



OPEN ACCESS

EDITED BY

Gavin T. L. Brown,
The University of Auckland, New Zealand

REVIEWED BY

Ana Remesal,
University of Barcelona, Spain
Serafina Pastore,
University of Bari Aldo Moro, Italy

*CORRESPONDENCE

Morgane Senden
✉ morgane.senden@uclouvain.be

RECEIVED 31 March 2023

ACCEPTED 25 May 2023

PUBLISHED 16 June 2023

CITATION

Senden M, De Jaeger D and Coertjens L (2023)
Safe and sound: examining the effect of a
training targeting psychological safety and trust
in peer assessment.
Front. Educ. 8:1198011.
doi: 10.3389/feduc.2023.1198011

COPYRIGHT

© 2023 Senden, De Jaeger and Coertjens. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Safe and sound: examining the effect of a training targeting psychological safety and trust in peer assessment

Morgane Senden^{1*}, Dominique De Jaeger² and Liesje Coertjens¹

¹Psychological Sciences Research Institute (IPSY), Université Catholique de Louvain, Louvain-la-Neuve, Belgium, ²Faculty of Motor Sciences, Institute of Neuroscience (IoNS), Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Introduction: Peer feedback can be very beneficial for student learning in higher education, yet students may feel uncomfortable providing and receiving peer feedback: they may for example not feel safe in the group or have little trust in their peers' abilities to provide feedback. Surprisingly, only few studies have investigated how students' feelings of discomfort can be reduced. To fill this gap, we created a 1-h training session using active learning methods. The training focuses on enhancing students' perceptions of psychological safety, trust in their abilities and in their peers' abilities to provide feedback.

Methods: The efficacy of this training was tested using a quasi-experiment with pre- and post-test design. Third-year bachelor students in physical education participated in a peer feedback activity to fulfill the requirement of an obligatory course. In 2019–2020, 47 students participated in a peer assessment activity without specific training on psychological safety and trust (control group), while in 2021–2022, 42 students received specific training before peer assessment (experimental group).

Results: Analyses include a comparison of the control and experimental groups with regard to (1) the evolution of their perceptions (psychological safety, trust in their abilities, and trust in their peers' abilities) for pre- to post-test, (2) the quality of the feedback they provided to their peers (3) and the improvement of students' work between the draft submitted for the peer activity and the final version submitted to the professor.

Discussion: Results do not support the training's efficacy, yet suggest pathways for future research.

KEYWORDS

peer assessment, psychological safety, trust, training, higher education, interpersonal factors, quasi-experiment, intervention

1. Introduction

Peer assessment is a practice that can contribute to achieving various goals of higher education. In addition to being an important tool for higher education students' learning, the use of peer assessment is in line with a more participatory culture of learning (Kollar and Fischer, 2010). Involving students in the assessment pushes them to take responsibility for their learning and to develop important skills such as self-regulatory practices (Carless et al., 2011; Planas Lladó et al., 2014). These practices are necessary to attain an important goal of higher education: enable students to become independent lifelong learners (Tai et al., 2018).

Topping (1998, p. 250) defined peer assessment as “an arrangement in which individuals consider the amount, level, value, worth, quality or success of the products or outcomes of learning of peers of similar status.” Several recent meta-analyses have shown that peer assessment not only has a positive impact on learning compared to the absence of any sort of assessment but also that it impacts student learning more positively than teacher assessment does (Double et al., 2020; Li et al., 2020; Wisniewski et al., 2020).

While peer assessment is an activity with a lot of potential for students’ learning, it is also an activity full of complexities, which involves interpersonal and intrapersonal factors (Panadero et al., 2023). Even though most students perceive the educational value of peer assessment (Mulder et al., 2014), they also express a series of concerns. Students may feel uncomfortable in both the assessee and the assessor’s roles. As assessor, they may not feel competent or they may find it difficult to be objective, while as an assessee they may fear their peers will be biased or will judge them (Hanrahan and Isaacs, 2001; Mulder et al., 2014; Wilson et al., 2015).

Although it is known that interpersonal and intrapersonal factors involved in peer assessment are important to consider because they can affect peer assessment activities (Panadero et al., 2023), few studies have investigated if an intervention could positively affect them. Moreover, it is not clear if the impact of interpersonal and intrapersonal factors on peer assessment outcomes occurs through an impact on the quality of received and provided feedback. To fill this gap, this study aims to investigate if a training targeting psychological safety and trust can enhance students’ perceptions, incites them to provide more elaborated feedback, and supports them to benefit more from the peer assessment. In the next sections, we will define the three factors this study focuses on, and detail research findings on their role in peer assessment and on how we could intervene.

1.1. The impact of psychological safety and trust in peer assessment

Trust in the self as an assessor, trust in the other as an assessor, and psychological safety will be the focus of this study. These three factors are strong predictors of peer assessment outcomes (van Gennip et al., 2010).

1.1.1. Psychological safety

The notion of psychological safety originally came from organizational psychology (Edmondson and Lei, 2014). In the working environment, where most of the research on psychological safety has been conducted, an emphasis is placed on the fact that psychological safety enables employees, teams, and organizations to learn (Edmondson and Lei, 2014). Therefore, it seems logical that the concept of psychological safety was subsequently studied in educational contexts. Psychological safety creates an environment where students feel comfortable discussing their performance and errors and asking for feedback which has a positive impact on their learning and their grades (Soares and Lopes, 2020). From there it becomes clear why psychological safety is a requirement for peer assessment. In the context of peer assessment, psychological safety is defined as “the extent to which students feel safe to give sincere feedback as an assessor and do not fear receiving inappropriate negative feedback or to respond to negative feedback” (Panadero et al., 2023, p. 5).

Students’ perceptions of psychological safety are a direct predictor of students’ conceptions of peer assessment and an indirect predictor of students’ perceived learning (van Gennip et al., 2010). Moreover, students with higher perceptions of psychological safety adopt deeper learning approaches in peer assessment and hold more cohesive conceptions of learning, which mean that they do not only see learning as an accumulation of knowledge but as an association of new content with prior knowledge (Cheng and Tsai, 2012). However, psychological safety is not associated with the perceived educational value of peer assessment (Rotsaert et al., 2017).

1.1.2. Trust in the self as an assessor and trust in the other as an assessor

Trust in the self as an assessor is defined as the “belief about the ability to perform peer assessment as assessor,” while trust in the other as an assessor is defined as the “confidence in a peer’s capability to perform a fair and/or accurate peer assessment” (Panadero et al., 2023, p. 5). A student can trust his or her ability but not his or her peers’ ability, or vice versa. It seems that most students tend to trust their abilities to assess their peers, while a smaller percentage of students also trusted their peers’ abilities (Cheng and Tsai, 2012).

Like psychological safety, trust in the self as an assessor and trust in the other as an assessor both positively predict students’ conceptions of peer assessment (van Gennip et al., 2010) and are associated with deeper learning approaches (Cheng and Tsai, 2012). However, only trust in the self is associated with more cohesive conceptions of learning. Moreover, students’ perceptions of trust in the self as an assessor and trust in the other as an assessor both affect their implication in peer assessment (Zou et al., 2018) and are associated with a higher perceived educational value of peer assessment (Rotsaert et al., 2017).

1.2. Interventions targeting psychological safety and trust

The relationships between levels of trust and psychological safety and peer assessment processes or outcomes, even though only correlational, highlight the important role that these perceptions can play. This raises the question of how to ensure that students feel psychologically safe and have trust in their peers and the self when participating in peer assessment.

The main suggestion in literature to overcome students’ concerns related to peer assessment is to use anonymity. Indeed, evidence suggests that students appreciate anonymity. They feel more comfortable during anonymous peer assessment, both as an assessor and as an assessee (Su, 2022), and they would rather be anonymous or use a nickname than use their real name (Yu and Liu, 2009). Moreover, anonymity can have positive impacts on students’ perceptions, such as reducing peer pressure and fear of disapproval (Vanderhoven et al., 2015). Although the use of anonymity in peer assessment has some positive effects, to the best of our knowledge, no study shows that anonymity has a positive impact on psychological safety and trust specifically. Studies have either found an absence of impact (e.g., Rotsaert et al., 2018) or even a negative one: a study found that pupils viewed their assessors more positively (which includes a higher trust in them) when real names were used in the peer assessment than when assessors were anonymous or used nicknames (Yu and Wu,

2011). According to Panadero and Alqassab's (2019) review of the effects of anonymity in peer assessment, anonymity seems to positively affect students' perceptions related to the peer assessment activity, but negatively affects their perceptions related to interpersonal factors, such as trust and psychological safety.

