel, E., Lischetzke, T., Weis, S., and Gollwitzer, M. (2015). Zur Validität von studentischen Lehrveranstaltungsevaluationen. *Diagnostica* 61, 124–135. doi: 10.1026/0012-1924/a000141

Fries, S., and Souvignier, E. (2015). "Training" in Pädagogische Psychologie. eds. E. Wild and J. Möller (Berlin, Heidelberg: Springer), 401–419.

Hagger, M. S., Cameron, L. D., Hamilton, K., Hankonen, N., and Lintunen, T. (2020). "Changing behavior: a theory- and evidence-based approach" in The handbook of behavior change. eds. K. Hamilton, L. D. Cameron, M. S. Hagger, N. Hankonen and T. Lintunen (Cambridge: Cambridge University Press), 1–14.

Hanauska, M. (2020). "Historical aspects of external science communication" in Science Communication. eds. A. Leßmöllmann, M. Dascal and T. Gloning (Berlin, Boston: De Gruyter Mouton), 585–600.

Holliman, R., and Warren, C. J. (2017). Supporting future scholars of engaged research. *Res. All* 1, 168–184. doi: 10.18546/RFA.01.1.14

Hunter, D. M., Jones, R. M., and Randhawa, B. S. (1996). The use of holistic versus analytic scoring for large-scale assessment of writing. *Can. J. Program Eval.* 11, 61–86. doi: 10.3138/cjpe.11.003

Klieme, E., and Hartig, J. (2008). "Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs" in Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft. eds. M. Prenzel, I. Gogolin and H.-H. Krüger (Wiesbaden: VS Verlag für Sozialwissenschaften), 11–29.

König, L. M., Altenmüller, M. S., Fick, J., Crusius, J., Genschow, O., and Sauerland, M. (2024). How to communicate science to the public? Recommendations for effective written communication derived from a systematic review Z. *Psychol.* 233, 40–51. doi: 10.1027/2151-2604/a000572

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

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776

Lewenstein, B. V., and Baram-Tsabari, A. (2022). How should we organize science communication trainings to achieve competencies? *Int. J. Sci. Educ. B* 12, 289–308. doi: 10.1080/21548455.2022.2136985

Liu, D., and Lei, L. (2019). "Technical vocabulary" in The routledge handbook of vocabulary studies. ed. S. Webb (New York: Routledge), 111–124.

Mayer, R. E., and Moreno, R. (2002). Aids to computer-based multimedia learning. *Learn. Instr.* 12, 107–119. doi: 10.1016/S0959-4752(01)00018-4

Montgomery, T. D., Buchbinder, J. R., Gawalt, E. S., Iuliucci, R. J., Koch, A. S., Kotsikorou, E., et al. (2022). The scientific method as a scaffold to enhance communication skills in chemistry. *J. Chem. Educ.* 99, 2338–2350. doi: 10.1021/acs.jchemed.2c00113

Paivio, A. (1971). "Imagery and language" in Imagery. ed. S. J. Segal (New York, NY: Academic Press), 7–32.

Pfeiffer, L. J., Knobloch, N. A., Tucker, M. A., and Hovey, M. (2022). Issues-360TM: an analysis of transformational learning in a controversial issues engagement initiative. *J. Agric. Educ. Ext.* 28, 439–458. doi: 10.1080/1389224X.2021.1942090

Pilt, E., and Himma-Kadakas, M. (2023). Training researchers and planning science communication and dissemination activities: testing the QUEST model in practice and theory. *J. Sci. Commun.* 22:2023. doi: 10.22323/2.220602042023

R Core Team (2021). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rakedzon, T., Segev, E., Chapnik, N., Yosef, R., and Baram-Tsabari, A. (2017). Automatic jargon identifier for scientists engaging with the public and science communication educators. *PLoS One* 12:e0181742. doi: 10.1371/journal.pone.0181742

Renkl, A., Mandl, H., and Gruber, H. (1996). Inert knowledge: analyses and remedies. *Educ. Psychol.* 31, 115–121. doi: 10.1207/s15326985ep3102_3

Richards, B. A., and Lillicrap, T. P. (2022). The brain-computer metaphor debate is useless: a matter of semantics. *Front. Comput. Sci.* 4:810358. doi: 10.3389/fcomp.2022.810358

Rodero, E. (2022). Effectiveness, attractiveness, and emotional response to voice pitch and hand gestures in public speaking. *Front. Commun.* 7:869084. doi: 10.3389/fcomm.2022.869084

Rodgers, S., Wang, Z., Maras, M. A., Burgoyne, S., Balakrishnan, B., Stemmle, J., et al. (2018). Decoding science: development and evaluation of a science communication training program using a triangulated framework. *Sci. Commun.* 40, 3–32. doi: 10.1177/1075547017747285

Rodgers, S., Wang, Z., and Schultz, J. C. (2020). A scale to measure science communication training effectiveness. *Sci. Commun.* 42, 90–111. doi: 10.1177/1075547020903057

Roth, W.-M. (2005). Talking science: language and learning in science classrooms. Lanham, MD: Rowman & Littlefield Publishers.

