Designing peer feedback in higher education

Building trust and psychological safety among students

Designing peer feedback in higher education

Building trust and psychological safety among students

Morgane Senden

Promotrice

Liesje COERTJENS (UCLouvain)

Thèse présentée en vue de l'obtention du grade de Docteur en sciences psychologiques et de l'éducation

Collection de thèses de l'UCLouvain, 2024

Comité d'accompagnement et jury

Catherine VAN NIEUWENHOVEN (UCLouvain, Présidente de jury)
Liesje COERTJENS (UCLouvain, Secrétaire du jury)
Vincent YZERBYT (UCLouvain)
Noémie BAUDOIN (UCLouvain)
Tijs ROTSAERT (Universiteit Gent, Belgium)
Maryam ALQASSAB (Open University, the Netherlands)

Couverture: illustration originale de Tess-h.be (2024)

Table of contents

Résumé

L'utilisation de l'évaluation par les pairs dans l'enseignement supérieur présente de multiples avantages : elle contribue au développement de l'apprentissage des étudiant·es et de leurs compétences en matière de feedback et elle les aide à développer les compétences nécessaires pour apprendre tout au long de leur vie. Malgré ces avantages, les étudiant·es sont souvent réticent·es à participer à des évaluations par les pairs et expriment de nombreuses inquiétudes. Ces inquiétudes découlent du fait que l'évaluation par les pairs est une activité sociale impliquant des processus interpersonnels. Cette thèse cible trois de ces processus qui jouent un rôle important dans l'évaluation par les pairs : la confiance en ses capacité à donner des feedbacks, la confiance en les capacités des pairs à donner des feedbacks et la sécurité psychologique. Bien qu'il soit reconnu que ces trois facteurs sont importants à prendre en compte lors de la conception des activités d'évaluation par les pairs, il existe peu d'études sur la manière de procéder. Par conséquent, à travers une revue de la littérature et trois études empiriques, cette thèse vise à mieux comprendre comment l'évaluation par les pairs peut être organisée pour optimiser les processus interpersonnels. Plus spécifiquement, elle vise à savoir si une formation, combinée ou non à une évaluation de la qualité des feedbacks reçus, peut augmenter les perceptions de confiance et de sécurité psychologique des étudiant·es.

Abstract

The use of peer feedback in higher education has multiple advantages: it fosters student learning, contributes to the development of feedback skills, and helps students develop the skills needed for lifelong learning. Notwithstanding these benefits, students are often reluctant to engage in peer feedback and express many concerns. These concerns emerge from the fact that peer feedback is a social activity involving interpersonal processes. This dissertation focuses on three interpersonal processes that play an important role in peer feedback: trust in the self as an assessor, trust in the other as an assessor, and psychological safety. Even though it is known that these three factors are important to consider when designing peer feedback activities, there is a scarcity of studies investigating how this could be done. Therefore, through a literature review and three empirical studies, this dissertation aims to better understand how peer feedback activities can be designed to optimize interpersonal processes. More specifically, it aims to discover if students' perceptions of trust and psychological safety can be optimized using an intervention consisting of training whether or not combined with a backward evaluation.

Remerciements

Tout d'abord, un grand merci à **Liesje** de m'avoir accompagnée et soutenue dans ce parcours doctoral. Peu importe les circonstances, j'ai toujours senti que tu étais disponible pour moi. Merci pour ton aide, ton expertise, tes conseils, tes encouragements. En plus d'un soutien constant, tu m'as aussi donné l'autonomie dont j'avais besoin pour m'approprier mon projet de thèse et me sentir capable de le réaliser. Merci pour ta confiance. Je n'aurai pas pu rêver de meilleure promotrice !

Un grand merci également aux membres de mon comité d'accompagnement. **Tijs**, thank you for your expertise and all your help. I have greatly appreciated your advice and thought-provoking comments, whether during the annual supervisory panel meetings or at the EARLI conferences. **Catherine**, un grand merci pour tes conseils et tes encouragements. Merci pour le temps que tu m'as accordé malgré un emploi du temps que je sais très chargé, ton aide m'a été particulièrement utile pour concevoir et analyser la partie qualitative de ma dernière étude. **Vincent**, merci pour ton regard extérieur sur ma thèse. Avoir dans mon comité quelqu'un qui vient d'un autre domaine de la psychologie et utilise d'autres méthodes était parfois un peu challengeant, mais surtout très enrichissant. Merci pour ton aide et tes conseils statistiques.

Merci beaucoup à Noémie et Maryam d'avoir accepté de faire partie de mon jury de thèse. **Maryam**, I first discovered you through the articles you wrote, which have all been sources of inspiration for my dissertation. I was grateful to have your expertise on my thesis jury. **Noémie**, merci pour l'intérêt que tu as porté à ma thèse, pour tes commentaires, et pour tes réflexions issues de ton regard de chercheuse, mais également d'enseignante du supérieur. Je me réjouis d'ailleurs de collaborer avec toi pour le cours QAPA.

Je n'aurai pas pu réaliser ma thèse sans l'aide des principales personnes concernées : un tout grand merci aux étudiant·es qui ont accepté de participer à mes études. Et un merci tout particulier aux étudiant·es qui ont donné de leur temps, parfois en pleine période de blocus, pour participer aux entretiens.

Outre les étudiant·es, je n'aurais pas pu mener cette thèse sans la collaboration de professeur·es qui ont eu la gentillesse de m'acceuillir dans leurs cours. **Dominique**, **Frédéric**, **Mariane** et **Bernadette**, merci d'avoir adapté vos cours pour y intégrer mes interventions et merci d'avoir fait en sorte que je puisses récolter des données dans les meilleures conditions possibles. Merci également à **Amandine**, **Laure** et **Margot** d'avoir pris à cœur de donner fidèlement les formations aux feedbacks lors des TP de psycho de l'apprentissage.

Merci aux stagiaires et mémorant·es qui ont contribué aux collectes et encodages de données, à la restranscription d'entretien et au codage de feedback : **Dorine**, **Johan**, **Dorian**, **Camille** et **Guillaume**, merci pour votre aide. Un merci particulier à **Annelise** pour ton aide précieuse pour l'analyse d'entretiens.

Merci aux bibliothécaires **Sophie** et **Céline** pour m'avoir aidé dans les premières étapes de ma revue de la littérature, ainsi qu'à **Aurélie**, **Nathalie** et **Alain** du SMCS pour leur aide au niveau statistique.

Aux chercheureuses avec qui j'ai pu échanger lors de conférences internationales ou des cours ICO, merci pour vos conseils, remarques et questions qui ont alimenté mes réflexions.

Avant de rejoindre l'équipe 2 du GIRSEF j'avais des craintes quant à l'entrée dans le monde de la recherche, souvent décrit comme très compétitif. J'ai très vite découvert que dans notre équipe c'était l'entraide qui primait. Je pense avoir particulièrement eu de la chance d'être entourée de collègues si géniaux·alles pendant ces années et de toujours avoir reçu l'aide dont j'avais besoin que ce soit pour avoir des retours sur un article en cours d'écriture (vive les solidareviews !), pour échanger sur des analyses, pour recevoir des conseils de lecture, et plein d'autres choses encore. Un immense merci à mes collègues pour tout cela ! Et puis quel bonheur d'être dans une équipe qui, malgré une quantité de travail importante, n'oublie jamais d'aussi de s'amuser. Je pense que peu de doctorant·es ont eu l'occasion pendant leur thèse de chercher des œufs cachés dans les couloirs, de fêter les pendaisons de crémaillères des bureaux ou encore de participer à des Xmas challenges… Merci pour cela ! Et un merci tout particulier à **l'esprit du couloir** (j'espère que tu ne finiras jamais ta thèse et continueras à hanter les couloirs pendant longtemps). Je garderai également en mémoire tous les supers team-building et les soirées, que ce soit en boite de nuit à Lausanne ou au cercle de psycho.

En environ six ans et demi, j'ai eu l'occasion de côtoyer un nombre important de collègues, et j'ai donc de nombreuses personnes à remercier. C'est ce que je vais faire ici dans un ordre plus ou moins chronologique. Tout d'abord, merci à **Benoit** de m'avoir proposé de travailler sur le harcèlement scolaire comme assistante de recherche. Merci pour ton accompagnement bienveillant pendant ces quelques mois qui a contribué à me faire aimer le monde de la recherche et me donner envie de réaliser une thèse. Merci à tous les collègues qui étaient déjà dans l'équipe quand je suis arrivée. Merci à **Mikaël** pour nos séances d'escalade et pour notre duo (pas vraiment) imbattable au badminton. Merci aussi pour tes attentions, que ce soit via des truffes pour fêter mon dernier TP de stat ou des bonbons à l'occasion de mon anniversaire. Merci à **Chloé** pour ton accueil chaleureux à mon arrivée, ainsi que pour tous les moments passés ensemble, notamment lors de notre année de cohabition dans le même bureau. Merci à **The Huy** pour les repas partagés dès midi pile, quand nous étions les seuls à avoir déjà faim, et merci de m'avoir appris à cuisiner de délicieux dumplings. Merci à **Simon** pour avoir veillé à ce que nous ne soyons pas débordé·es grâce à ton travail en tant que représentant assistant et pour tous tes conseils de fin de thèse. Merci à **Florence D** pour ta joie de vivre contagieuse. Merci à **Michaël** pour ton expertise et ta disponibilité. Merci à **Elise** pour les moments partagés dans l'effective team. Merci à **Agnès** pour tes conseils et encouragements.

J'ai eu la chance de ne pas être la seule à avoir débuté ma thèse en septembre 2018. **Christine**, merci d'avoir été là à chaque étape, d'avoir partagé les moments de doute du début

de thèse, les dernières semaines d'écriture intense, et très bientôt la célébration de nos défenses publiques. Ça a été un réel plaisir de t'avoir dans le bureau juste à côté pendant ces six années, d'entendre ton rire à travers le mur, et de te voir régulièrement débarquer dans mon bureau. Merci pour tes attentions, tes encouragements, toutes les fois où tu m'as ramenée à Bruxelles, nos discussion et de manière plus générale, merci pour tous les moments passés ensemble. **Quentin**, merci pour ta créativité (merci d'ailleurs pour m'avoir aidé à trouver ce titre de thèse), merci pour les midis d'équipe et les partages de recettes, merci pour ton investissement important dans la vie de l'équipe et merci d'être toujours là quand on a besoin de toi. **Thibault**, merci pour ta disponibilité, ton expertise en statistique et ta passion communicative pour la recherche. **Amandine**, merci tous les partages autour des cours et le soutien lors des corrections.

Merci à tous les collègues que j'ai eu le plaisir de voir arriver dans l'équipe en cours de route. **Margot F**, merci d'être une personne si attentive et dévouée avec qui c'est tellement agréable de travailler. **Florence V** et **Charlie** (impossible de ne pas vous mettre ensemble), merci de mettre de l'animation dans tout le couloir, merci de me faire rire, merci pour tous nos échanges. Vous êtes de supers collègues et si un jour vous ouvrez votre refuge pour hérissons, je viendrai travailler avec vous sans hésitation ! **Sarah** et **Justine**, un grand merci pour tout ce que vous faites pour l'équipe. **Sarah**, merci pour ta bienveillance et pour ton humour. **Justine**, merci pour toutes tes supers initiatives, j'adore trouver des petits canards pendant mes pauses ! **Laure**, merci d'être une personne à la fois drôle (merci pour tes farces, jeux de mots et dad jokes) et sur laquelle on peut toujours compter (merci pour ton investissement dans ton travail d'assistante). **Alice**, merci pour ton côté maman attentionée de l'équipe (surtout en semaine d'écriture !). Merci aussi à toutes les quatre de m'avoir fait découvrir que j'étais printemps et merci pour tous les moments passés ensemble. **Eve**, merci pour ton implication dans la vie de l'équipe, merci pour tes engagements et merci d'apporter un bout de culture militante au bureau. **Marie-France**, merci pour ta bonne humeur et ta générosité. **François**, merci pour les parties de jeux de société. **Maurice** merci pour ta force tranquille. **Caroline R**, merci pour les TP donnés ensemble et le soutien en période de corrections.

Et puis merci aux dernières arrivées que j'ai eu l'occasion de côtoyer pendant moins longtemps, mais grâce à qui je suis sûre que l'équipe gardera son ADN dans les années à venir. **Zoé**, merci pour ton investissement dans la vie de l'équipe, que ce soit au sein de la cellule verte, de la PEI, ou en donnant des coups de main à l'esprit du couloir. **Marie**, merci pour tes conseils au badminton (dommage que tu ne sois pas arrivée des années plus tôt, j'aurais pu devenir un pro et Mikaël et nous serions un duo vraiment imbattable). **Julie**, merci pour ton implication dans ton travail d'assistante et merci de toujours proposer ton aide. **Astrid**, merci pour ta générosité et tes délicieux chips d'orties. **Noémie D**, merci pour ta gentillesse.