To the best of our knowledge, besides anonymity, the only existing intervention that aims to increase students' perceptions of psychological safety and trust before or during peer assessment is the use of role-playing described by Ching and Hsu (2016). In their study, students had to choose a stakeholder's role during peer assessment and provided feedback to their peers from this stakeholder's perspective. Students reported a high level of psychological safety and trust during this kind of peer assessment and they provided feedback of high quality. However, as there were no control group or pre-test measures, it remains up to present unclear if students' level of trust and psychological safety would have been lower without the role-play. Moreover, psychological safety and trust were measured with different types of items than in other studies (e.g., van Gennip et al., 2010; Rotsaert et al., 2018), and may not reflect the same concepts. In addition, the peers' work that had to be assessed in Ching and Hsu (2016)'s study is well suited to be assessed while adopting a stakeholder's role, but for other types of student productions (e.g., mathematical problem solving), it could be less adapted and the use of role-play might not work in these cases. If role-play is not possible or relevant during peer assessment, it could be included in a training session that would take place before peer assessment.

There is evidence that including a training session is an effective way to ease some tensions related to peer assessment. Students who followed a training that aimed to control the peer pressure they can feel during a peer feedback activity saw more value in the peer assessment and experienced less peer pressure than the students who did not follow it (Li, 2017). To the best of our knowledge, no training has ever been designed to enhance students' perceptions of psychological safety or trust in the context of peer assessment. For psychological safety, which is important in various situations, inspiration can be found beyond this specific context. A study in the context of group work found no effect of a 50-min training session that aimed to increase students' perceptions of psychological safety before working in small groups (Dusenberry and Robinson, 2020). This absence of effect could be explained by two factors: the training was not context-specific and the consisted mostly of non-active learning methods. By overcoming these limits, effective training for psychological safety could be developed.

A training enhancing perceptions of trust and psychological safety could also positively affect students' learning and performances. Indeed, in Li (2017)'s study, students who had participated in the training obtained higher scores for their revised work compared to students who did not receive the training. The mechanisms by which students' perceptions of psychological safety and trust could affect learning gains following peer assessment are unclear, due to a lack of research linking interpersonal and intrapersonal factors and peer assessment learning outcomes (Panadero et al., 2023).

1.3. The present study

This study aims to investigate if students' perceptions of psychological safety, trust in the self as an assessor, and trust in the

other as an assessor can be increased through a training targeting these factors. To this end, we designed a 1-h training session and provided it to a group of students before they participated in a peer assessment activity. This cohort of students was compared to a cohort of students who did not receive the training. We assumed that providing a training on these aspects would increase students' perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor.

Our first hypothesis is therefore:

H1: Perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor increase more in the experimental group than in the control group.

Another aim of this study is to investigate the relationship between students' perceptions of psychological safety and trust and their learning gains following peer assessment. Indeed, there is a lack of research on how students' perceptions can hinder or facilitate learning through peer assessment (Panadero et al., 2023). It is known that intrapersonal and interpersonal factors can affect students' engagement in peer feedback (Zou et al., 2018), but not how this engagement will affect students' learning gains. Prior research has found that there is a relationship between the quality of feedback students provide and the improvement in their final work following peer assessment (Li et al., 2010). It could mean that an increase in trust and psychological safety would affect students' implication in the peer assessment, which would be reflected by the quality of feedback provided during the peer assessment and by an improvement in performance due to the peer assessment. Thus, our second and third hypotheses are the following:

H2: Students in the experimental group provide feedback of higher quality to their peers than students in the control group.

H3: Students' grades improve more after the peer assessment activity in the experimental than in the control group.

2. Materials and methods

2.1. Study design and participants

To test the effect of the training on psychological safety and trust, we designed an intervention study, using a quasi-experiment with pre and post-test design. The first cohort of students followed the course in 2019–2020. These students were the control group; they participated in the peer assessment as it was implemented in the course, without any modification. The second cohort of students followed the course in 2021–2022. These students were the experimental group; before participating in the peer assessment, they received specific training on psychological safety and trust.

The participants were third-year physical education students enrolled in a course on acrobatic sports didactics at a French-speaking Belgian university. In physical education, an important part of courses is practical, and students are used to (1) being active, (2) interacting with each other, and (3) receiving formative feedback from peers and professors. Fifty-one students followed the course in 2019–2020, and 56 students followed it in 2021–2022. Of these 107 students, 100 (94%)

agreed to participate in the study, but among them, 11 were absent during the training session or did not participate in the peer assessment activity (both mandatory for the course). Therefore, the sample consisted of 89 students, 47 in the first cohort (control group) and 42 in the second cohort (experimental group). Sixty-nine percent of the participants were men. On average, participating students were 21.5 years old ($SD = 1.79$). Most participants (89%) indicated having already participated in peer assessment before, yet the professor of the course confirmed that students have not received training prior to this. There was no difference in terms of gender, age, or prior experiences with peer assessment between the control and the experimental groups.

2.2. Procedure and intervention

The study took place in the context of a course on the didactics of acrobatic sports. This course was built on prerequisites that students have acquired in previous courses (e.g., gymnastics, biomechanics), but students were novices in didactics. In parallel, they followed courses on other sports' didactics. The course aimed to teach students how to develop relevant learning situations in acrobatic sports. To this end, students had to create an instruction sheet describing a learning exercise. This sheet consisted of a diagram illustrating the exercise, one or two instructions explaining how to perform it, and one or two criteria for success (an example is provided in [Supplementary material](#)). They also had to create a video of the exercise.

After completing the first draft of their sheet, students participated in the peer assessment activity on Moodle. Using a rubric created by the professor (second author), each student assessed the instruction sheet of seven of their peers (randomly assigned). The rubric is available in [Supplementary material](#). For each criterion (diagram, video, instructions, success criteria, and overall coherence), students chose a level in the rubric and wrote qualitative feedback. After the peer assessment activity, students could improve their sheet before submitting it for the final assessment by the course's professor. A description of all design elements of the peer assessment is available as [Supplementary material](#) using [Panadero et al. \(2023\)](#)'s reporting instrument.

As a course requirement, students from both groups participated in a 1-h session that prepared them for the peer assessment: the procedure of the peer assessment activity was detailed, they learned how to give effective feedback based on [Hattie and Timperley \(2007\)](#)'s feedback model, and they had the opportunity to practice in class. Directly after this session, a researcher (first author) provided the training to the experimental group. This 1-h training aimed to increase students' perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor.

The training included three stages (see [Figure 1](#)). The first stage took place a week before the training session. To discover students' thoughts on peer assessment and adapt the training based on it, we asked them to answer a few questions (such as "What are the disadvantages of peer assessment?") using an interactive platform. The second stage included role-plays and discussions, two methods recommended to transform attitudes ([Blanchard and Thacker, 2012](#)). For the role-plays, students were split into groups of six and received a detailed roadmap with two role-play scenarios and instructions on how to play these scenarios and how to discuss them afterward. After

performing and discussing the two role-plays, they stayed in sub-groups to summarize their discussions. The third and final stage was an open discussion with the entire class to summarize all the ideas. Based on the discussion, we created a mind map that students received a few days later. A detailed presentation of the training is available in the [Supplementary material](#). The presence of a training session for the experimental group was the only difference between the groups, otherwise, the procedure was identical (see [Figure 2](#)). Students filled out questionnaires about their perceptions (pre-test and post-test) during the first and the last sessions of the course.