Rubega, M. A., Burgio, K. R., MacDonald, A. A. M., Oeldorf-Hirsch, A., Capers, R. S., and Wyss, R. (2021). Assessment by audiences shows little effect of science communication training. *Sci. Commun.* 43, 139–169. doi: 10.1177/1075547020971639

Sadler, T. D., Romine, W. L., and Topçu, M. S. (2016). Learning science content through socio-scientific issues-based instruction: a multi-level assessment study. *Int. J. Sci. Educ.* 38, 1622–1635. doi: 10.1080/09500693.2016.1204481

Sadoski, M., Goetz, E. T., and Rodriguez, M. (2000). Engaging texts: effects of concreteness on comprehensibility, interest, and recall in four text types. *J. Educ. Psychol.* 92, 85–95. doi: 10.1037/0022-0663.92.1.85

Schiebel, H., Stone, R., Rivera, E. A., and Fairfield, J. (2021). Developing science communication skills in early career scientists. *Limnol. Oceanogr. Bull.* 30, 35–38. doi: 10.1002/lob.10417

Seakins, A., and Fitzsimmons, A. (2020). Mind the gap: can a professional development programme build a university's public engagement community? *Res. All* 4, 291–309. doi: 10.14324/RFA.04.2.11

Sevian, H., and Gonsalves, L. (2008). Analysing how scientists explain their research: a rubric for measuring the effectiveness of scientific explanations. *Int. J. Sci. Educ.* 30, 1441–1467. doi: 10.1080/09500690802267579

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton: Mifflin and Company.

Shulman, H. C., Bullock, O. M., and Riggs, E. E. (2021). The interplay of jargon, motivation, and fatigue while processing COVID-19 crisis communication over time. *J. Lang. Soc. Psychol.* 40, 546–573. doi: 10.1177/0261927X211043100

Shulman, H. C., Dixon, G. N., Bullock, O. M., and Colón Amill, D. (2020). The effects of jargon on processing fluency, self-perceptions, and scientific engagement. *J. Lang. Soc. Psychol.* 39, 579–597. doi: 10.1177/0261927X20902177

Simmons, J. P., and Simonsohn, U. (2017). Power posing: *P*-curving the evidence. *Psychol. Sci.* 28, 687–693. doi: 10.1177/0956797616658563

Smith, J. G., and McPherson, M. L. (2020). A cross-campus professional development program strengthens graduate student leadership in environmental problem-solving. *Elementa* 8:085. doi: 10.1525/elementa.085

Spitzberg, B. H. (2009). Axioms for a theory of intercultural communication competence. *Annu. Rev. Engl. Lang. Teach.* 14, 17–29.

Spitzberg, B. H. (2013). (Re)Introducing communication competence to the health professions. *J. Public Health Res.* 2, 126–135. doi: 10.4081/jphr.2013.e23

Spitzberg, B. H. (2015). "22. Assessing the state of assessment: communication competence" in Communication competence. eds. A. F. Hannawa and B. H. Spitzberg (Berlin, München, Boston: De Gruyter Mouton), 559–584.

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., and McCulley, L. (2013). Intervention fidelity in special and general education research journals. *J. Spec. Educ.* 47, 3–13. doi: 10.1177/0022466911419516

Thielsch, M. T., and Hirschfeld, G.. (2010). Münsteraner Fragebogen zur Evaluation von Seminaren - revidiert (MFE-Sr). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). Available online at: https://zis.gesis.org/skala/Hirschfeld-Thielsch-M%C3%BCnsteraner-Fragebogen-zur-Evaluation-von-Vorlesungen-(MFE-V) (Accessed August 9, 2020).

Uttl, B., White, C. A., and Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning are not related. *Stud. Educ. Eval.* 54, 22–42. doi: 10.1016/j.stueduc.2016.08.007

Vickery, R., Murphy, K., McMillan, R., Alderfer, S., Donkoh, J., and Kelp, N. (2023). Analysis of inclusivity of published science communication curricula for scientists and STEM students. *CBE Life Sci. Educ.* 22:ar8. doi: 10.1187/cbe.22-03-0040

Weinert, F. E. (2001). "Concept of competence: a conceptual clarification" in Defining and selecting key competencies (Ashland, OH, US: Hogrefe & Huber Publishers), 45–65.