Un tout grand merci également aux membres des équipes administrative : **Isabelle**, **Anaïs**, **Régeane**, **Sigried**, **Dominique**, **Martine**, **Achraf**, **Roland**, **Sabine**, **Sandra**, **Solange**. Un merci particulier à **Marianne** pour ton aide pour la mise en page de ma thèse.

Merci également à toutes les autres personnes que j'ai toujours eu plaisir à croiser dans les couloirs du 1er étage : **Olivier**, **Thomas**, **Claire**, **Margot D**, **Caroline V**, **Joséphine**, **Rindra**, **Denis**, **Antoine**, **Sébastien**, **Jean-Pierre**, **Coryse**, **Marc**, **Virginie**, **Isabel**, **Stéphane**, **Frédéric**, **Vanessa**, **Margharita**, **Véronique**, **Nathalie**, **Emmanuelle**, **Pavana**, **Anne**, **Christelle**, **Pascalia**.

Hors du bureau et du monde de la recherche, j'ai été énormément soutenue et entourée au cours de la réalisation de cette thèse. Merci à mes ami·es, et particulièrement à **Sofie** et **Laura**, pour vos encouragements. Merci à toute ma famille ainsi qu'à ma belle-famille pour votre soutien. Un grand merci à **Tess** pour la magnifique illustration de ma couverture de thèse. Merci à **mes parents** d'avoir développé ma curisoité, d'avoir constemment veillé à ce que l'école et l'apprentissage soit toujours source de plaisir, et de m'avoir toujours soutenue et encouragé dans mes études. Merci aussi **maman** pour ton soutien dans les dernières étapes de ma thèse, par exemple en m'aidant à vérifier ma bibliographie. Merci à **Aurore** et **Iris** d'être des sœurs si géniales, merci d'être toujours présentes, merci pour tout ce qu'on partage. Et finalement merci à toi **Andreas**. Depuis notre bachelier en psycho, je ne sais compter combien de mes écrits tu as relu et pour combien de présentations tu m'as écoutée m'entrainer. Ton soutien a été particulièrement important dans ces intenses derniers mois de thèse. Merci d'être là, merci d'être toi.

Chapter 1

General introduction

General introduction

It is now well-established that feedback can be a powerful tool to achieve the main goal of education: foster students' learning (e.g. P. Black & Wiliam, 1998, Hattie & Timperley, 2007). Feedback is one of the most effective educational interventions: extensive evidence supports its high impact on learning and it is low cost compared to other educational interventions (Education Endowment Foundation, 2023). Most skills can only be acquired with practice in a supportive environment incorporating feedback (Sadler, 1989). Effective feedback helps learners reduce the gap between their current performance and desired performance by informing them about the learning goal to attain, their current level of performance, and how to improve their current performance (Hattie & Timperley, 2007; Sadler, 1989).

Classically, feedback is often provided by teachers but, with the massification of higher education, teachers rarely have the opportunities to provide feedback to students and alternatives to teachers' feedback are needed (Huisman et al., 2019). An alternative to teachers' feedback that has multiple advantages is peer feedback. Peer feedback can be as effective as teacher feedback to foster students' learning (Double et al., 2020: Wisniewski et al., 2020). Its potential lies in the fact that peer feedback inspires students to take an active role in their learning, to learn from the feedback they receive, and to learn to better assess themselves, which is a skill that is considered essential to life-long learning (Liu & Carless, 2006; Carless et al., 2011)

Given the benefits of peer feedback, it can be expected that students are motivated to participate in it. This is partly the case; most students report that they like peer assessment and find it useful (Hanrahan & Isaacs, 2001; Mulder et al., 2014). However, they also express a series of concerns that are often related to the social nature of peer assessment (e.g., Carvalho, 2013; Wilson et al., 2015). If peer assessment is used without taking into consideration its social nature – and students' concerns related to it – it could negatively affect its validity (Topping, 2017).

This raises the question of how to design peer feedback activities to minimize the negative effects of interpersonal processes. In this dissertation, I will focus on three interpersonal processes that play an important role in peer feedback (e.g., K. Cheng & Tsai, 2012; van Gennip, Segers & Tillema, 2010): trust in the other as an assessor, trust in the self as an assessor, and psychological safety. In the current dissertation, I aim to investigate how peer feedback activities can be designed to ensure that higher education students feel psychologically safe and have trust in their abilities and in their peers' abilities.

The focus is put on higher education for three main reasons. First, it is more complicated for teaching staff to provide formative feedback to students in higher education than at other educational levels given the important students-to-teacher ratio, which means that alternative feedback practices are particularly needed (Huisman et al., 2019; Liu & Carless, 2006). Second, peer feedback can allow students to develop feedback skills, which is considered

particularly important for higher education students who need to become independent lifelong learners (Tai et al., 2018; Wanner & Palmer, 2018). Third, higher education may be considered the optimal level to organise peer feedback given the levels of maturity and competence of young adults (Panadero & Brown, 2017).

In this general introduction, I will first define peer feedback and discuss its benefits and students' perceptions and concerns related to it. Then, I will dive into the interpersonal processes of peer feedback, while paying particular attention to trust and psychological safety. Next, I will argue that training could be an effective way to increase students' perceptions of trust and psychological safety. Finally, I will present the dissertation's general objective and provide an overview of the four studies that were conducted to achieve this objective.

1. Peer feedback

1.1 Definition of peer feedback

To understand what peer feedback is, it is necessary to understand what feedback is. Various definitions of feedback have been proposed (e.g. Hattie & Timperley, 2007; Sadler, 1989; Shute, 2008). In this dissertation, I will use the definition proposed by Lipnevich and Panadero (2021, p. 25):

> Feedback is information that includes all or several components: students' current state, information about where they are, where they are headed and how to get there, and can be presented by different agents (i.e., peer, teacher, self, task itself, computer). This information is expected to have a stronger effect on performance and learning if it encourages students to engage in active processing

This definition, which was developed as part of a literature review on feedback models and theories, is more comprehensive than previous definitions because it combines the five key elements identified in fourteen previous definitions (e.g., Evans, 2013; Narciss, 2008; Nicol & MacFarlane-Dick, 2006). It integrated the ideas that (1) feedback is information, (2) feedback intends to close a gap between a goal and a student's current performance, (3) feedback can be provided by different agents, (4) feedback has to be processed, and (5) students are active and not passive recipients of feedback.

Peer feedback refers to activities with a formative purpose only. It is often considered as a sub-category of peer assessment given that peer assessment can refer to assessments that are only formative (i.e. peer feedback), only summative (i.e., peer grading) or have both a formative and summative purpose (Liu & Carless, 2006). Various definitions of peer assessment have also been proposed (e.g., Falchikov, 1995; Topping, 2017). In this dissertation, I will follow the definition of peer assessment proposed by Alqassab et al. (2023, p. 2), which reflects current perspectives on peer assessment in the research community:

> [Peer assessment is] a learning phenomenon where individuals or social constellations (e.g., pair, group, team, community)—within a physical and/or virtual environment— exchange, react to, interact and/or act upon information about individual performance and/or individual

> contributions to the process and/or product of a social constellation, with the purpose to accomplish implicit or explicit shared and individual learning goals (e.g., domain-specific knowledge or skills, social skills, etc.)"

Therefore, peer feedback can refer to "information" from peers when following the feedback definition of Lipnevich and Panadero (2021), or to "a learning phenomenon" when considering it a subcategory of peer assessment as defined by Alqassab et al. (2023). The existence of these two different meanings echoes a frequent debate among feedback researchers: two conflicting meanings of the word "feedback" exist, an information-centric understanding and a processes-centric understanding (Lipnevich & Panadero, 2021; Winstone et al., 2022). In the information-centric understanding of feedback, feedback is considered as information presented to the learners (e.g., a written comment from a teacher) that they can use to improve their work or their learning strategies. In the process-centric understanding of feedback, feedback is considered as an interactive process between a learner and someone else (e.g., a teacher or a peer) and includes the actions undertaken by the learner to obtain, understand, and act upon the received information (Lipnevich & Panadero, 2021; Winstone et al., 2022).

Considering that the two meanings have value and to avoid confusion, I will follow the suggestion by Winstone et al. (2022) to specify whether I talk about "feedback information" or "feedback process". Therefore, in this dissertation, "peer feedback-information" will be used to refer to feedback as defined by Lipnevich and Panadero (2021), but specifically provided by peers, and "the peer feedback-process" will be used to refer to a formative form of peer assessment as defined by Alqassab et al. (2023). I will use "peer assessment" when summative assessment is also involved or when the studies mentioned did not clarify the formative or summative purpose of the assessment.

Compared to teacher feedback, the peer feedback-process and peer feedback-information present specific characteristics (Panadero et al., 2018). On the one hand, these specific characteristics make the peer feedback-process particularly beneficial for students' learning, as will be developed in the next section (1.2). On the other hand, these characteristics also cause specific concerns for students, as will be developed in section 1.3.

1.2 Benefits of peer feedback

The educational value of the peer feedback-process primarily lies in the fact that it has a positive impact on students' learning. Three recent meta-analyses have investigated the effectiveness of peer assessment and they all concluded that it is an effective way to foster students' learning (Double et al., 2020; Huisman et al., 2019; H. Li et al., 2020). Two of these meta-analyses have even found that the effect of peer assessment on learning is greater than that of teacher assessment (Double et al., 2020; H. Li et al., 2020), while the third meta-analysis showed a comparable positive effect of teacher and peer assessment (Huisman et al., 2019).

This positive effect of the peer feedback-process on learning can be explained by two main reasons. First, when participating in peer feedback activities, students receive feedback-

information. Whether it comes from a teacher or a peer, feedback-information has the potential to foster learning by guiding students to close the gap between their actual and desired performance (Hattie & Timperley, 2007). A meta-analysis of 435 studies (Wisniewski et al., 2020) found that feedback-information has a moderate size effect on students' learning. This effect is heterogeneous – some feedback-information being more powerful than others (e.g., feedback-information that contains more information has more impact) – but shows the potential of feedback-information. A significant moderator of the effect of feedback-information is its direction: student-to-student feedback-information is more efficient than teacher-to-student feedback-information (Wisniewski et al., 2020). Students are often able to explain what they have just learned in a more accessible and encouraging way than teachers (Nicol & Macfarlane-Dick, 2006), which could explain the superiority of the feedback-information they provide.

The second reason explaining the positive impact of the peer feedback-process on students' learning is that students perform the role of assessee (i.e., receptor of feedback-information) and assessor (i.e., provider of feedback-information) and can learn from both. Students do not only revise their work based on the feedback-information they receive but also based on the feedback-information they provide (Colognesi & Deschepper, 2018; Nicol et al., 2014). Results from quasi-experimental studies directly comparing the role of assessee and assessor suggest that it is the latter role (which consists of both seeing exemplars of their peers' assignments and providing feedback to these peers) that allows students to learn the most (Çevik, 2015; Lundstrom & Baker, 2009).

Besides fostering students' learning, the peer feedback-process can contribute to the development of students' feedback skills (Carless et al., 2011). According to the student feedback literacy framework (Molloy et al., 2020), student feedback literacy consists of 31 components, which can be categorised into seven groups: commits to feedback as improvement, appreciates feedback as an active process, elicits information to improve learning, processes feedback information, acknowledges and work with emotions, acknowledges feedback as a reciprocal process, and enacts outcomes of processing feedback information. These various components are considered necessary for the feedback-process to reach its full potential (Carless & Boud, 2018). Through the peer feedback-process, students can develop these features because it provides them with opportunities to take responsibility in the feedback-process, to dialogue around feedback-information, to make judgments, and to act upon feedback-information (Carless & Boud, 2018). Therefore, participating in peer feedback activities can help students to benefit more from every feedback-process. Moreover, it also helps them to better assess themselves and to become independent lifelong learners (Liu & Carless, 2006; Tai et al., 2018).

Finally, from a practical point of view, the peer feedback-process has the advantage of being feasible even with many students. Even though implementing peer feedback activities can be time-consuming for teachers, the time investment only slightly increases as the number of students increases, contrary to the time required by teachers to provide feedback to each student which is directly multiplied by the number of students. Therefore, the peer

feedback-process offers a scalable solution for students in higher education to receive fast and numerous feedback-information (Er et al., 2021).