2.3. Measurements

2.3.1. Survey data

Students' perceptions (trust in the self as an assessor, trust in the other as an assessor, psychological safety, and the importance of anonymity) were measured with a French translation of the scales of [Rotsaert et al. \(2018\)](#). These scales have been shown to be reliable and their items operationalize well the concepts as defined in the introduction of this study. Items were measured using a seven-point Likert scale ranging from 1 (totally disagree) to 7 (totally agree). Macdonald's ω suggest good reliability (see [Table 1](#)).

2.3.2. Students' feedback

To operationalize the quality of feedback, we used the distinction between verification and elaboration feedback ([Shute, 2008](#)). Verification feedback only tells students if their productions are correct or not, while elaboration feedback contains additional information to help students to arrive at the correct answer. Effective feedback contains both verification and elaboration ([Shute, 2008](#)).

To compare feedback provided by students from the control group to those provided by students from the experimental group, feedback from all students was coded. First, we divided each feedback into units of analysis following [Strijbos et al. \(2006\)](#)'s procedure. This resulted in a database of 2,874 units of meaning. Each unit of meaning was then coded according to the criterion it refers to (examples are provided in [Table 2](#)) and to the feedback style (examples are provided in [Table 3](#)). To code feedback according to their style (verification or elaboration), we relied on the coding scheme developed by [Alqassab et al. \(2018\)](#) that we adapted to the context of the present study. According to this scheme, elaboration feedback can be one of five types: correction, confirmation, justification, question, or suggestion. Given the very small number of meaning units belonging to the "confirmation" category, this category was omitted. Some units of meaning did not fit into any of these categories (e.g., "Good luck for the next step!") and were classified as "other."

To test inter-rater agreement, a second coder coded a random subsample of 287 units of meaning (i.e., 10% of the total). As the verification and correction feedback were more frequent than the other types, resulting in unbalanced marginals, the Gwet's AC_1 is more reliable than the Cohen's κ ([Black et al., 2016](#)). The Gwet's AC_1 indicated a moderate to substantial agreement [Gwet' $AC_1 = 0.61$ (95% CI, 0.546–0.674), $p < 0.001$] ([Gwet, 2014](#)).

2.3.3. Students' grades

To investigate the effect of the training session on students' performance, the course professor assessed both the first draft and the

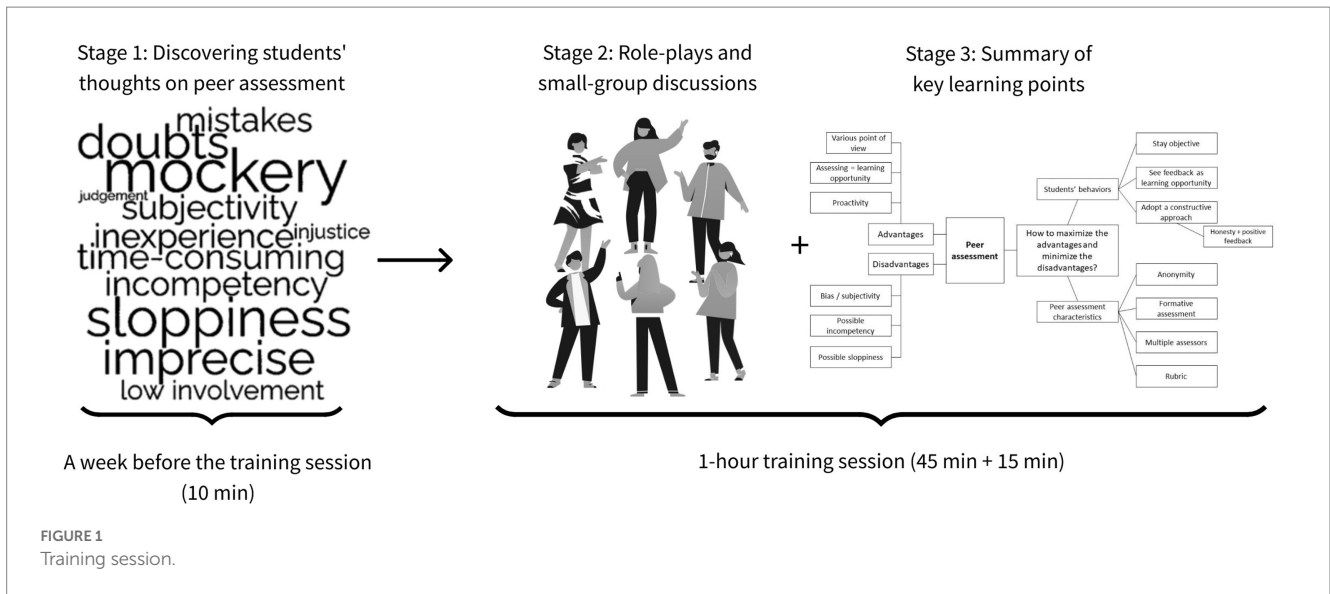
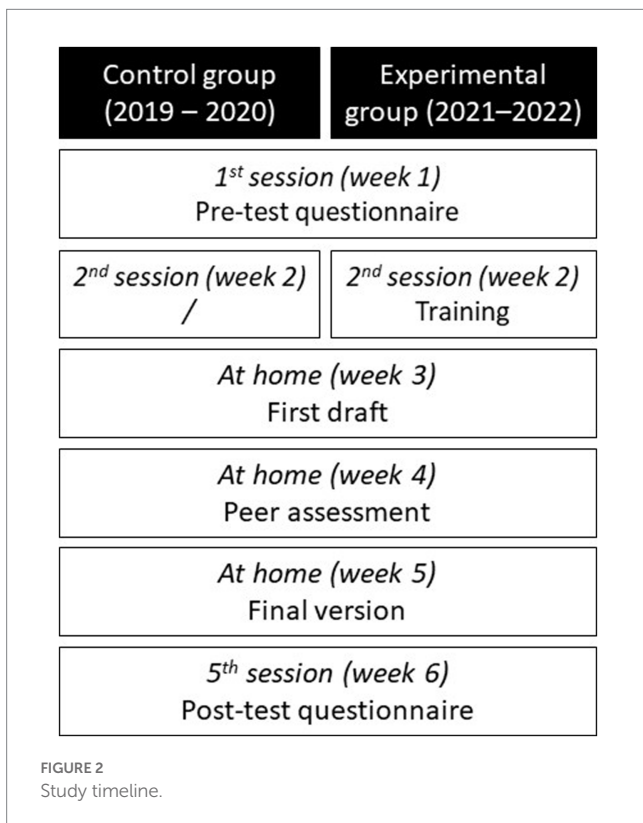


FIGURE 1 Training session.



final version of students' instruction sheets. The first draft was assessed for research purposes only, students did not receive their grades and it did not affect whether they passed the course or not. The criteria provided to the students were slightly different from those that the professor used, in order to guide them more in the peer assessment. For example, the video, which was only there to help the global comprehension of the exercise, was in students' criteria to ensure they would watch it. The criteria used by the professor to grade both versions were the following: "diagram," "relevance of the situation," "instructions," and "success criteria." Each criterion was scored out of 5, giving an overall score out of 20.

2.4. Data analysis

2.4.1. Survey data

There were some missing data for the questionnaire. In each cohort, some students were absent (e.g., due to illness) at the pre or post-test. We created an online version of the questionnaire and shared the link with absent students, but nevertheless, there was 17% of missing data at the pre-test and 7% at the post-test. There were no item-level missing data.

To handle these missing data, we used multiple imputations, which is considered the gold standard method when data are missing at random (Enders, 2010). The variables used for the imputations were the age, gender, and group (experimental or control), as well as the perceptions available (if the data were missing for the pre-test, we used post-test data and inversely). The analyses were conducted in R (version 4.1.0) with the "mice" package. Following von Hippel's (2020) two-stage calculation, we calculated that we needed at least 22 imputations. Therefore, we did 30 imputations. We used Rubin's rules (Rubin, 1987) to pool the results.

Once the data were imputed and pooled, linear mixed-effects models were conducted to test our hypothesis according to which an interaction effect would be present, with a higher increase in the experimental group.

2.4.2. Students' feedback

To compare the proportion of feedback provided by students in the control group and the experimental group, we calculated a score between 0 and 1 for each student regarding each type of feedback. This score was calculated by dividing the number of feedback elements provided by the total number of feedback units that could have been provided. A score of 0 means that the student did not provide any feedback of this type, while a score of 1 means that the student provided feedback of this type at every opportunity. For example, a student who assessed seven peers on six criteria and provided 34 verification feedback had a score of 0.81 (34/42) for verification feedback. To investigate the impact of the training on the type of feedback given, Student's *t*-tests and Mann-Whitney's *U* tests were conducted on Jamovi (version 2.2.5).

TABLE 1 Reliability of survey scales.

Scales	Number of items	Macdonald's ω			
		Control group		Experimental group	
		Pre-test	Post-test	Pre-test	Post-test
Psychological safety Example item: In this group, I can share my opinion without hesitation	4	0.91	0.91	0.86	0.90
Trust in the self as an assessor Example item: I feel I am able to assess my peers	7	0.92	0.85	0.90	0.93
Trust in the other as an assessor Example item: I think my peers are able to give me objective feedback	6	0.69	0.85	0.88	0.82

TABLE 2 Example of feedback for each criterion.

Criterion	Example
Diagram	For the shape of the diagram, there is just the fact that one pushes on the palm of the hand and not on the fingers
Video	Properly executed movement
Overall coherence	Overall coherence: Good, nothing to say, simple and clear
Instructions	The instructions are very clear and precise, we know at what moment to push the shoulders and to pull the heels
Success criteria	Another criterion for success that you could have used instead is to reach far with your hands on the jumping table
Global	Great work!

TABLE 3 Example of feedback for each type.

Type	Example
Verification	Success criteria: very good
Correction	I think it would be more correct if the arms (Figures 2, 3) and legs (Figure 3) were aligned
Justification	[To say "maintain muscular tension" is not an instruction] in the sense that in almost all exercises you have to maintain muscular tension
Question	Where do we need to jump? And how?
Suggestion	I would have written for example: "do not exceed 1/3 of the big mattress"

2.4.3. Students' grades

Given that students' grades did not follow a normal distribution, we conducted Wilcoxon tests to investigate if there was an improvement in students' performance. These analyses were conducted on Jamovi (version 2.2.5). Then, to test if the improvement was more important in the experimental group than in the control group, we calculated difference scores between the two versions (for the global version and each criterion) and conducted Mann-Whitney's U tests to compare the difference scores between the control and experimental groups.

Spearman correlations between all study variables are presented in Table 4.

3. Results

3.1. Impact of the training on students' perceptions

The mean scores and standard deviation of students' perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor are displayed in Table 5.

The level of psychological safety, trust in the self as an assessor, and trust in the other as an assessor were high at the

pre-test. It means that students already felt safe and had positive perceptions of their own assessment abilities and those of peers before the training and the peer assessment activity.

3.1.1. Psychological safety

For psychological safety, the results show an absence of a significant effect. Contrary to our hypothesis, there was no significant interaction effect [$t(1432.829) = 1.181, p = 0.238$]. Moreover, the two groups were comparable [$t(14187.907) = -0.957, p = 0.338$] and, in both groups, the levels of psychological safety were not altered significantly from pre-test to post-test [$t(362.903) = 1.161, p = 0.246$].

3.1.2. Trust in the self as an assessor

Contrary to our hypothesis, there was no significant interaction effect for trust in the self as an assessor [$t(1377.802) = -1.164, p = 0.245$]. It is important to note that the two groups were not comparable. Levels of trust in the self as an assessor were higher in the experimental than in the control group [$t(8621.171) = -2.054, p = 0.040$]. In both groups, levels of trust in the self as an assessor did not change statistically from pre-test to post-test [$t(359.655) = 1.327, p = 0.185$].

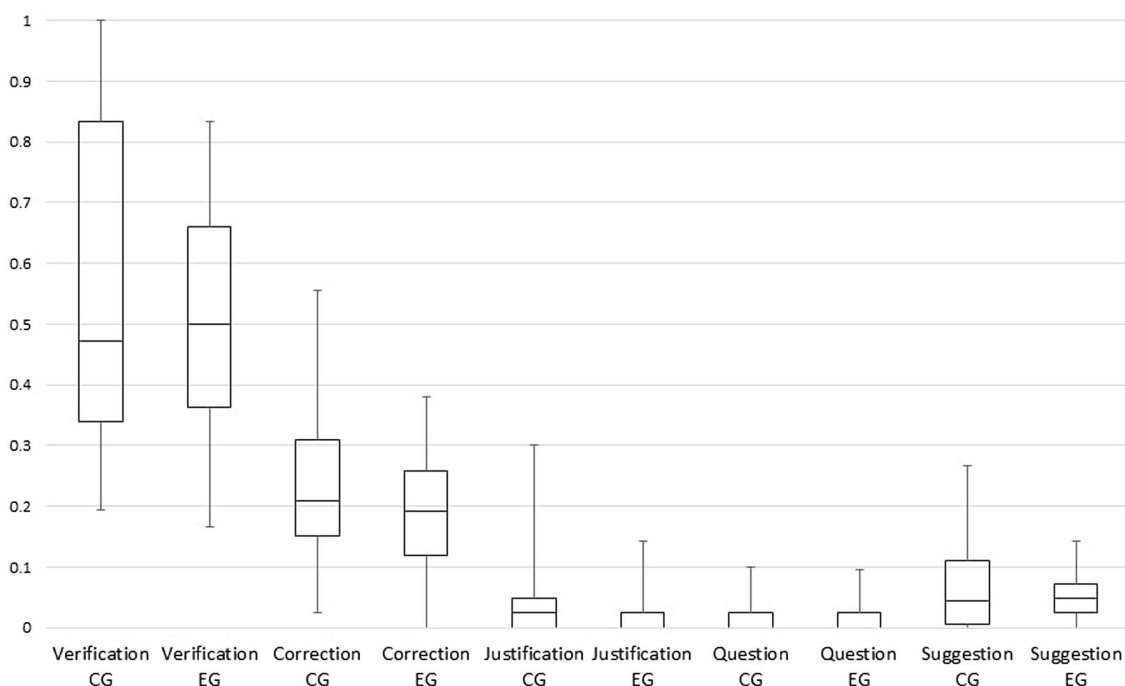


FIGURE 3
Box-plots of students' scores for each feedback type in the control group and the experimental group. CG, Control Group; EG, Experimental Group.

3.1.3. Trust in the other as an assessor

There was no significant interaction effect for trust in the other as an assessor [$t(980.787) = 1.282, p = 0.200$]. As for trust in the self as an assessor, the two groups were not comparable, but the difference was in the other direction: levels of trust in the other as an assessor were higher in the control than in the experimental group [$t(867.910) = -5.997, p < 0.001$]. In both groups, levels of trust in the other as an assessor did not change statistically from pre-test to post-test [$t(798.583) = -0.422, p = 0.673$].

3.2. Impact of the training on the feedback provided by students

As shown in Figure 3, the pattern of provided feedback is rather similar for both groups. Verification feedback is frequently provided, but with important variability among students. Correction feedback is the second most frequent feedback, again with quite some variability. The other types of feedback (justification, question, and suggestion) are rarely provided.

Scores for verification, justification, question, and suggestion did not follow a normal distribution. Therefore, some Mann–Whitney U tests were conducted to test for differences between the two groups. There was no significant effect for verification feedback ($U = 1,030, p = 0.472$), question feedback ($U = 1,072, p = 0.632$), and suggestion feedback ($U = 963, p = 0.218$). However, there was a significant effect for justification feedback ($U = 840, p = 0.020$), but this effect was in the opposite direction compared to our initial hypotheses: students in the control group provided more feedback of this type than students in the experimental group did. A t -test was conducted for correction feedback given that, for this type of feedback, the scores followed a normal distribution. The results were similar to those for justification feedback:

students in the control group provided more correction feedback than students in the experimental group [$t(93) = 2.13; p = 0.036$].

3.3. Impact of the training on students' grades

First, the impact of the training on students' performance was investigated with the global grade. Figure 4 illustrates the evolution of students' grades for both groups.