1.3 Students' perceptions and concerns about the peer feedback-process

Most students seem to recognize the value of the peer feedback-process. Indeed, several studies investigating students' perceptions reported that they find peer assessment to be useful: they feel that it motivates them to produce high-quality work and helps them improve their learning (Hanrahan & Isaacs, 2001; Mulder et al, 2014). However, students' perceptions are not only positive. They also express negative feelings linked to the peer feedback-process, such as fear that their peers will be biased or will not put enough effort into their assessments (e.g., Hanrahan & Isaacs, 2001), discomfort with having their work disclosed to their peers (e.g., Mulder et al., 2014) or concerns about the fairness of peer assessment (e.g., Zou et al., 2018). These concerns are not anecdotal. For example, a study reported that 40% of students found peer assessment to be an uncomfortable experience (Stanier, 1997), while according to another study, 29% of students believed that their peers did not engage enough in the activity and 19% did not trust their peers as assessors (Mostert & Snowball, 2013).

Students' concerns about the peer feedback-process are not without consequences: despite the benefits of the peer feedback-process, students can be reluctant to fully participate in it and strongly prefer to receive feedback-information from teachers rather than from peers (Alharbi & Alqefari, 2022; Mahvelati, 2021). Some studies suggest that students' perceptions of peer assessment affect the way they are involved in the activity or the outcomes of peer assessment. For example, a study found that students with more positive attitudes toward peer assessment provided longer feedback-information comments (Zou et al., 2018), while another study found that students who were more motivated to participate in peer assessment made fewer errors while assessing their peers (Guo & Lei, 2020).

2. Interpersonal processes of the peer feedback-process

As seen in the previous section, the peer feedback-process has a lot of potential for enhancing student learning, but it is also associated with concerns among students. These concerns can have negative consequences on the peer feedback-process and are therefore important to understand. Even though they are varied, they emerge from a common point: the fact that the peer feedback-process is a social experience (van Gennip et al., 2009; Wilson et al., 2015). Indeed, the peer feedback-process necessarily involves a student providing feedback and another student receiving this feedback. This interaction between students exchanging feedback produces thoughts, actions, and emotions that are important to consider (Panadero, 2016). The present section focuses on interpersonal processes of the peer feedback-process.

2.1 Conceptualisation of interpersonal processes in the peer feedback-process

The research on interpersonal processes of the peer feedback-process is fairly recent. Some of the first researchers to study peer assessment from a social perspective were van Gennip and colleagues (2009; 2010). Based on the literature on collaborative learning and group work, they identified four interpersonal variables that could be of importance in peer assessment (psychological safety, value congruency, interdependence, and trust) and conducted a literature review to investigate the impact they have on peer assessment outcomes (van Gennip et al., 2009). However, they did not find any study allowing them to answer this question, which led them to conclude that there was a lack of studies investigating the impact of interpersonal variables on peer assessment. This conclusion was also shared by another review published the same year deploring the lack of attention to interpersonal processes in peer assessment practices (Strijbos et al., 2009).

To confirm their hypothesis that these four variables are of importance, van Gennip, Segers, and Tillema (2010) investigated their impact on secondary-vocational students' conceptions of peer assessment and perception of learning. They found that a high level of psychological safety and trust (in the self and in the other) was positively associated with perceived learning and that the conceptions of peer assessment mediated this effect. For value congruency the pattern was more complex: it had both a positive direct association and a negative indirect association with students' perceived learning (more unanimity led to less positive conceptions of peer assessment, which in turn led to lower perceived learning). There was no significant relationship between interdependence and conceptions of peer assessment or between interdependence and perceived learning.

In the last fifteen years, the research interest in the interpersonal processes of peer feedback-process has expanded with more researchers diving into this field (e.g., K. Cheng & Tsai, 2012; Panadero, 2016; Rotsaert et al., 2018). Besides "interpersonal variables" (van Gennip, Segers & Tillema, 2010, p. 281), other terms have been used to label these processes, such as "social aspects" (Panadero, 2016, p. 247) or "social factors" and "social process" (Topping, 2017, p. 13). Moreover, additional interpersonal variables have been studied, such as perceived fairness, experienced peer pressure, experienced fear of disapproval, and feelings of discomfort (Raes et al., 2015; Vanderhoven et al., 2015). Yet, most of the terms used to label interpersonal processes of peer feedback-process have not been clearly defined and it is unclear which concepts they refer to.

Recently, a conceptualization in two categories has been proposed (Panadero et al., 2023). In a literature review on the subject, which highlighted a lack of clear and agreed-upon definitions, Panadero et al. (2023) proposed to talk about intrapersonal and interpersonal factors of peer assessment. Intrapersonal factors refer to "elements of students' own processes or that influence these individual processes", while interpersonal factors refer to "elements of the peer interaction or that influence the interaction in the peer assessment process" (Panadero et al., 2023, p. 1057). These authors consider that there are four intrapersonal factors (motivation, emotion, trust in the self as an assessor, and fairness /

(dis)comfort) and five interpersonal factors (social connections, trust in the other as an assessor, psychological safety, value diversity/congruence, and interdependence).

The conceptualization proposed by Panadero and colleagues (2023) has the advantages of setting clear boundaries to what interpersonal factors are and of clearly defining these factors, but it has the disadvantage of being quite restrictive. Indeed, some interpersonal processes that have been studied by other researchers are not included, such as fear of disapproval or peer pressure (e.g., L.Li, 2017; Vanderhoven et al., 2015). Moreover, the choice of the factors seems to be European-centric. It has been argued that perceptions of and reactions to peer feedback-process are influenced by students' culture (e.g., Fithriani, 2018). For example, people coming from collectivist cultures, which are more common in Asian countries, tend to work toward maintaining group harmony and mutual face-saving (Hofstede, 1989) and several studies conducted with Chinese students confirmed that students evoked concerns about face-saving during the peer feedback-process (S. Kim & Lan, 2021, Xue et al., 2023; F. Zhang et al., 2023). Moreover, according to the same studies, these concerns can impact students' preferences for the organisation of the activity. Face-saving is therefore an important interpersonal process in Asian countries, yet it is not included in the interpersonal factors identified by Panadero and colleagues (2023).

In this dissertation, I will use the term "interpersonal processes" as a broader term that encompasses the previously used terms (e.g., interpersonal variables, interpersonal factors, social aspects). I propose the following working definition of interpersonal processes: processes that emerge from the fact that peer feedback-process involves interactions between students.

2.2 The importance of trust and psychological safety

The current dissertation focuses on three interpersonal processes more specifically: trust in the self as an assessor, trust in the other as an assessor, and psychological safety. The choice was based on three reasons. First, these interpersonal processes seemed to be of particular importance according to professors organising peer feedback activities. Before conducting the first study of this dissertation, I conducted eleven unstructured interviews with university professors at UCLouvain who (used to) implement peer feedback activities in their courses to have an idea of the obstacles they encountered. The professors confirmed that students are often reluctant to participate in the peer feedback-process and reported feeling uncomfortable or worried. The most frequent concerns that students reported to them were linked to trust in the other (e.g., students doubted that their peer would be able to provide fair and objective feedback-information and when they received feedback-information from both peers and teaching assistants, they only paid attention to the latter). Other concerns that students reported according to the interviewed professors could be linked to trust in the self (e.g., students asking who they were to provide feedback to other students) and psychological safety (e.g., students explaining that they do not feel safe to participate in peer feedback activities with peers they do not know well).

Second, concerns associated with trust in the other, trust in the peer, and psychological safety are often reported in the literature, which suggested that what professors reported in the interviews reflected students' experiences (e.g., L. Li & Steckelberg, 2013, McConlogue, 2012; Wilson et al., 2015). For example, in a study conducted on first-year students in Taiwan, approximately half of the students reported that they did not trust their peers' abilities to provide feedback, a quarter of students that they did not trust their own abilities, and a fifth of students that they did not feel psychologically safe during the peer feedback-process (K. Cheng & Tsai, 2012).

Finally, trust in the other, trust in the self, and psychological safety seem to play important roles in the peer feedback-process as they are associated with important outcomes of peer assessment, such as students' perceived learning, a proxy of real learning (van Gennip, Segers & Tillema, 2010). The link between trust and peer assessment outcomes will be developed in more detail in the next two sections (section 2.3 for trust in the self and in the other and section 2.4 for psychological safety).

2.3 Trust in the self and in the other as an assessor

In the context of the peer feedback-process, the term "trust" was first used in the sense of the trust that students have in their ability to assess the work of other students – trust in oneself – and in their peers' ability to do the same – trust in peer – (van Gennip, Segers & Tillema., 2010). Later, more precise definitions were proposed by Panadero and colleagues (2023), which defined trust in the self as an assessor as a "belief about the ability to perform peer assessment as an assessor" and trust in the other as an assessor as the "confidence in a peer's capability to perform a fair and/or accurate peer assessment" (Panadero et al., 2023, p. 1057). Trust in the self and trust in the other as an assessor are two distinct concepts (van Gennip, Gijbels et al., 2010): students can trust their abilities but not their peers' abilities, and vice versa. Some studies suggest that students are more concerned about their peers' abilities to provide feedback-information than about their own abilities, even though these concerns are also non-negligible (e.g., K. Cheng & Tsai, 2012; L. Li & Steckelberg, 2013).

In the definition of trust in the other as an assessor (Panadero et al., 2023), two aspects are present: the trust in peers' abilities to perform (1) a fair and (2) an accurate assessment. This mention of fair and accurate assessment echoes Carless's (2013) idea that, in the context of peer assessment, trust is a concept with two main components. Indeed, Carless (2013) explained that two of the components of trust identified by Reina and Reina (2015) are relevant in the context of the feedback-process: competence trust and communication trust. Competence trust refers to the trust that a person can efficiently carry out a task, while communication trust refers to the trust that a person will share information transparently with the interests of others at heart (Reina & Reina, 2015).

Applied to the context of the peer feedback-process, competence trust refers to the belief that peers are able to provide accurate feedback-information (i.e. understand the course content, know how to apply the assessment criteria, etc.), which corresponds to the second aspect of Panadero et al.'s (2023) definition of trust in the other as an assessor (i.e., providing

accurate feedback-information). Communication trust refers to the belief that peers will be fair and provide honest feedback-information, with the sole purpose of helping other students to learn, which corresponds to the first aspect of Panadero et al.'s (2023) definition of trust in the other as an assessor (i.e., providing fair feedback-information). This dual aspect of trust is supported by studies (e.g., K. Cheng & Tsai, 2012; McConlogue, 2012; Wilson et al., 2015) showing that some students not only cast doubt on their peers' knowledge (competence trust) but also on their peers' ability to make objective judgments and that students who mistrust their peers often complain that peer assessment is unfair (communication trust).

In contrast to Panadero et al. (2023) who did not include the aspect of fairness in the definition of trust in the self as an assessor, I consider that this dual aspect is also relevant for the self. Indeed, students can also doubt that they will be able to be objective when providing feedback-information to their peers (e.g., Hanrahan & Isaacs, 2001). Therefore, I will use the following working definitions of trust:

- Trust in the self as an assessor is the belief that students have about their abilities to provide fair and accurate feedback-information to their peers.
- Trust in the other as an assessor is the belief that students have about their peers' abilities to provide fair and accurate feedback-information to them.

In line with other researchers (e.g., K. Cheng & Tsai, 2012; van Gennip et al., 2009), and contrary to Panadero et al.'s (2023) classification, I consider that both trust in the self as an assessor and trust in the other as an assessor are interpersonal processes. In both cases, the question of trust emerges from interactions between students who provide and receive peer feedback-information.

A few studies have investigated the impact of trust in the self and in the other on students' perceptions and on peer assessment outcomes. Trust in the self as an assessor and trust in the other as an assessor are both positively associated with students' conceptions of peer assessment and students' perceived learning (van Gennip, Segers & Tillema, 2010). In addition, trust in the self and trust in the other are both associated with deeper learning approaches and trust in the self is associated with a tendency to see learning as linking new content to prior knowledge and not just as an accumulation of knowledge (K. Cheng & Tsai, 2012). Finally, compared to students who do not trust themselves and their peers, students with higher perceptions of trust in the self and trust in the other value peer assessment more (Rotsaert et al., 2017) and students with more positive attitudes toward peer assessment, which includes trust in the self, provide longer feedback-information comments to their peers (Zou et al., 2018).

2.4 Psychological safety

Psychological safety is a notion that initially came from organisational psychology and is defined as the "perceptions of the consequences of taking interpersonal risks in a particular context" (Edmondson & Lei, 2014, p. 24). The interpersonal risks mentioned in the definition are the risks of being perceived as ignorant, incompetent, negative, or disruptive

(Edmondson, 1999). If a working environment is perceived as psychologically safe, employees should be able to adopt risky behaviours – such as admitting errors, asking for help, criticizing events, or seeking information – without worrying about possible negative reactions from other members of their team (Edmondson, 2003). These risky behaviours are essential to teams and individual learning, hence psychological safety is seen as a prerequisite for learning (Edmondson, 1999). Indeed, evidence supports that psychological safety is associated with feedback-information seeking in a working context (De Stobbeleir et al., 2020) and with academic performance in an educational context (Soares & Lopez, 2020).