We started by testing if, for both groups taken together, there was an improvement in performance using a Wilcoxon test. It was the case ($W = 546, p < 0.001, r = -0.491$), which confirmed that students' work in both groups improved after the peer assessment.

A Mann–Whitney U test was conducted to investigate if the improvement was more important in the experimental than in the control group. The effect was not significant ($U = 915, p = 0.548$), implying that we could not show that the grade improvement was different in the experimental ($M = 0.29, SD = 0.74$) and in the control group ($M = 0.40, SD = 1.03$).

Observing Figure 4, it seems like students from the control group performed better than those in the experimental group. Mann–Whitney's U tests were conducted to check if these differences were significant. Students from the control group received a higher grade for their first draft ($U = 588, p < 0.001$) and for their final version ($U = 559, p < 0.001$).

Secondly, the analyses were conducted for each criterion. Wilcoxon tests' results indicated that there was a significant grade improvement for the criteria "diagram" ($W = 347, p = 0.003$) and "relevance of the situation" ($W = 70, p = 0.006$), but not for the criteria "instructions" ($W = 365, p = 0.167$) and "success criteria" ($W = 354, p = 0.599$). The

TABLE 4 Correlation matrix.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Psychological safety pre-test															
2. Psychological safety post-test	0.76**														
3. Trust self pre-test	0.40**	0.37**													
4. Trust self post-test	0.53**	0.68**	0.63**												
5. Trust other pre-test	0.12	0.05	-0.06	-0.2											
6. Trust other post-test	0.29*	0.18	-0.10	0.13	0.65**										
7. Improvement global grade	-0.09	-0.18	-0.02	-0.12	0.06	-0.12									
8. Improvement grade "diagram"	-0.18	-0.04	-0.08	0.00	-0.08	-0.15	0.67**								
9. Improvement grade "relevance of the situation"	0.04	0.03	0.12	0.09	0.40**	0.20	0.19	0.01							
10. Improvement grade "instructions"	-0.05	-0.16	-0.12	-0.16	-0.04	-0.26	0.51**	0.15	-0.11						
11. Improvement grade "success criteria"	0.13	-0.12	0.13	-0.10	0.06	0.01	0.53**	0.01	0.04	0.11					
12. Verification feedback	0.00	-0.05	-0.05	-0.08	-0.03	0.04	-0.08	-0.04	-0.11	0.02	-0.14				
13. Correction feedback	-0.03	-0.07	-0.11	-0.15	-0.10	0.00	-0.06	-0.01	-0.19	-0.01	-0.03	0.57**			
14. Justification feedback	0.05	0.03	0.00	-0.02	0.18	0.24*	-0.02	0.03	-0.02	-0.06	0.04	0.36**	0.29**		
15. Question feedback	-0.12	0.03	0.02	0.10	-0.11	0.13	-0.24*	-0.14	-0.15	-0.07	-0.20	0.10	0.21*	0.11	
16. Suggestion feedback	0.04	0.00	-0.14	-0.12	-0.02	0.18	0.03	0.09	-0.17	0.09	0.02	0.30**	0.19	0.37**	0.19

* $p < 0.05$; ** $p < 0.01$.

difference score between the two versions was calculated for the two criteria with a significant improvement and Mann–Whitney U tests were conducted to investigate if the improvement was different in both groups (see Table 6). There is only a significant difference for the criterion "relevance of the situation." This difference is in the opposite direction of our hypothesis: students in the control group improved more than students in the experimental group did.

4. Discussion

4.1. General discussion

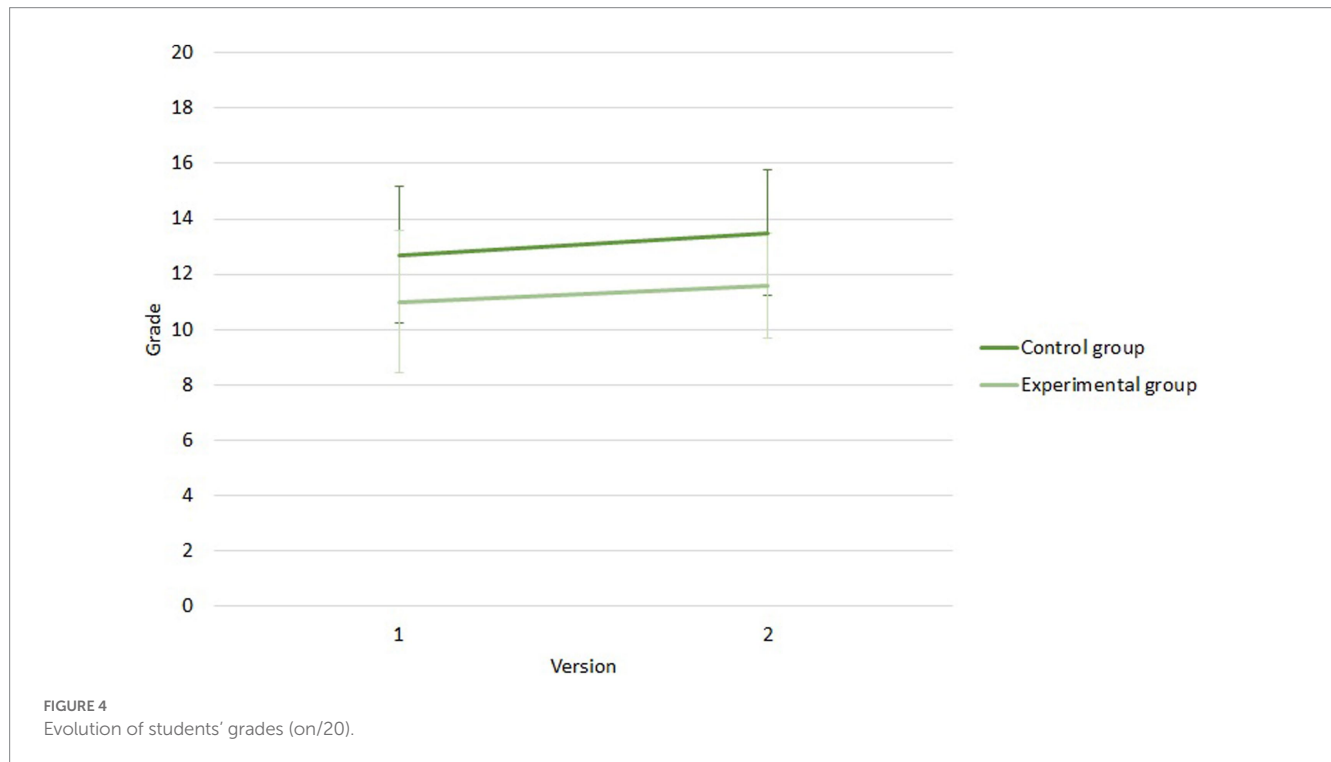
As a reminder, this study aims to investigate if a training on psychological safety and trust can reduce students' concerns

regarding peer assessment. Using an intervention design, a 1-h training session was provided to a group of students ($N = 42$), while a previous cohort served as a control group ($N = 47$). We examined whether the training session affected (1) students' perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor, (2) the type of feedback they provided to their peers, and (3) their performance improvement following the peer assessment.

Our first hypothesis was that students' perceptions of psychological safety, of trust in the self as an assessor, and of trust in the other as an assessor would increase more in the experimental than in the control group. The results did not support this hypothesis; in both groups, levels of students' perceptions did not evolve significantly from pre-test to post-test. Our training did not have a visible impact on students' perceptions. This result is not coherent with another

TABLE 5 Means and standard deviations of students' perceptions.

	Control group		Experimental group	
	Pre-test	Post-test	Pre-test	Post-test
Psychological safety	5.24 (0.92)	5.44 (1.00)	5 (1.22)	5.42 (1.14)
Trust in the self as an assessor	4.96 (0.68)	5.17 (0.78)	5.40 (0.85)	5.35 (0.93)
Trust in the other as an assessor	5 (0.62)	4.99 (0.85)	4.07 (1.04)	4.29 (1.09)

TABLE 6 Means, standard deviations, and Mann–Whitney *U* tests' results for difference scores.