In the context of the peer feedback-process, psychological safety can be 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. 1057). Psychological safety is required in peer feedback activities (van Gennip et al., 2009) because the peer feedback-process is a situation that carries interpersonal risks for students; for example, they can be judged as incompetent by providing feedback-information with mistakes or be judged as negative by providing critical feedback.

The above-mentioned studies that have investigated the impact of trust on the peer feedback-process also investigated the impact of psychological safety. 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, Segers & Tillema, 2010). Moreover, students who report higher levels of psychological safety adopt deeper learning approaches in peer assessment and tend to see learning as linking new content to prior knowledge and not just as an accumulation of knowledge (K. Cheng & Tsai, 2012). However, psychological safety is not associated with the perceived educational value of peer assessment (Rotsaert et al., 2017).

Even if the studies on trust and psychological safety only revealed correlations, with no way of affirming which variable causes which other variable, their findings support the idea that trust and psychological safety play an important role in the peer feedback-process (K. Cheng & Tsai, 2012; Rotsaert et al., 2017; van Gennip, Segers et Tillema, 2010). This raises the question of how to design peer feedback activities to ensure that students feel psychologically safe and have high levels of trust in the self and in the other as an assessor.

3. Training as an impactful instructional design element of the peer feedback-process

When teachers organise peer feedback activities, they have to choose how to design these activities. The eleven university professors I interviewed to help me design the studies for this dissertation seem to be primarily concerned with the issue of anonymity. They discussed the potential advantages and disadvantages of concealing the identity of assessors and assessees from their peers. The issue of anonymity has also been well studied in the literature (for a review, see Panadero & Alqassab, 2019). However, it is far from being the only choice to make as peer feedback activities can be designed in different ways: they can be summative

or formative only, face-to-face or online, voluntary or compulsory, organised in or out of class, etc.

To better understand this diversity, several typologies of peer assessment have been proposed (e.g., Gielen et al., 2011; Topping, 1998). The more recent one, proposed by Alqassab et al. (2023), is based on previous typologies and contemporary reviews. They identified important design elements of peer assessment and regrouped them into four categories: context (e.g., subject domain), instructional design (e.g., number of assessors), outcomes (e.g., students' performances), and moderators/mediators (e.g., gender). The second category, instructional design elements, is particularly interesting because they refer to "what designers can manipulate" (Alqassab et al., 2023, p. 18). Contrary to other elements that are out of control, such as students' age or gender for example, teaching staff can have a direct impact on these elements when designing peer feedback activities.

Among the design elements identified (Alqassab et al., 2023), one seems particularly promising to ensure trust and psychological safety: the provision of training. To the best of my knowledge, no previous study has ever investigated the efficacy of a training session targeting students' perceptions of psychological safety and trust. However, such training could be an efficient way to ensure that students have high levels of psychological safety, trust in the self as an assessor, and trust in the other as an assessor for three reasons that will be detailed in the upcoming sections. First, it has been found that general peer feedback-process training – training that does not specifically target interpersonal processes – positively affects peer assessment outcomes (e.g., van Zundert et al., 2010), which shows the potential of peer feedback training. Furthermore, some evidence suggests that general peer feedback-process training could also positively affect students' perceptions of trust (Ocampo et al., 2023; Parikh & Sheehan, 2016). Finally, L. Li (2017) found that peer feedback-process training specifically designed to ease interpersonal concerns had a positive impact on students' perceptions of peer pressure, which suggests that a training session could also effectively target other interpersonal processes.

3.1 Efficacy of peer feedback-process training

To prepare students to actively engage in peer assessment activities and to develop their peer assessment skills, training is often designed (van Zundert et al., 2010). Important skills in the context of peer assessment are the ability to identify what is important to assess, the ability to judge the performance of a peer, and the ability to provide feedback-information for future learning (Sluijsmans et al., 2004). To achieve these aims, most peer assessment training features similar components, such as general discussion on peer assessment, lectures on how to provide effective feedback-information, group and individual exercises, modelling, and practice sessions (e.g., Alqassab et al., 2018a; Min, 2005; Ocampo et al., 2023)

Multiple pieces of evidence suggest that training is beneficial for various outcomes of peer assessment (e.g. Min, 2005; Sluijsmans et al., 2004). The meta-analysis conducted by H. Li et al. (2020) revealed that the presence of a training is the most critical moderator of

the efficacy of peer assessment to foster students' learning. Moreover, a review of peer assessment (van Zundert et al., 2010) reported that training has a positive impact on the inter-rater agreement between students, on students' attitudes toward peer assessment, and on students' peer assessment skills (as evidenced by the quality of the provided feedback).

3.2 Impact of peer feedback-process training on trust

Two studies suggest that general training could also positively impact the trust that students have in their own and in their peers' abilities (Ocampo et al., 2023; Parikh & Sheehan, 2016). Parikh and Sheehan (2016) provided a short training session to explain to students how to provide feedback-information to their peers. Students participating in this study reported high levels of trust in the self as an assessor and trust in the other as an assessor after the training. However, in the absence of a pre-test and a control group in the study, it is not possible to be sure that it was the training that positively affected students' perceptions of trust.

This limit was overcome in Ocampo et al.'s (2023) study. In this study, students were randomly allocated to a control or an experimental condition. Students in the control group received a 30-minute training to learn how to use a peer feedback tool (*Eduflow*), while students in the experimental group received a 180-minute training that also allowed them to practice providing peer feedback-information. After the training, students were led to believe they provided feedback-information on their peers' work. However, they all provided feedback-information to the same three writing samples created by the research team. The gender that was indicated in the writing samples was manipulated to investigate if the assessee's gender was of importance. The study revealed a positive impact of the training on students' perceptions of trust in the self as an assessor, but only for men providing feedback-information to women.

3.3 Impact of specific training on interpersonal processes

First evidence from the only study on the subject suggests that a training specifically designed to ease interpersonal tensions can positively affect some interpersonal processes. L. Li (2017) created a training session that aimed to ease the pressure students can experience from their peers during non-anonymous peer feedback activities. First, students watched a video explaining the purpose, characteristics, and learning benefits of the peer feedback-process. The reasons why students could feel pressure and tension during the peer feedback-process were also discussed in the video, alongside an explanation of what was put in place to mitigate the possible negative impact of peer pressure. Next, students had the opportunity to express their concerns through an online forum, and then, to discuss these concerns with their teacher. Finally, students were encouraged to discuss any remaining concerns in small groups until they were all addressed.

To investigate the efficacy of this training, students were assigned to one of three groups: (1) a group where students were anonymous and did not receive the specific training, (2) a

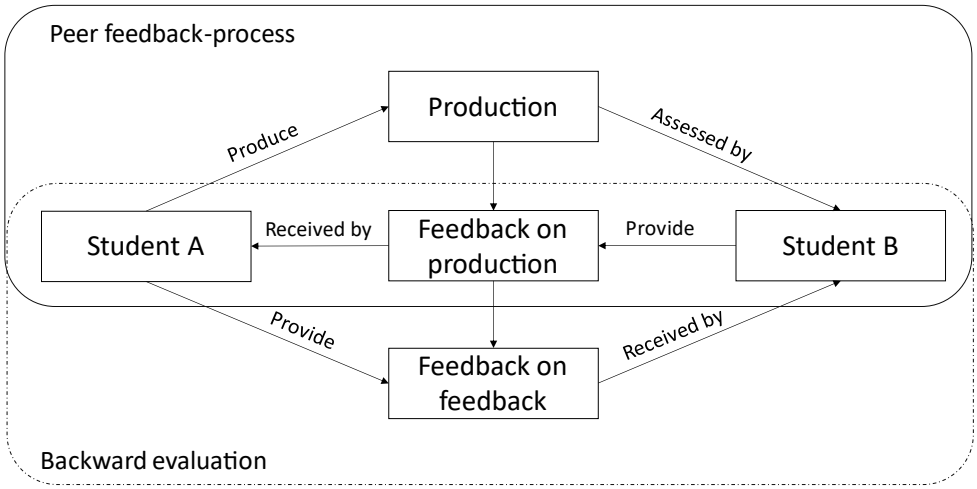
group where students were non-anonymous and did not receive the specific training, and (3) a group where students were non-anonymous and received the specific training. Students who were anonymous or who received the training improved their performance more than students who were non-anonymous and did not receive the training. Moreover, students who received the training valued the peer feedback-process more and experienced less peer pressure than students who did not receive the training, regardless of whether they were in the anonymous or non-anonymous group (L. Li, 2017).

Taken together, these results lead us to believe that creating a training session targeting trust and psychological safety could be an effective way to ensure that students feel safer and have enough trust in their own abilities and in their peers' abilities prior to a peer feedback activity. However, a limitation of peer feedback-process training is that it is often limited to one 2- or 3-hour session (e.g., Alqassab et al., 2018a; Min, 2005; Ocampo et al., 2023), which does not allow the combination of various training methods, although a combination of training methods enhances training effectiveness (Martin et al., 2014). A comprehensive training with several sessions could be more effective but seems unrealistic to implement in higher education's busy course schedules. Another option is to reinforce training with other activities that share the same objectives but do not take time in class, such as backward evaluation.

4. Backward evaluation as a promising complement to training

Backward evaluation is a term used "to describe the feedback that an author provides to a reviewer about the quality of the review" (Luxton-Reilly, 2009, p. 226). When a backward evaluation is used an additional step is added to the peer feedback-process (see Figure 1): Student A produces a work that is reviewed by Student B, Student B provides feedback-information to Student A, and Student A provides feedback-information on the received feedback-information to Student B so that they can reflect on the quality of the feedback-information they provided (Misiejuk & Wasson, 2021). It is hypothesised that backward evaluation helps students to provide meaningful feedback-information, ensures that students process the received feedback-information, and increases students' engagement and reflection (Luxton-Reilly, 2009; Misiejuk & Wasson, 2021).

The value of backward evaluation lies in the fact that it increases the level of interactions between students by allowing students to react to the feedback they received and to know how the feedback they provided was perceived. In this sense, backward evaluation can be considered a first step towards a dialogic feedback approach (Zhao et al., 2024). Conceptualising the feedback-process as a dialogue where students can discuss the feedback-information with their teachers or peers instead of seeing it as a mere transmission of information is established as an important principle of effective feedback practices (Nicol & Macfarlane-Dick, 2006). In the context of the peer feedback-process more specifically, the presence of dialogue around feedback-information helps students understand each other and increases their active engagement (Zhu & Carless, 2018).



Note. Inspired from Misiejuk & Wasson (2021).

Figure 1. Backward evaluation

Most studies on backward evaluation implement it for research purposes, using backward evaluation data (i.e., students' perceptions of the received feedback-information) to answer other research questions (Misiejuk & Wasson, 2021). Therefore, the impact of backward evaluations on the peer feedback-processes or its outcomes is understudied. Moreover, as a wide range of terms is used to design backward evaluation (e.g., backward feedback, feedback-to-feedback, reciprocal feedback, back-review feedback, double-loop feedback, meta-feedback, etc., Misiejuk & Wasson, 2021), studies on the subject can be difficult to find. To the best of my knowledge, there are no studies investigating the impact of backward evaluation on interpersonal processes of the peer feedback-process. However, results from two studies (M. Kim, 2009; Ng & Yu, 2023) suggest that backward evaluation could enhance the peer feedback-process. Findings from a quasi-experimental study revealed that students who participated in a backward evaluation improved their performance more than students who participated in the peer feedback activity without a backward evaluation. They also reported more motivation to participate in peer feedback activities and more metacognitive awareness of the learning process (M. Kim, 2009). Moreover, in a qualitative study, students reported that knowing there would be a backward evaluation incited them to be more respectful and serious during the peer feedback activity (Ng & Yu, 2023).

Backward evaluation could particularly be a useful complement to a training targeting the interpersonal processes of peer feedback-process to increase students' perceptions of trust in the self and the other as an assessor. Receiving feedback-information on their feedback-information could help students to better judge their abilities as assessor, and, if needed, help them to understand how they could provide more effective feedback-information. This could increase their perceptions of trust in the self as an assessor. Moreover, to judge the received feedback-information, students must read it, process it, and reflect on it (Misiejuk & Wasson, 2021), three essential actions that students do not always do otherwise. Indeed, studies

suggest that many students do not even read the feedback-information they receive (Orsmond et al., 2005; Sinclair & Cleland, 2007). By forcing students to take these steps, backward evaluation could push them to see the merits of the feedback-information provided by their peers and to develop higher levels of trust in the other as an assessor.

5. Dissertation's objective and studies overview

This **general introduction (Chapter 1)** illustrates that the question of interpersonal processes of peer feedback-process is a burgeoning research area in which there is still much to investigate. It is known that interpersonal processes, such as trust and psychological safety, play a role in the peer feedback-process and should be considered when designing peer feedback activities (e.g., K. Cheng & Tsai, 2012; Topping, 2017; van Gennip, Segers & Tillema, 2010). However, the question of how peer feedback activities should be designed to optimize trust and psychological safety is still open given the scarcity of studies on the subject. This question is of interest to both researchers and practitioners. From a research perspective, it helps to reduce a knowledge gap: how to design peer feedback activities to ensure trust and psychological safety. From a practical perspective, it can help teacher teams to design effective peer feedback activities. Therefore, the broad research question underlying this dissertation is the following:

How can peer feedback activities be designed to ensure that higher education students feel psychologically safe and trust their own and their peers' abilities?

The first step to answering this research question is to take stock of the current state of knowledge. Hence, **Study 1** (**Chapter 2**) is a **systematic literature review** reviewing studies that have investigated how the instructional design elements of peer assessment influence interpersonal processes. Given the small number of studies that could directly inform our general research question, a broad approach was adopted for this review that included studies conducted in any educational context, in both formative and summative peer assessment, and targeting any interpersonal processes. This review is necessary because previous literature reviews on the interpersonal processes of peer assessment (Panadero, 2016; Panadero et al., 2018; Panadero & Alqassab, 2019, Panadero et al., 2023; Strijbos et al., 2009; Topping, 2017; van Gennip et al., 2009) provided little information on how peer assessment could be designed to minimise the negative effects of interpersonal processes. The review's results allow us to discover which instructional design elements have already been manipulated to affect interpersonal processes, and with which results.

The results of the first version of this study were used to inform and guide studies 2, 3, and 4: as they highlighted the potential positive role of training, I decided to create a training session targeting interpersonal factors. There have been three versions of this training session, one for each study (studies 2, 3, and 4). In each empirical study, students had to provide feedback on assignments produced by students individually, never on group work, because the peer feedback-process concerning group work involves more complex processes (Gabelica et al., 2012). As there is currently little research on the impact of design elements of peer feedback-process on interpersonal processes, it seems important to start by

understanding how these interpersonal processes can be affected in less complex situations before introducing more complexity.

In **Study 2** (**Chapter 3**), the training was designed to be implemented online. Having the possibility to provide peer feedback training online is important given that peer feedback is often used in online courses, such as MOOC, and that students have little incentive to take the peer feedback-process seriously in these contexts (Suen, 2014). Study 3 consists of a **description of the training and an explanation of its design process**. Moreover, through the analyses of data from questionnaires and semi-directed interviews investigating students' perceptions of the training, some insights into the way the training was received by 27 second-year bachelor students in motricity sciences were provided.

To test the efficacy of a face-to-face version of the training session, **Study 3** (**Chapter 4**) is a **quasi-experimental study** conducted with 89 second-year bachelor students in motricity sciences. Data from two cohorts of students were collected: a control group in 2019-2020 and an experimental group that received the training in 2021-2022. Quantitative analyses were conducted to investigate the impact of this training session on (1) students' perceptions of trust and psychological safety, (2) the quality of feedback-information provided by students, and (3) students' performances.

In **Study 4** (**Chapter 5**), an improved version of the intervention was designed and it was tested with a methodology that allowed me to overcome Study 3's limitations. Two important adjustments were made to have a stronger intervention: (1) psychological safety was set aside and the training was adapted to focus more heavily on trust and (2) the intervention was reinforced with a backward evaluation. Concerning the methodology, students from the same cohorts were drawn to compose more comparable control and experimental groups and a mixed-method approach was used to better understand students' experiences of and reactions to the intervention. Therefore, study 4 is a **mixed-method quasi-experimental** study conducted on 189 master education students. Quantitative data was analszed to investigate if students' perceptions of trust evolved more positively in the experimental group than in the control group and if students in the experimental group were more inclined to participate in following peer feedback activities than students in the control group. Qualitative data were analysed to explain the results of the quantitative analyses and to better understand students' perceptions and experiences.

Finally, **Chapter 6** is a **general discussion** in which I integrate the results of the four studies and discuss them, discuss the dissertation limitations, propose perspectives for further research, and provide implications for practice.

This dissertation is a collection of articles. Even though the studies presented in Chapters 2 to 5 are successive based on the findings of each study, each one was written to be read on its own. Hence, repetition and overlap between chapters are inevitable. An overview of the studies is available in Table 1.

Table 1. Overview of the studies

	Objectif	Interpersonal processes	Study type	Context and sample	Method and data collection	Variables measured
Study 1 (Chapter 2)	Reviewing studies that have investigated how the instructional design elements of peer assessment influence interpersonal processes	All interpersonal processes	Literature review	/	/	/
Study 2 (Chapter 3)	Designing a training session targeting trust and psychological safety	1) Trust in the self 2) Trust in the other 3) Psychological safety	Design and pilot study	Bachelor in motricity sciences at UCLouvain, 27 students	Mixed-method study Questionnaires and semi-directed interviews	1) Reactions to the intervention: students' perceptions of the training session 2) Attitudes: / 3) Behaviours:/
Study 3 (Chapter 4)	Investigating the efficacy of a training session targeting trust and psychological safety	1) Trust in the self 2) Trust in the other 3) Psychological safety	Intervention study with a pre-and post-test quasi-experimental design (in 2 cohorts)	Bachelor in motricity sciences at UCLouvain, 89 students (47 in the control group and 42 in the experimental group)	Quantitative study Questionnaires, feedback provided by students, and students' performances before and after the peer feedback activity	1) Reactions to the intervention: / 2) Attitudes: students' perceptions of trust in the self, trust in the other, and psychological safety 3) Behaviours: type of feedback provided by students and students' performance improvement
Study 4 (Chapter 5)	Investigating the efficacy of an intervention composed of a training session targeting trust and a backward evaluation	1) Trust in the self 2) Trust in the other	Intervention study with a pre-and post-test quasi-experimental design	Master in educational sciences at UCLouvain, 189 students (94 in the control group and 95 in the experimental group)	Mixed-method study Questionnaires, participation rates in the peer feedback activities, and semi-directed interviews	1) Reactions to the intervention: students' perceptions of the training session 2) Attitudes: students' perceptions of trust in the self and trust in the other 3) Behaviours: participation rate in peer feedback activities

Chapter 2

The impact of peer assessment design on interpersonal processes

A Systematic review

Peer assessment can have a very positive impact on students' learning. However, it also poses some challenges such as trust issues, concerns about its fairness, feelings of discomfort, etc. It is known that the interpersonal processes of peer assessment can affect its outcomes. However, it is less clear how peer assessment should be implemented to prevent this. In this article, we reviewed 27 studies that have investigated how the instructional design elements of peer assessment influence interpersonal processes. A literature search was carried out on Scopus, PsycINFO, and Eric with a search string composed of three parts with keywords linked to (1) peer assessment, (2) educational contexts, and (3) interpersonal processes. The 14552 retrieved records were screened in terms of the inclusion criteria, which resulted in 27 studies being included and analysed through a narrative synthesis. The results revealed that the three most commonly-investigated design elements were privacy, format, and training. Privacy seems to have a positive impact on some interpersonal processes, but a detrimental effect on others. Concerning format, studies suggest that increased amounts of guidance are associated with a positive impact on interpersonal processes, but this result must be interpreted with caution due to the limitations of the studies' research designs. The use of training seems to be beneficial in minimising the negative effects of interpersonal processes. Finally, no conclusion could be drawn on three other design elements that were only investigated in one study each, and with weak research designs. These results allowed us to draw important recommendations for practitioners implementing peer assessment and to identify gaps and needs that could inform future research.

Keywords

Peer evaluation, literature review, interpersonal processes, instructional design elements

The impact of peer assessment design on interpersonal processes

A systematic review

1. Introduction

1.1 Peer assessment and its challenges

Formative assessment has the power to foster students' learning (Wisniewski et al., 2020). A key component of formative assessment – feedback – can help students close the gap between their actual and desired performances. Consequently, feedback is considered one of the most critical influences on student learning (Hattie & Timperley, 2007). Involving students in formative assessment by asking them to provide feedback to their peers can be particularly profitable for students' learning, given that peers are often conceptualised as potential mediators of students' learning (Panadero et al., 2018).

The term "peer assessment" is used to define situations in which students provide feedback to their peers. More precisely, peer assessment can be defined as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners, then learn further by giving elaborated feedback and discussing their judgment with peers to achieve a negotiated agreed outcome" (Topping, 2017, p. 2). Other definitions of peer assessment have been proposed (e.g., Falchikov, 1995; Topping, 1998), but Topping's (2017) definition has the advantage of clearly delimiting the concept of peer assessment, which makes it easier to determine if studies implemented peer assessment and, therefore, should be included in the review.

Recent meta-analyses have found that students can learn more from peer assessment than from teacher assessment (Double et al., 2020; Wisniewski et al., 2020). There are multiple reasons why peer assessment is an effective way to enhance students' learning. Students are often able to explain what they have just learned in a more accessible and encouraging way than teachers (Nicol & Macfarlane-Dick, 2006). Moreover, with peer assessment, students perform both the role of assessee and assessor. Some studies suggest that it is the latter role that allows them to learn the most (e.g., Çevik, 2015; Lundstrom & Baker, 2009).

However, peer assessment does not automatically increase students' learning. Indeed, even though meta-analyses found that peer feedback has a medium (Double et al., 2020) to strong effect (Wisniewski et al., 2020) on students' learning, the effect sizes of each study are heterogeneous (Wisniewski et al., 2020). This variability is not specific to peer feedback. The effect of teacher feedback is also highly variable (Wisniewski et al., 2020), which suggests that feedback is a complex process with many elements involved that can increase or hinder its efficacy. The complexity of feedback was recently captured by the MISCA meta-

model developed by Panadero and Lipnevich (2022). According to this model, five elements are important to take into consideration when implementing or researching feedback. The first element, student characteristics (i.e., gender, prior knowledge, self-efficacy, etc.), is an overlapping element situated in the centre of the model as the learner belongs at the very centre of the feedback process. The four other elements are the message (i.e., the information contained in the feedback), the implementation (i.e., the purpose of the feedback and the processes it triggers in the learner), the context (i.e., the instructional context in which the feedback takes place and the presentation of the feedback), and the agents (i.e., the agents involved in the feedback process, their role, and their interactions). These five elements are interrelated, and need to be aligned to allow the learner to use the feedback in such a way as to enhance their learning (Panadero & Lipnevich, 2022).

The "agents" element of the MISCA model is central to peer assessment (Panadero & Lipnevich, 2022). In terms of this element, the source of feedback plays an important role. Indeed, the specific characteristics of peer assessment pose specific challenges. Students tend to be reluctant to participate in peer assessment and to ignore the feedback they receive (e.g., Wichmann, 2018; Winstone et al., 2017), which risks hindering its efficacy. This can be explained by two fundamental aspects of peer assessment: students do not have the same status and expertise as their teachers (Gielen et al., 2010) and they can have various social connections with their peers (Harris & Brown, 2013). A series of challenges can arise from this, such as trust issues (e.g., Hanrahan & Isaacs, 2001), concerns about the fairness of peer assessment (e.g., Zou et al., 2018), feelings of discomfort during the peer assessment process (e.g., Stanier, 1997), the importance of "face-saving" (e.g., S. Yu et al., 2016), or bias due to the relationships between peers (e.g., Domínguez et al., 2016). These challenges have in common that they emerge from the interpersonal processes of peer assessment.

Different terms are used in the literature to label the interpersonal processes of peer assessment, such as "social aspects" (Panadero, 2016, p. 247), "social factors" and "social process" (Topping, 2017, p. 13), "interpersonal variables" (van Gennip, 2010, p. 281). None of these terms have been clearly defined, and it is not clear which individual variables are associated with each of these terms. In their seminal paper, van Gennip and colleagues (2009) identified four "interpersonal variables": trust, psychological safety, value congruency, and interdependence. In subsequent research, other researchers have studied other variables related to the interpersonal processes of peer assessment, such as perceived fairness, fear of disapproval, and feeling of discomfort (e.g., Raes et al., 2015; Vanderhoven et al., 2015). More recently, Panadero et al. (2023) have proposed a conceptualization in two categories: *intrapersonal factors* and *interpersonal factors*. The latter refers to "elements of the peer interaction or that influence the interaction in the peer assessment process" (Panadero et al., 2023, p. 1057) and only includes five variables (social connections, trust in the other as assessor, psychological safety, value diversity, interdependence). This conceptualization has the advantage of setting clear boundaries to the notion of interpersonal factors, but is quite restrictive. For example, notions that are more relevant in an Asian context than in a European context such as face-saving or group harmony (e.g., Fithriani, 2018; S. Yu et al., 2016), are not included in this conceptualization. In this review, we will adopt a broader conception of

interpersonal processes and include all variables that have been described as "social" or "interpersonal" by the researchers studying them. As a result, some processes that are not included in Panadero et al.'s (2023) conceptualization are included in this review (e.g., face-saving, peer pressure). Moreover, we will not strictly follow Panadero et al.'s (2023) distinction between intra and interpersonal factors, as we consider any process that results from, or is strongly influenced by, the peer interactions that take place in peer assessment to be interpersonal. Therefore, four processes that would be considered intrapersonal factors in Panadero et al.'s (2023) conceptualisation (i.e., emotions due to the interpersonal processes, comfort, trust in the self as an assessor, and perceived fairness) are included in this review. These processes are also usually considered as interpersonal processes by the researchers investigating them (e.g., van Gennip et al., 2009; van Heerden & Bharuthram, 2021; Vanderhoven et al., 2015).