	<i>M</i> (SD) in the control group	<i>M</i> (SD) in the experimental group	Mann–Whitney <i>U</i>	Sig.	Effect size (<i>r</i>)
Difference score for the global grade	0.80 (2.07)	0.57 (1.48)	915	0.548	0.07
Difference score for “diagram”	0.42 (1.00)	0.24 (0.98)	873	0.324	0.12
Difference score for “relevance of the situation”	0.31 (0.66)	0.02 (0.35)	708	0.004	0.28

study that also aimed at making students feel at ease before peer assessment but targeting perceptions of peer pressure (Li, 2017).

This absence of effect may be explained by students' positive perceptions at the pre-test; mean scores were around five on a 7-point scale. Two factors could explain these positive perceptions. First, the majority of students in the sample (89%) had prior peer assessment experiences and such prior experiences have been found related to less negative attitudes toward peer assessment (Wen and Tsai, 2006). Second, some studies show that students' perceptions of peer assessment may vary according to students' majors (e.g., Zou et al., 2018). In this study, participants studied physical education, a major in which students know each other well. They are used to practicing sports together, both during and outside courses, and they excel at different sports, which allows them to help each other out a lot.

A consequence of the initial positive perceptions is that they could have been more difficult to affect. Indeed, most learnings follow a diminishing-returns curve: it begins slowly, increases exponentially, and then slows importantly when approaching mastery (Ritter and Schooler, 2001). This learning curve has been found for the acquisition of numerous intellectual and perceptual-motor skills (Rosenbaum et al., 2001). It implies that the most important improvement happens near the start, which suggests that it is at this time that a training will have its biggest impact. Even though in the present study we targeted attitudes and not skills, this reasoning may suggest that students' perceptions may have been too high at the pre-test for the training to have a visible impact.

There was one exception regarding the high scores at the pre-test: trust in the other in the experimental group showed

moderate scores at the pre-test. These were significantly lower compared to the levels of trust in the other as an assessor for the students in the control group. A possible explanation for these differences between groups could be the COVID-19 pandemic. The data collection for the control group ended just before the first lockdown in Belgium, while the data collection for the experimental group took place after on-and-off lockdowns for almost 2 years. The use of online teaching reduced contact with peers and had a deep impact on the way students could interact with each other (Alger and Eyckmans, 2022), which likely reduced the proximity students in physical education usually experience.

Our second hypothesis was that students in the experimental group would provide more elaborated feedback (including verification and an elaboration element such as correction, justification, question, or suggestion) than students in the control group. This hypothesis was not supported either; students in the experimental group did not provide any type of feedback in a higher proportion than students in the control group did. The absence of effects supporting our hypothesis is not surprising given the absence of visible effects of the training on students' perceptions. Our initial hypothesis was that the training would increase students' perceptions, which would incite them to provide more elaborated feedback. However, given that the training did not seem to affect students' perceptions, this assumption did not hold anymore.

In addition, significant effects in the opposite direction of our hypotheses were observed: students in the control group provided more correction and justification feedback than students in the experimental group. This effect may be explained by the initial performance differences between the two groups. The higher grades received by students from the control group suggest they had a better understanding of the course, which could explain why they seemed able to give more elaborated feedback than students in the experimental group. It is known that higher-performing students tend to provide more useful feedback to their peers (Wu and Schunn, 2023). This explanation is supported by the positive correlations between students' grades for their first draft and the proportion of question feedback provided.

Our third hypothesis was that students' performance would improve more in the experimental than in the control group. We started by looking at whether there was a performance improvement. As expected, in both groups, students' grades were higher for their final version than for their first draft. More precisely, students improved their grades for the criteria "diagram" and "relevance of the situation," but not for the criteria "instructions" and "success criteria." The particularity of these last two criteria is that students must master both higher-order aspects (e.g., find the most relevant instruction for the exercise) and lower-order aspects (e.g., write the instruction with adequate vocabulary and without spelling mistakes). If students only acted upon peer feedback on lower-order aspects – which prior research has shown that students tend to do (e.g., Aben et al., 2022; Van Meenen et al., 2023)—and left important problems with higher-order aspects, the improvement would not result in a higher grade from the professor.

The comparison of the performance improvement in the control and the experimental group did not support our hypothesis that students from the experimental group would improve more than those in the control group. The performance improvement was similar in both groups for the global grade. Here too, one could argue that it

is due to the absence of a visible impact of the training on students' perceptions. Our initial hypothesis was that the training would increase the perceptions of students from the experimental group, which would allow them to gain more from the peer assessment than students from the control group. Given that the training did not seem to affect students' perceptions, the assumption that students from the experimental group would improve more did not hold anymore.

On the contrary, considering that students in the control group trusted their peers more, they are expected to benefit more from their feedback (van Gennip et al., 2010). We would therefore expect a larger improvement in the control than in the experimental group. This was the case, but only for the criterion "relevance of the situation" (a criterion for which an improvement is only possible if important changes were made). For this criterion, students who trust their peers may have considered their feedback more carefully and, therefore, improved their performance. An element that supports this is the moderate positive correlation between students' perceptions of trust in the peer as an assessor and the performance gain for the criteria "relevance of the situation."

4.2. Pathways for future research

All results taken together, we could not find evidence supporting the training's efficacy. There may be several reasons for this finding, which should be investigated in future research.

First, some limitations of the study may have prevented us to find an effect of the training. As the study was conducted in a natural setting, it was impractical and unethical to use random sampling (Cohen et al., 2007). Therefore, we used a quasi-experiment with pre-and post-test design with two cohorts of students. Although this design is relatively strong, having the two groups participating in the study in different years hindered the comparability between the control and the experimental groups, especially in this case with a pandemic occurring in between. Students in the experimental group participated in this study after experiencing several lockdowns during which the majority of their classes were held online. Online teaching due to the COVID-19 pandemic deeply impacted interactions between students (Alger and Eyckmans, 2022) and students' learning (Di Pietro, 2023), which could explain the differences between the two groups at pre-test. To ensure group comparability while remaining ethical, future studies could be conducted in a course with a large cohort of students and multiple peer assessment opportunities. In this context, a waiting list control group would be possible (Elliott and Brown, 2002); the training could be introduced at different times for different groups of students so that it would be possible to compare them while still allowing every student to benefit from the training before the outcomes of peer assessment could impact their grades.

Another limitation is linked to the population of the study. As explained before, the perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor of students participating in the study were already quite positive at the pre-test, which could have made it more difficult to affect them (Ritter and Schooler, 2001). Future studies could investigate the impact of the training on students who have negative perceptions of psychological safety and trust at baseline. These studies could be conducted with students for whom more concerns related to peer assessment could be expected, such as students who never experienced peer assessment

before (Wen and Tsai, 2006) or students who are studying humanities (Praver et al., 2011).

Second, it is possible that the training had both positive and negative impacts on students' perceptions, which canceled each other out. During the training, students had to opportunity to think about and discuss their concerns linked to peer assessment. If students feel safe and had trust in themselves and their peers before the training, which seems to be the case, the training may have challenged these positive perceptions at first, before positively affecting them. For trust in the self as an assessor more particularly, the training may have both allowed students to acquire skills in providing feedback and to realize that they may have overestimated their skills before. Indeed, according to the Dunning-Kruger effect, unskilled people overestimated themselves due to their lack of skills (Kruger and Dunning, 1999).

Third, the training could be more effective with a smaller students per instructor-ratio. In the study, the relatively large number of students made it difficult for the instructor to visit each group, while they were role-playing, which is important to keep students on task, answer their questions, and provide suggestions (Bolinger and Stanton, 2020). Moreover, to ensure student participation, group discussions were also conducted in small groups and, therefore, could not be facilitated by an instructor, contrary to what is recommended (Blanchard and Thacker, 2012). Future studies could investigate the efficacy of an improved version of the training, with more time for role-plays and discussions, diverse training methods, and more instructors to guide the students.