Given the challenges of peer assessment, it is important to implement peer assessment activities in a way that reduces the negative effects of interpersonal processes (Panadero et al., 2023). However, to be able to do this, it is necessary (1) to understand what the effects of these interpersonal processes are and (2) to understand how the negative effects of interpersonal processes can be minimized through peer assessment design. Previous literature reviews on the interpersonal processes of peer assessment provide a partial answer to these two questions.

1.2 Previous literature reviews: What do we already know?

Over the last two decades, the impact of the interpersonal processes of peer assessment has attracted a great deal of attention. Research on the subject has been synthesized in no less than seven different literature reviews (Panadero, 2016; Panadero et al., 2018; Panadero & Alqassab, 2019, Panadero et al., 2023; Strijbos et al., 2009; Topping, 2017; van Gennip et al., 2009). These literature reviews aim to provide us with a better understanding of the role that interpersonal processes play in peer assessment, and the impact they have on its outcomes. Most of them only focus on the "agents" element of the MISCA model from Panadero and Lipnevich (2022) without considering the other elements. In the first systematic literature review on the subject, van Gennip and her colleagues (2009) aimed to understand the extent to which the outcomes of peer assessment are related to interpersonal processes. They only found four studies measuring interpersonal processes, none of which linked interpersonal processes to the learning outcomes of peer assessment. The review concluded that more research on the subject was needed. A similar conclusion was drawn in a thematic review published the same year, which deplored the lack of attention to interpersonal processes in peer assessment practices (Strijbos et al., 2009).

Following these calls, more researchers have investigated the interpersonal process of peer assessment, which has allowed subsequent reviews to be based on a more substantial body of evidence and to draw some conclusions. The systematic review conducted by Panadero (2016) revealed that the interaction between peer assessment and interpersonal processes should not be underestimated, and that the use of a formative approach to peer

assessment could mitigate the negative effects of interpersonal processes. The same conclusions were highlighted in Panadero et al.'s (2018) thematic review of the main dissertation in the field. In a thematic review aimed at teachers, Topping (2017) warns that interpersonal processes of peer assessment can hinder its reliability and validity, and urges teachers to consider interpersonal processes. In addition, a review on the effect of anonymity in peer assessment (Panadero & Alqassab, 2019) showed the mixed impact of anonymity on interpersonal processes. Although empirical evidence is limited, this review suggests that anonymity has beneficial effects on some aspects (e.g., reducing peer pressure), but can have detrimental effects on others (e.g., reducing perceptions of fairness).

More recently, Panadero et al. (2023) conducted an update of the review published in 2016 (Panadero, 2016) aimed at synthesizing the results of research on the relationships between intra- and inter-personal factors of peer assessment. Even though the number of studies investigating these factors has increased, a major conclusion of this systematic review is that some of the limitations of these studies make it difficult to synthesize the results. These limitations were (1) a poor reporting of peer assessment intervention characteristics, (2) a lack of clear and agreed-upon definitions of the factors investigated, and (3) a lack of strong research design and a lack of replication studies. They proposed different ways to overcome these limitations in future research. First, to improve the reporting of peer assessment intervention, they developed an instrument that researchers can use to systematically describe the design elements of the peer assessment activities implemented in studies. Second, to address the under-conceptualization of the field, they proposed definitions of the main intra- and inter-personal factors of peer assessment, and presented the idea of bi-directionality of the relationship between peer assessment and intra- and interpersonal processes. This idea emphasizes the fact that intra- and inter-personal factors of peer assessment can both affect and be affected by peer assessment, and that it is important to take into consideration the directionality of the effects. Third, they urged researchers in the field to conduct studies with stronger research designs and to conduct more replication studies.

While all the published reviews highlight the limitations of existing studies on the subject, they also allow us to better understand the impact that interpersonal processes can have on peer assessment processes and outcomes. For example, studies included in these reviews found that the social connections between students influence the scores they provide to their peers (e.g., Domínguez et al., 2016), that low perception of fairness is associated with less engagement in peer assessment (e.g., Zou et al., 2018), or that high levels of trust in the other as an assessor is associated with more effective approaches to learning during peer assessment (K. Cheng & Tsai, 2012). These results suggest that interpersonal processes of peer assessment are one of the factors that explain the variability in the efficacy of peer assessment in fostering learning, and that they have to be considered when designing peer assessment activities.

1.3 What is missing? Purpose of the present review study

The existing reviews have emphasised the importance of considering the interpersonal processes of peer assessment, but give little indication of how such assessment should be designed to take into account these interpersonal processes. The MISCA model (Panadero & Lipnevich, 2022) emphasises the importance of considering the interconnections between the different elements of feedback. To address the question of how to consider interpersonal processes such as to ensure they do not harm peer assessment outcomes, the interconnection between the "agents" element and the "context" element seems particularly important. The only review that incorporates the context element of feedback is one on the impact of anonymity (a contextual element) conducted by Panadero and Alqassab (2019), which revealed the mixed impact of anonymity. However, even though anonymity is an important aspect of the context of peer assessment, it is far from being the only one.

Indeed, when implementing peer assessment, the teaching team has a lot of decisions to make, such as matching the assessor and the assessee (e.g., randomly, by competence level…), the intended output of peer assessment (e.g., comments, scores…), and the guidance that will be offered to students (e.g., rubrics, prompts…). Alqassab et al. (2023, p. 18) label these decisions as relating to instructional design (i.e. "instructional design refers to what designers can manipulate"). These authors developed a typology including twenty-two elements (e.g., matching, training, directionality, frequency, privacy…). The present review is the first to examine the literature using this typology of instructional design elements as we strive to understand how peer assessment can be designed to minimize the negative effect of interpersonal processes.

In the present article, we set out to review studies that have investigated how the instructional design elements of peer assessment influence interpersonal processes. Echoing Panadero et al.'s (2023) idea of bi-directionality, we are interested here in investigating how peer assessment influences interpersonal processes, not how interpersonal processes influence peer assessment. Indeed, our aim is to understand how peer assessment can be designed to minimise the negative effects of interpersonal processes. The other direction (i.e. the influence of interpersonal processes on peer assessment) has already been synthesised in several literature reviews (e.g., Panadero, 2016; van Gennip et al., 2009). Therefore, this study will review studies that have investigated the impact of peer assessment design elements (dependent variable) on interpersonal processes of peer assessment (independent variable).

RQ1. What are the characteristics of the studies investigating the impact of instructional elements on the interpersonal processes of peer assessment?

RQ2. Which instructional design elements have been manipulated and which interpersonal processes have been measured in these studies?

RQ3. What are the impacts of the manipulated instructional design elements on interpersonal processes?

This review will be of interest to both researchers and practitioners. For researchers, the results will give an overview of which instructional design elements have been manipulated to influence which interpersonal processes, and which relationships are still to be investigated. For practitioners, the results can provide guidance on how to design peer assessment activities.

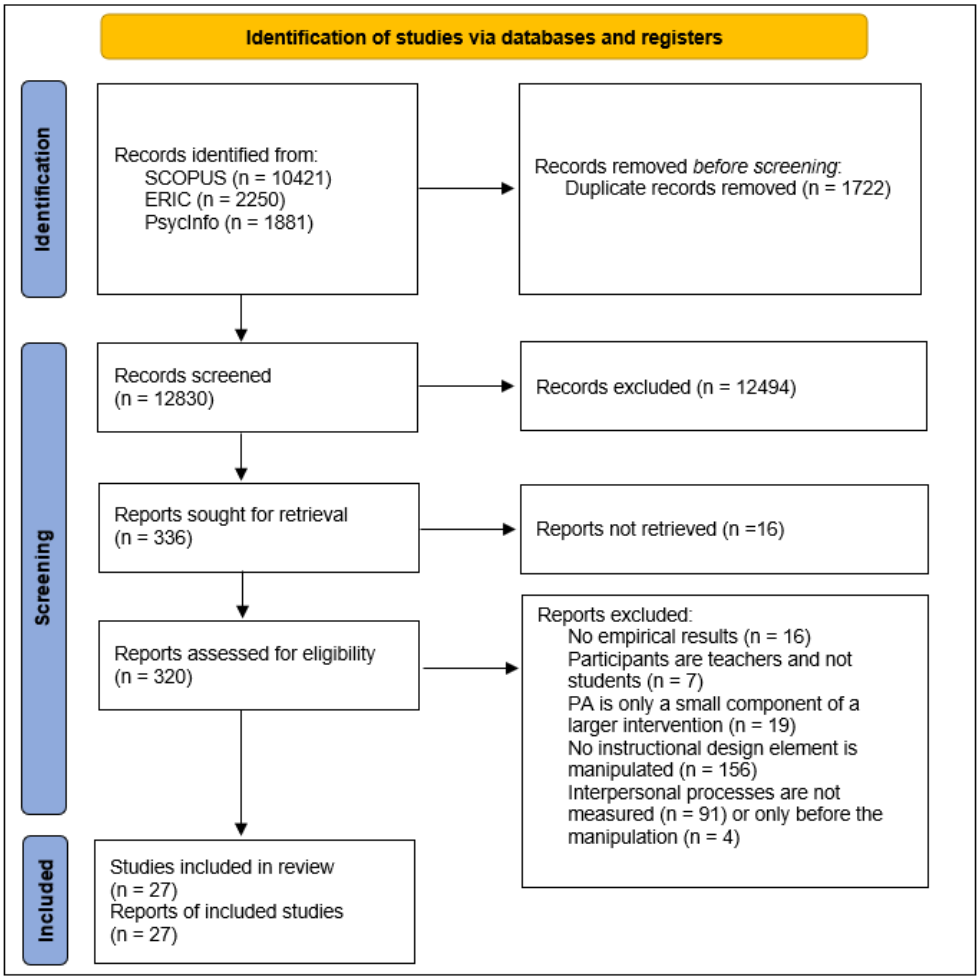
2. Methodology

2.1 Search strategy

As it is recommended that at least two databases be used (Bramer et al., 2017; Ewald et al., 2022), the literature search was carried out using three complementary databases: Scopus (a general database), PsycINFO (a database specialised in the field of psychology), and ERIC (a database specialised in the field of education). The search strings consisted of three connected parts. The first was composed of peer assessment and its most common synonyms ("peer assessment", "peer evaluation", "peer feedback", "peer grading", "peer marking" and "peer review"). These search terms were included with and without hyphenation. The second part was composed of keywords linked to educational contexts (e.g., "student", "school", "college", "elementary education", etc.). We set out to include studies conducted in primary, secondary, and higher education. The keywords related to the first two parts were searched for in the titles, abstracts, and keywords of the articles. For the third part, related to interpersonal processes, a broader approach was adopted. To capture all relevant articles, a wide selection of search terms was used (e.g., "interpersonal", "social", "relation", "psychological safety", "trust", "fear of disapproval", "friendship", etc.), and these terms were searched for in the full text. When available (i.e., for PsycINFO and Eric), the thesaurus was also used for each part of the search string. The complete search string is available in the Appendix 1.

A first search was conducted in July 2020, and an update was conducted in September 2023. The search was limited to peer-reviewed journal articles published in English. No time restrictions were placed; the results thus included all studies from these three databases until September 2023. In total, 14552 records were identified and exported to a reference management software that identified 1722 duplicates. These duplicates were manually removed, which left 12830 unique records (see Figure 1).

In addition, a snowballing method was used to search for additional publications. The references cited in the included articles were checked, as well as the other literature reviews on the subject (Panadero, 2016; Panadero et al., 2018; Panadero & Alqassab, 2019; Panadero et al., 2023; Strijbos et al., 2009; Topping, 2017; van Gennip et al., 2009). No additional records were identified through this snowballing procedure.



Note. From: Page et al. (2021)

Figure 1. Flow diagram of selection of relevant publications for the review.