In addition, choices made to keep training time-and cost-efficient may have reduced its efficacy. Our objective was to create a training that could easily be implemented in any higher education course. Therefore, the training had been designated so that it was short enough to fit into the busy course schedules and so that a single instructor could train a group of about 50 students. At the same time, to overcome a limitation of the training on psychological safety developed by Dusenberry and Robinson (2020), our training almost exclusively contained active-learning methods, namely role-play and group discussion. Active-learning methods and interactions between learners are indispensable for learning, but they require time (Martin et al., 2014), which was limited in our 1-h session. Students only participated in two role-plays (one as active participants, and one as observers), while it is recommended to have three sets of role-plays (Blanchard and Thacker, 2012). Moreover, with a 1-h session, it was not possible to combine more than two training methods, although this combination enhances training effectiveness (Martin et al., 2014). Although short training is effective to impact peer pressure (Li, 2017), longer training may be required to impact psychological safety and trust, because as Hunt et al. (2021) argued it is difficult to foster psychological safety.

It is also possible that given this difficulty to foster psychological safety (Hunt et al., 2021), a training is not an effective way to affect it. The only other study we know of that tried to affect psychological safety through training did not find a significant effect either (Dusenberry and Robinson, 2020). Even when designing the training with this study's limitations in mind to overcome, we did not find a significant impact of the training on students' perceptions of psychological safety. This would suggest that, contrary to peer pressure (Li, 2017), psychological safety is a factor that cannot be effectively enhanced through a training, and that other kinds of interventions are needed.

To the best of our knowledge, the present study is the first that aims to incite trust in the self and the peer as an assessor through training. While the results of the present study indicate an absence of effect, more studies are warranted. Concluding that trust cannot be positively affected by a training appears premature.

To have an intervention that is both feasible and effective, a possibility could be to use guidance during peer assessment, eventually in combination with a 1 or 2-h training session prior to the peer assessment exercise. Prior research found that guidance can enhance the peer assessment process (e.g., Bloxham and Campbell, 2010; Harland et al., 2017; Misiejuk and Wasson, 2021). Three types of guidance could also enhance students' perceptions of psychological safety and trust. The first is the use of backward evaluation, the feedback that an assessee provides to his/her assessor about the quality of the received feedback (Misiejuk and Wasson, 2021). The second is the use of cover sheets, sheets that allow an assessee to ask specific questions and explain what type of feedback they would like to receive (Bloxham and Campbell, 2010). The third is the use of a rebuttal, a text that assessee must write to explain why they found the received feedback relevant or not and justify how they acted upon it (Harland et al., 2017). These sorts of guidance could help students realize they are not passive recipients of feedback, but that they can critically appraise the feedback they received and that they can improve themselves in providing feedback, and therefore, could enhance their perceptions. Future studies could investigate if, combined with a training session, these kinds of guidance have positive impacts on students' perceptions of psychological safety and trust.

Besides testing the effectiveness of the training we developed, a second aim of this study was to investigate the impact of an increase in students' perceptions of psychological safety and trust in the self and the other as assessor on peer assessment outcomes. Given that we could not find any impact of the training on students' perceptions, we have no evidence that an increase in psychological safety or trust would result in higher quality feedback or bigger performance improvement. However, there was a correlation between students' perceptions of trust in the peer as assessor and their performance gain for one of the criteria ("relevance of the situation"). Previous research found that trust in the other as an assessor is a predictor of students' learning following a peer assessment activity when learning is measured by asking students if they feel they have learned (van Gennip et al., 2010). Our result corroborates this positive relationship between trust in the other and learning, with the learning being measured by students' grade improvement in our study.

The relationship between perceptions of trust in the other and performance improvement did not seem to be explained by the quality of feedback given by students (there was no significant correlation in our study). However, the quality of provided feedback is not the only variable that could mediate the relationship between students' perceptions and their performance improvement, feedback uptake could also play an important role. Some studies show that peer feedback uptake tends to be low and that students make few revisions following peer assessment (e.g., Winstone et al., 2017; Berndt et al., 2018; Aben et al., 2022; Bouwer and Dirkx, 2023; Van Meenen et al., 2023). Possibly, this low feedback uptake is linked to low levels of psychological safety and trust, but empirical evidence is lacking on this. Future research on the relationship between students' feedback uptake on the one hand and their perceptions of psychological safety and trust on the other hand is warranted.

4.3. Conclusion

This study aimed to investigate the efficacy of a training session created to enhance students' perceptions of psychological safety and trust before a peer assessment activity. Findings suggested that students' perceptions are related to students' performance improvement, which confirms that it is important to develop interventions targeting these perceptions. Contrary to our hypotheses, we could not find evidence of the training's effectiveness on students' perceptions, students' performances, or the type of feedback provided by students. This absence of visible impact could be explained by limitations of the study, or by limitations of the training in itself. Future studies could investigate if an improved longer version of the training would be more effective, or if the training should be combined with other guidance during peer assessment.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Commission d'éthique IPSY, Université catholique de Louvain. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MS, DJ, and LC conceptualized the research project, defined the research question, and developed the research design. MS and DJ

implemented the intervention and collected the data, under the supervision of LC. MS performed the statistical analysis with the help of LC. MS wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to thank Maryam Alqassab for sharing her detailed coding scheme, Aurélie Bertrand helping us with the analyses, and Quentin Brouhier for coding feedback.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2023.1198011/full#supplementary-material>

References

- Aben, J., Timmermans, A. C., Dingyloudi, F., Mascareño Lara, M., and Strijbos, J. (2022). What influences students' peer-feedback uptake? Relations between error tolerance, feedback tolerance, writing self-efficacy, perceived language skills and peer-feedback processing. *Learn. Individ. Differ.* 97:102175. doi: 10.1016/j.lindif.2022.102175
- Alger, M., and Eyckmans, J. (2022). "I took physical lessons for granted": a case study exploring students' interpersonal interactions in online synchronous lessons during the outbreak of COVID-19. *System (Linköping)* 105:102716. doi: 10.1016/j.system.2021.102716
- Alqassab, M., Strijbos, J., and Ufer, S. (2018). The impact of peer solution quality on peer-feedback provision on geometry proofs: evidence from eye-movement analysis. *Learn. Instr.* 58, 182–192. doi: 10.1016/j.learninstruc.2018.07.003
- Berndt, M., Strijbos, J.-W., and Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *Eur. J. Psychol. Educ.* 33, 31–49. doi: 10.1007/s10212-017-0343-z
- Black, N., Mullan, B., and Sharpe, L. (2016). Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. *Health Psychol. Rev.* 10, 341–357. doi: 10.1080/17437199.2016.1168268
- Blanchard, P. N., and Thacker, J. W. (2012). *Effective training: Systems, strategies, and practices*. 5th Edn. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Bloxham, S., and Campbell, L. (2010). Generating dialogue in assessment feedback: exploring the use of interactive cover sheets. *Assess. Eval. High. Educ.* 35, 291–300. doi: 10.1080/02602931003650045
- Bolinger, A. R., and Stanton, J. V. (2020). *Role-play simulations*. Cheltenham: Elgar.
- Bouwer, R., and Dirckx, K. (2023). The eye-mind of processing written feedback: unraveling how students read and use feedback for revision. *Learn. Instr.* 85:101745. doi: 10.1016/j.learninstruc.2023.101745
- Carless, D., Salter, D., Yang, M., and Lam, J. (2011). Developing sustainable feedback practices. *Stud. High. Educ.* 36, 395–407. doi: 10.1080/03075071003642449
- Cheng, K., and Tsai, C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australas. J. Educ. Technol.* 28, 599–618. doi: 10.14742/ajet.830
- Ching, Y., and Hsu, Y. (2016). Learners' interpersonal beliefs and generated feedback in an online role-playing peer-feedback activity: an exploratory study. *Int. Rev. Res. Open Dist. Learn.* 17, 105–122. doi: 10.19173/irrodl.v17i2.2221
- Cohen, L., Manion, L., and Morrison, K. (2007). *Research methods in education*. 6th Edn. London: Routledge/Taylor & Francis Group.
- Di Pietro, G. (2023). The impact of covid-19 on student achievement: evidence from a recent meta-analysis. *Educ. Res. Rev.* 39:100530. doi: 10.1016/j.edurev.2023.100530
- Double, K., McGrane, J., and Hopfenbeck, T. (2020). The impact of peer assessment on academic performance: a meta-analysis of control group studies. *Educ. Psychol. Rev.* 32, 481–509. doi: 10.1007/s10648-019-09510-3
- Dusenberry, L., and Robinson, J. (2020). Building psychological safety through training interventions: manage the team, not just the project. *IEEE Trans. Prof. Commun.* 63, 207–226. doi: 10.1109/TPC.2020.3014483
- Edmondson, A. C., and Lei, Z. (2014). Psychological safety: the history, renaissance, and future of an interpersonal construct. *Annu. Rev. Organ. Psych. Organ. Behav.* 1, 23–43. doi: 10.1146/annurev-orgpsych-031413-091305