In addition, a snowballing method was used to search for additional publications. The references cited in the included articles were checked, as well as the other literature reviews on the subject (Panadero, 2016; Panadero et al., 2018; Panadero & Alqassab, 2019; Panadero et al., 2023; Strijbos et al., 2009; Topping, 2017; van Gennip et al., 2009). No additional records were identified through this snowballing procedure.

2.2 Selection process

Titles and abstracts were scanned using the following criteria: (1) The study was published in English in a scientific peer-reviewed journal, (2) the study involved empirical research, (3) the study was conducted in an educational context and involved students (4) the study investigated the interpersonal processes of peer assessment, (5) the peer assessment

was performed on a learning task and with students from the same educational level and (6) the study manipulated at least one instructional design element.

The first criterion aimed to exclude studies that were of insufficient scientific quality to be accepted through a peer-review process. With regard to the second criterion, reviews, opinion pieces, and studies presenting peer assessment tools were excluded. Quantitative, qualitative, and mixed-methods studies were included. The third criterion restricted the selection to studies conducted in an educational context (from primary school to higher education) and excluded studies done in a professional context (e.g., studies on peer reviews between researchers). The fourth criterion allowed us to only select studies that investigated at least one interpersonal process of peer assessment. As explained in the introduction, we adopted a broad definition of interpersonal processes of peer assessment and included all factors that have been considered as "social" or "interpersonal" by the researchers investigating them. The fifth criterion was used to remain in line with Topping's (2017) definition of peer assessment. It excluded studies in which the peer assessment was conducted between students from different levels (e.g., master's students assessing bachelor's students' work), and studies in which peer assessment was not used as a learning tool (e.g., peer assessment only used to weight teacher's scores of group work). Finally, in line with our research question, studies were included if they manipulated at least one of the twenty-two elements of instructional design described in the typology of Alqassab et al. (2023). When the title and abstract did not provide sufficient information to determine if the study fully satisfied the inclusion criteria, the publication remained in the selection. If it became clear in the next step that a publication did not match the inclusion criteria, it was removed.

The screening of titles and abstracts was conducted by the first author. To verify its reliability, a second coder screened 10% of the references (1300 references). As most references did not meet the inclusion criteria, resulting in unbalanced marginals, Gwet's AC_1 is the most reliable indicator (N. Black et al., 2016). According to the cut-offs discerned by Landis & Koch (1977), the Gwet's AC_1 indicated an almost perfect agreement [Gwet's AC_1 = 0.98 (95% CI, 0.969–0.985), $p < .001$] (Gwet, 2014).

Based on the title and abstract scan, 12494 records were excluded. Full-text versions of the remaining publications were obtained, with the exception of sixteen publications that could not be retrieved, even after a personal request to the authors. The full texts of the remaining 320 references were screened, and 293 publications were excluded based on the eligibility assessment (the reasons for exclusion are detailed in Figure 1). This resulted in 27 articles that were included in the review and subsequently analysed.

2.3 Data extraction and data analysis

As a variety of types of studies are included in the review, they were analysed through a narrative synthesis, following the guidelines of Petticrew and Roberts (2006). This narrative analysis consisted of three steps. First, the relevant information for each article was recorded using a data extraction form. This data form was inspired by other literature reviews (Heitink et al., 2016; Lindvall & Ryve, 2019) and consisted of the following sections:

- General information (authors' names, publication date, publication title)
- Context (country, level of education, domain of education)
- Study aim (research question, interpersonal processes investigated as named in the study, instructional design elements manipulated as categorized by Alqassab et al., 2023)
- Research design (study type, study design, instruments)
- Sample (sample type, sample size, sample characteristics, sampling method)
- Results and conclusion

The characteristics of the study were descriptively analysed to answer our first research question (i.e., What are the characteristics of the studies investigating the impact of instructional elements on the interpersonal processes of peer assessment?). In the second step, each study was carefully read, and its results were explored in depth. In the third step, the results were analysed across the individual studies. The studies were organised based on (1) the instructional design elements they manipulated and (2) the interpersonal processes they measured (which allowed us to answer RQ2: Which instructional design elements have been manipulated and which interpersonal processes have been measured in these studies?). For each set of studies, the results were examined to judge whether they were consistent or contradictory across the studies. This allowed us to answer the third research question (What are the impacts of manipulated instructional design elements on interpersonal processes?) through a cross-study synthesis.

During the third step, the quality of the studies was also considered. The overall quality of the included studies was assessed, using three checklists proposed by Efron & Ravid (2019), one for quantitative studies, one for qualitative studies, and one for mixed-method studies. A second coder coded a subset of the quantitative, qualitative, and mixed-method studies. Gwet's AC_2 (using linear weights) was run to determine the agreement between the two raters on the quality appraisal with regard to 9 articles. The agree.coeff2.r script (Gwet, 2014) on R was used. There was a moderate to substantial agreement between the two raters, $AC_2 = .573$ (95% CI, .481 to .666), $p < .001$.

The aim of the quality appraisal is not to exclude any study, but to identify the weaknesses and strengths of each study to (1) consider them when synthesising the studies' results and (2) have an idea of the quality of the studies in the field to identify the limitations and perspectives. Indeed, even studies with limitations can provide valuable information Indeed, even studies with important limitations can provide valuable information (Booth et al., 2016; Walsh & Downe, 2005), such as examples of innovative ways to design peer assessment activities, but it is important to consider the studies' quality. The critical appraisal of all 27 quantitative, qualitative, and mixed-method studies can be found in Appendices D, E, and F, respectively.

3. Results

3.1 RQ1: Description of included studies

The characteristics of the 27 included studies are quite homogeneous, as can be seen in Table 1. First, the studies have principally been conducted in three regions of the world: East and South Asia (China, Taiwan, Japan, Vietnam, and the Philippines), Europe (Belgium, Turkey, and Spain), and the United States of America. Second, most studies have been conducted in higher education, with only two studies conducted in secondary education, and one in primary education. Third, the majority of studies have been conducted in the context of education or English as a foreign language. Fourth, all the studies were published in the last fifteen years.

Concerning study design, most studies used a mixed method (14 studies), followed by a quantitative method (9 studies), and a qualitative method (4 studies). There is an overrepresentation of correlational studies (16 studies), with only eight quasi-experimental, and three experimental studies. The sample size ranges from 12 to 58 students for qualitative studies and from 16 to 209 students for mixed-method or quantitative studies.

3.2 RQ2: Investigated instructional design elements and interpersonal processes

To present their results, we organised studies based on the instructional design elements they manipulated (see Tables 2 and 3). The most commonly-investigated design element was privacy (15 studies), followed by the peer assessment format (6 studies) and the presence of training (4 studies). The matching of assessors and assessees, the product of assessment, and the unit of assessment (assessors) are each investigated in one study. The interpersonal processes of peer assessment investigated in the studies are, in order of frequency, *(dis)comfort* (11 studies), *trust* (8 studies), *emotions due to interpersonal processes* (7 studies), *perceived fairness* (5 studies), *peer pressure* (4 studies), *psychological safety* (3 studies), *friendship marking* (2 studies), *perceptions towards assessors* (2 studies), *fear of disapproval* (2 studies), *face-saving* (2 studies), and *value congruency* (1 study). Most studies investigated at least two interpersonal processes.

Table 1. Main characteristics of the included studies

Authors and date	Context			Research design		Sample size
	Country	Education setting	Subject	Study type	Study design	
Akçay et al. (2021)	Turkey	Higher education	Education	Qualitative	Correlational	N = 58
Cheng et al. (2023)	China	Higher education	Education	Mixed method	Experimental	N = 34
Ching & Hsu (2016)	USA	Higher education	Education	Mixed method	Correlational	N = 16
Güler (2017)	Turkey	Higher education	Education	Mixed method	Experimental	N = 84
Kaya (2021)	Turkey	Higher education	EFL	Qualitative	Correlational	N = 12
Kim (2019)	Japan	Higher education	EFL	Mixed method	Correlational	N = 64
Kim & Lan (2021)	China	Higher education	EFL	Mixed method	Correlational	N = 57
Li (2017)	USA	Higher education	Education	Quantitative	Quasi-experimental	N = 77
Lim et al. (2021)	USA	Higher education	Education	Mixed method	Correlational	N = 81
Lin (2018)	Taiwan	Higher education	Education	Quantitative	Experimental	N = 32
Ocampo et al. (2023)	Philippines	Higher education	Psychology	Quantitative	Quasi-experimental	N = 145
Panadero et al. (2013)	Spain	Higher education	Education	Quantitative	Quasi-experimental	N = 209
Parikh & Sheehan (2016)	USA	Higher education	Pharmacy	Quantitative	Correlational	N = 144
Praver et al. (2011)	Japan	Higher education	EFL	Quantitative	Correlational	N = 86

Raes et al. (2015)	Belgium	Higher education	Education	Mixed method	Quasi-experimental	N = 51
Rotsaert et al. (2018)	Belgium	Higher education	Education	Mixed method	Correlational	N = 46
Senden et al. (2023)	Belgium	Higher education	Education	Quantitative	Quasi-experimental	N = 89
Su (2023)	China	Higher education	EFL	Mixed method	Correlational	N = 40
Thanh & Gillies (2010)	Vietnam	Higher education	/	Mixed method	Correlational	N = 145
Tzeng et al. (2021)	USA	Higher education	Medicine	Mixed method	Correlational	N = 32
van Heerden & Bharuthram (2021)	South Africa	Higher education	Academic writing	Mixed method	Correlational	N = 64
Vanderhoven et al. (2015)	Belgium	Secondary education	Oral skills	Mixed method	Quasi-experimental	N = 69
Xue et al. (2023)	China	Higher education	Academic writing	Qualitative	Correlational	N = 12
Yu (2012)	Taiwan	Secondary School	Science	Quantitative	Quasi-experimental	N = 101
Yu & Liu (2009)	Taiwan	Higher education	Engineering	Mixed method	Correlational	N = 40
Yu & Wu (2011)	Taiwan	Primary education	Science	Quantitative	Quasi-experimental	N = 243
Zhang et al. (2022)	China	Higher education	Academic writing	Qualitative	Correlational	N = 12

3.3 RQ3: Impact of privacy on interpersonal processes

Peer assessment can be organised in terms of various levels of privacy (Alqassab et al., 2023): it can be entirely public (i.e., all names are disclosed to all participants), confidential (i.e., names of the assessee and the assessor are only disclosed to peers with whom they interact), single-blind (anonymity of the assessor or the assessee), or double-blind (anonymity of both assessor and assessee). Five studies have investigated the impact of double-blind anonymity (Kaya, 2021; Su, 2023; F. Yu, 2012; F. Yu & Liu, 2009; F. Yu & Wu, 2011), while seven studies have investigated the impact of single-blind anonymity, with only the identity of the assessor being concealed (Güler, 2017; L. Li, 2017; Lim et al., 2021; Lin, 2018; Raes et al., 2015; Rotsaert et al., 2018; Vanderhoven et al., 2015). Finally, three studies have combined the effects of anonymity and face-to-face contact by comparing online anonymous peer feedback with face-to-face non-anonymous peer feedback (Akçay et al., 2021; S. Kim, 2019; S. Kim & Lan, 2021). The results of these studies are summarized in Table 2.

Double-blind anonymity

In the three studies conducted by Yu and colleagues (F. Yu & Liu, 2009; F. Yu & Wu, 2011; F. Yu, 2012), the authors obtained mixed results on the impact of disclosing students' names or anonymity (either with a nickname or no name). In F. Yu and Liu's (2009) study on emotions relating to the interpersonal processes, higher education students could choose the identity revelation mode they wanted to use as assessor and as assessee. They expressed a strong preference for the anonymous mode, both as an assessor and as an assessee, because, as they reported, having their identity concealed helps them avoid feelings of unease, tensions, shameful feelings, and humiliating situations. The two other studies both investigated the impact of the different identity revelation modes on students' perceptions toward their assessors. In F. Yu & Wu's (2011) study, eight classes of primary school pupils were assigned to one of four conditions: real names, no name, nicknames, or self-choice (i.e., students could choose any of the previous modes). The perceptions toward their assessors were more positive in the real names and self-choice conditions which suggest that anonymity, at least when imposed, impairs assessees' perceptions of the assessors. However, these results were not replicated in F. Yu's (2012) study, conducted with secondary school pupils: no significant differences were found between the conditions regarding assessees' perceptions of the assessors.

The impact of double-blind anonymity on interpersonal processes was also investigated in two other studies: Kaya's (2021) study showed a mixed impact of privacy on emotions due to the interpersonal processes, while Su's (2023) study revealed a positive impact on comfort. In both studies, higher education students experienced non-anonymous peer feedback, followed by anonymous peer feedback to compare their perceptions. In Kaya's (2021) study, students felt tense when they were identifiable, while anonymity alleviated this tension by making them feel free to provide objective feedback. However, anonymous peer feedback also conveyed negative emotions (anger, upset, and disappointment) because, in

the anonymous condition, students attributed the poor performance of their peers to a lack of effort, which was not the case in the non-anonymous condition. In a study by Su (2023), students felt significantly more comfortable when providing and receiving anonymous peer feedback than was the case with non-anonymous peer feedback. They reported feeling more at ease and relaxed under conditions of anonymity.