- Elliott, S. A., and Brown, J. S. L. (2002). What are we doing to waiting list controls? *Behav. Res. Ther.* 40, 1047–1052. doi: 10.1016/S0005-7967(01)00082-1
- Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among multiple raters* 4). Gaithersburg, MD: Advanced Analytics, LLC.
- Hanrahan, S. J., and Isaacs, G. (2001). Assessing self-and peer-assessment: the students' views. *High. Educ. Res. Dev.* 20, 53–70. doi: 10.1080/07294360123776
- Harland, T., Wald, N., and Randhawa, H. (2017). Student peer review: enhancing formative feedback with a rebuttal. *Assess. Eval. High. Educ.* 42, 801–811. doi: 10.1080/02602938.2016.1194368
- Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487
- Hunt, D., Bailey, J., Lennox, B., Crofts, M., and Vincent, C. (2021). Enhancing psychological safety in mental health services. *Int. J. Ment. Heal. Syst.* 15:33. doi: 10.1186/s13033-021-00439-1
- Kollar, I., and Fischer, F. (2010). Peer assessment as collaborative learning: a cognitive perspective. *Learn. Instr.* 20, 344–348. doi: 10.1016/j.learninstruc.2009.08.005
- Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134. doi: 10.1037/0022-3514.77.6.1121
- Li, L. (2017). The role of anonymity in peer assessment. *Assess. Eval. High. Educ.* 42, 645–656. doi: 10.1080/02602938.2016.1174766
- Li, L., Liu, X., and Steckelberg, A. L. (2010). Assessor or assessee: how student learning improves by giving and receiving peer feedback. *Br. J. Educ. Technol.* 41, 525–536. doi: 10.1111/j.1467-8535.2009.00968.x
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., and Tywoniu, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assess. Eval. High. Educ.* 45, 193–211. doi: 10.1080/02602938.2019.1620679
- Martin, B. O., Kolomito, K., and Lam, T. C. M. (2014). Training methods: a review and analysis. *Hum. Resour. Dev. Rev.* 13, 11–35. doi: 10.1177/1534484313497947
- Misiejuk, K., and Wasson, B. (2021). Backward evaluation in peer assessment: a scoping review. *Comput. Educ.* 175:104319. doi: 10.1016/j.compedu.2021.104319
- Mulder, R. A., Pearce, J. M., and Baik, C. (2014). Peer review in higher education: student perceptions before and after participation. *Act. Learn. High. Educ.* 15, 157–171. doi: 10.1177/1469787414527391
- Panadero, E., and Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assess. Eval. High. Educ.* 44, 1253–1278. doi: 10.1080/02602938.2019.1600186
- Panadero, E., Alqassab, M., Fernández Ruiz, J., and Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assess. Evaluat. Higher Educ.* ahead-of-print, 1–23. doi: 10.1080/02602938.2023.2164884
- Planas Lladó, A., Soley, L. F., Fraguell Sansbelló, R. M., Pujolras, G. A., Planella, J. P., Roura-Pascual, N., et al. (2014). Student perceptions of peer assessment: an interdisciplinary study. *Assess. Eval. High. Educ.* 39, 592–610. doi: 10.1080/02602938.2013.860077
- Praver, M., Rouault, G., and Eidswick, J. (2011). Attitudes and affect toward peer evaluation in EFL reading circles. *Read. Matrix* 11, 89–101.
- Ritter, F. E., and Schooler, L. J. (2001). "The learning curve," in *International encyclopedia of social and behavioral sciences*. eds. N. J. Smelser and P. B. Baltes (Amsterdam: Pergamon).
- Rosenbaum, D. A., Carlson, R. A., and Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annu. Rev. Psychol.* 52, 453–470. doi: 10.1146/annurev.psych.52.1.453
- Rotsaert, T., Panadero, E., Estrada, E., and Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Stud. Educ. Eval.* 53, 29–40. doi: 10.1016/j.stueduc.2017.02.003
- Rotsaert, T., Panadero, E., and Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *Eur. J. Psychol. Educ.* 33, 75–99. doi: 10.1007/s10212-017-0339-8
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795
- Soares, A. E., and Lopes, M. P. (2020). Are your students safe to learn? The role of lecturer's authentic leadership in the creation of psychologically safe environments and their impact on academic performance. *Act. Learn. High. Educ.* 21, 65–78. doi: 10.1177/1469787417742023
- Strijbos, J., Martens, R. L., Prins, F. J., and Jochems, W. M. G. (2006). Content analysis: what are they talking about? *Comput. Educ.* 46, 29–48. doi: 10.1016/j.compedu.2005.04.002
- Su, W. (2022). Masked ball for all: how anonymity affects students' perceived comfort levels in peer feedback. *Assess. Evaluat. Higher Educ.* ahead-of-print, 1–11. doi: 10.1080/02602938.2022.2089348
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., and Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *High. Educ.* 76, 467–481. doi: 10.1007/s10734-017-0220-3
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Rev. Educ. Res.* 68, 249–276. doi: 10.3102/00346543068003249
- van Gennip, N., Segers, M., and Tillema, H. (2010). Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learn. Instr.* 20, 280–290. doi: 10.1016/j.learninstruc.2009.08.010
- Van Meenen, F., Masson, N., Catrysse, L., and Coertjens, L. (2023). Taking a closer look at how higher education students process and use (discrepant) peer feedback. *Learn. Instr.* 84:101711. doi: 10.1016/j.learninstruc.2022.101711
- Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., and Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Comput. Educ.* 81, 123–132. doi: 10.1016/j.compedu.2014.10.001
- von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociol. Methods Res.* 49, 699–718. doi: 10.1177/0049124117747303
- Wen, M. L., and Tsai, C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *High. Educ.* 51, 27–44. doi: 10.1007/s10734-004-6375-8
- Wilson, M. J., Diao, M. M., and Huang, L. (2015). 'I'm not here to learn how to mark someone else's stuff': an investigation of an online peer-to-peer review workshop tool. *Assess. Eval. High. Educ.* 40, 15–32. doi: 10.1080/02602938.2014.881980
- Winstone, N. E., Nash, R. A., Parker, M., and Rowntree, J. (2017). Supporting Learners' Agentic engagement with feedback: a systematic review and a taxonomy of Recipience processes. *Educ. Psychol.* 52, 17–37. doi: 10.1080/00461520.2016.1207538
- Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Front. Psychol.* 10:3087. doi: 10.3389/fpsyg.2019.03087
- Wu, Y., and Schunn, C. D. (2023). Assessor writing performance on peer feedback: exploring the relation between assessor writing performance, problem identification accuracy, and helpfulness of peer feedback. *J. Educ. Psychol.* 115, 118–142. doi: 10.1037/edu0000768
- Yu, F., and Liu, Y. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *Br. J. Educ. Technol.* 40, 1109–1123. doi: 10.1111/j.1467-8535.2008.00905.x
- Yu, F., and Wu, C. (2011). Different identity revelation modes in an online peer-assessment learning environment: effects on perceptions toward assessors, classroom climate and learning activities. *Comput. Educ.* 57, 2167–2177. doi: 10.1016/j.compedu.2011.05.012
- Zou, Y., Schunn, C. D., Wang, Y., and Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assess. Eval. High. Educ.* 43, 800–811. doi: 10.1080/02602938.2017.1409872