Single-blind anonymity

The results from the three studies that examined the effects of the provision of anonymous feedback on peer pressure are also mixed. Vanderhoven et al. (2015) compared the perceptions of students who anonymously or no-anonymously provided feedback to their peers in class. Students felt significantly less peer pressure when they provided anonymous feedback (through a classroom-response technology) than when they provided non-anonymous feedback (through card raising). However, this impact of anonymity on peer pressure was not found in L. Li's (2017) and Raes et al.'s (2015) studies. In L. Li's (2017) studies, students were also assigned to provide anonymous or non-anonymous peer feedback, but the perceptions of peer pressure did not differ significantly between the groups. In Raes et al.'s (2015) study, students gave scores to their peers using classroom-response technology, participated in an oral feedback exchange moderated by the teacher, and wrote comments they uploaded online. In the first group, the uploaded feedback was anonymous, whereas it was non-anonymous in the second group. There was no significant difference in students' perceptions of peer pressure in the two groups, even though their answers to open questions suggested that an increase in anonymity is associated with a decrease in peer pressure.

Single-blind anonymity, like double-blind anonymity, seems to positively affect students' perceptions of comfort. Indeed, two studies (Lim et al., 2021; Raes et al., 2015) found a positive effect, while the third did not find any effect (Vanderhoven et al., 2015). In Lim et al.'s (2021) study, the assessee was asked to leave the room while their peers provided feedback on their performance (feedback that will be communicated afterwards by the teacher). Students reported that, compared to previous experiences of peer assessment, they found this type of peer assessment more comfortable because, as an assessor, they were able to express themselves without being afraid or hurting someone's feelings, and, as an assessee, they know the feedback they receive is authentic. In Raes et al.'s (2015) study, anonymity had a significant effect on comfort: students felt more comfortable when they provided anonymous feedback through classroom-response technology than when they provided non-anonymous oral or written feedback. In Vanderhoven et al.'s (2015) study, students' perceptions of discomfort were not significantly different whether they provided anonymous or non-anonymous peer feedback.

Güler (2017), Lin (2018), and Vanderhoven et al. (2015) all compared the impact of the anonymity of the assessor on students' perceptions of fairness, but found different results. There was no significant difference between students in the anonymous and non-anonymous conditions in the studies of Güler (2017) and Vanderhoven et al. (2015), while students'

perceptions of fairness were significantly higher for students in the non-anonymous condition in the study of Lin (2018).

Finally, Rotsaert et al. (2018) manipulated privacy in yet another way and found mixed results. They investigated the effect of a transitional approach to anonymity. Each week, students provided written feedback on their peers' presentations through mobile-response technology followed by an oral discussion moderated by the teacher. In the first two weeks, the written feedback was anonymous, while in the final two weeks, it was non-anonymous. Students' perceptions of friendship marking, trust, psychological safety, fear of disapproval, and value congruency were already quite positive prior to the intervention. Rotsaert et al.'s (2018) findings indicated a decrease in fear of disapproval, which was also found in Vanderhoven et al.'s (2015) study. In addition, Rotsaert and colleagues (2018) found an increase in friendship marking. None of the other students' perceptions evolved with the transition to anonymity, which suggests that, according to the authors, the transitory approach revealing students' identities had no deleterious effects.

Anonymity and contact

The last three studies found different effects depending on the interpersonal process investigated (comfort, trust, face-saving, peer pressure, or perceived fairness) when comparing online anonymous peer feedback with face-to-face non-anonymous peer feedback (Akçay et al., 2021; S. Kim, 2019; S. Kim & Lan, 2021). In the study of S. Kim (2019) and S. Kim and Lan (2021), one group of randomly assigned students experienced first non-anonymous face-to-face peer assessment and then anonymous peer assessment with no interaction, while another group experienced the activities in reverse order. In the study by S. Kim (2019), conducted on Japanese students, assessors felt more comfortable providing negative feedback in the anonymous condition. However, they also felt comfortable providing negative feedback face-to-face because, through the interactions, hedging was easier, and they could work collaboratively with their peers to make them realise their mistakes. As assessees, they trusted their peers more in the non-anonymous condition, because the interactions helped the assessors to formulate accurate feedback. S. Kim and Lan (2021) conducted their study with Chinese students. They hypothesised that students' concerns about face-saving would hamper their abilities to provide constructive feedback when non-anonymous. The results confirm that face-saving matters to students, yet, students preferred face-to-face feedback: they were keen to provide high-quality feedback, to save face, and they found it was easier to do so when they could interact with their peers. In Açkay et al.'s (2021) study, all the students first experienced face-to-face peer feedback, and then single-blind peer feedback concealing the identity of the assessor. Some students felt pressure to provide positive feedback and higher grades when it was non-anonymous. Anonymity was seen as a solution to this: the students felt that they could provide the deserved scores and comments when anonymous. However, other students viewed anonymous peer feedback as being less fair, precisely because they thought that the pressure felt during face-to-face feedback positively contributed to the objectivity of the peer assessment, even though the authors did not clarify why this might be.

Table 2. Privacy and interpersonal processes investigated in the included studies

Privacy	Impact	Emotions due to the interpersonal processes	Perceptions toward assessors	Comfort	Peer pressure	Perceived fairness	Fear of disapproval	Friendship marking	Trust	Psychological safety	Value congruency	Face-saving
Double-blind anonymity	Positive	Kaya (2021) ; Yu & Liu (2009)		Su (2023)								
	Negative	Kaya (2021)	**Yu & Wu (2011)**									
	Absent		**Yu (2012)**									
Single-blind anonymity	Positive			Lim et al. (2021); **Raes et al. (2015)**	**Vanderhoven et al. (2015)**		Rotsaert et al. (2018); **Vanderhoven et al. (2015)**					
	Negative					**Lin (2018)**		Rotsaert et al. (2018)				
	Absent			**Vanderhoven et al. (2015)**	**Li (2017)**; **Raes et al. (2015)**	**Güler (2017)**: **Vanderhoven et al. (2015)**			Rotsaert et al. (2018)	Rotsaert et al. (2018)	Rotsaert et al. (2018)	
Anonymity and absence of contact	Positive			Kim (2019)	Akçay et al. (2021)							
	Negative					Akçay et al. (2021)			Kim (2019)			Kim & Lan (2021)

Note. Studies with an experimental or quasi-experimental design are in bold. Studies are highlighted in green if the researchers found that the design element had a desirable effect on interpersonal processes and in red if they found an undesirable effect.

3.4 RQ3: Impact of peer assessment format on interpersonal processes

The format of peer assessment was manipulated in six studies. Format refers to the degree of guidance of the peer assessment activity, i.e., did the students receive a rubric, some prompts, etc. to guide them? (Alqassab et al., 2023).

The studies of Xue et al. (2023) and F. Zhang et al. (2022) found a positive impact of guidance on face-saving and a mixed impact on emotions due to the interpersonal processes. These two publications relate to the same data collection, which aimed to compare the impact of three assessment tools offering different types of guidance (Moodle, rain Classroom, and WeChat) on face-saving and emotions due to interpersonal processes. Moodle is a learning management system that pushes students to follow the instructions thoughtfully when providing feedback, Rain Classroom is a tool created to project presenters' slides on the audience's mobile application while allowing them to provide quick real-time feedback, and WeChat is a social media platform that attends to the participants' affective factors. All students had the opportunity to use the three tools on multiple occasions to provide peer feedback. Students evoked concerns about face-saving (F. Zhang et al., 2022) and communication anxiety (Xue et al., 2023). They thought the most appropriate tool to deal with these issues was WeChat because it allowed them to insert emojis and memes in their feedback to soften harsh feedback, which the teaching teams encouraged students to do. However, according to the students involved, WeChat also created an emotional burden because students needed to be connected to their peers' social media accounts to be able to use it.

Two studies (Ching & Hsu, 2016; Tzeng et al., 2021) found a positive impact of a peer assessment format with more guidance on students' perceptions of comfort, while a third study did not find any effect (Panadero et al., 2013). It is worth noting that in the two studies identifying a positive impact, there was no real control group. Tzeng et al. (2021) introduced a method called "Observation-Reaction-Feedback" that aimed to guide students when they provided feedback on their peers' clinical skills. Students could choose to use the method or not. Students who chose to use it were those least comfortable with peer feedback at the pre-test. At the post-test, their level of comfort caught up with the levels of other students, and, according to their answers to open questions, this was thanks to the method. In Ching and Hsu's (2016) study, students played a self-select stakeholder role while providing feedback to their peers (i.e., they provided feedback from a particular perspective). Students reported feeling comfortable while participating in the peer feedback activity. In their quasi-experimental study, Panadero et al. (2013) compared students' perceptions of comfort when using a rubric or not. In two classrooms students received a rubric to provide feedback to their peers, while in the two other classrooms, they did not. There was no difference in students' perceptions of comfort between the two groups.

In their study on the use of rubrics, Panadero et al. (2013) also compared students' perceptions of fairness and their tendency to overmark their friends. Even though peer ratings

were globally more valid when students used a rubric, friendship marking was only present in these classrooms: when students did not use a rubric, they tended to overmark all their peers, not just their friends. There was no impact in terms of the use of the rubric on students' perceptions of fairness.

Finally, the presence of increased amounts of guidance seem to positively impact students' perceptions of trust and psychological safety. In L. Cheng et al.'s (2023) experimental study, students were randomly assigned to a group that was guided by a general script during the peer feedback, or to a group that received a more detailed regulation script for dialogic feedback with prompts. Students' attitudes toward the climate of trust were significantly more positive in the group with the more detailed regulation script for dialogic peer assessment. In the interviews, students explained that the prompts allowed them to ensure they shared a common understanding of the assessment criteria, which raised their feelings of trust. This positive impact on trust was also present in Ching and Hsu's (2016) study, alongside a positive impact on psychological safety. Students reported feeling psychologically safe and confident in themselves and others when participating in a peer feedback activity where they had to provide feedback from the perspective of a chosen stakeholder.

3.5 RQ3: Impact of training on interpersonal processes

The impact of training on the interpersonal processes of peer assessment was investigated in four studies, two of them investigating the impact of a general training session (Ocampo et al., 2023; Parikh & Sheehan, 2016), while the other two investigated the impact of a training session specifically designed to target interpersonal processes (L. Li, 2017; Senden et al, 2023).

The studies investigating the impact of a general training session found it had a positive impact on comfort (Ocampo et al., 2023) and on trust (Ocampo et al., 2023; Parikh and Sheehan, 2016). In Parikh and Sheehan's (2016) study, students received a 30-minute instructional lecture explaining how to provide feedback to their peers. Most students reported trusting their abilities and that of their peers to conduct peer assessments after receiving the training. However, there was no control group or use of a pre-test to compare the results. Ocampo et al. (2023) also investigated the effect of general training on students' perceptions, but used a quasi-experimental design and a longer training session. Students in the control group received a 30-minute training on how to use the peer feedback tool, while students in the experimental group received a 180-minute training that also allowed them to practice peer feedback. All students assessed the same three writing samples created by the research team, even though they were led to believe they were real samples from their peers. The researchers manipulated the gender that was indicated on the samples. Students' perceptions were significantly different between trained and untrained assessors, but only for male assessors providing feedback to women: trained men felt more comfortable and confident than untrained men when providing feedback to women.

The studies investigating the impact of a specific training session found a positive effect on peer pressure (L. Li, 2017), but not on trust and psychological safety (Senden et al., 2023). In L. Li's (2017) study, already mentioned in the section on privacy, students in the training group received a training session specifically conceived to address the peer pressure students can feel when their identities are revealed. Students who received the training felt significantly less peer pressure than students who did not receive it, regardless of whether their identities were revealed or not. Training seems therefore more effective than privacy when it comes to addressing feelings of peer pressure. Senden et al. (2023) also created a training session focusing on the interpersonal processes of peer assessment, but targeting trust and psychological safety. Students received either a 1-hour general training or a 2-hour training with a part designed to address their concerns regarding trust and psychological safety. Students' perceptions of psychological safety, trust in the self, and trust in the other were already quite high at the pre-test and remained so in both groups.

3.6 RQ3: Impact of other instructional design elements on interpersonal processes

The three remaining studies each manipulated a different instructional design element: the unit of assessment (Thanh & Gillies, 2010), the output of assessment (Praver et al., 2011), and the matching of assessor and assessees (van Heerden & Bharuthram, 2021). These three studies were correlational in nature.

Thanh and Gillies (2010) found a positive impact when it came to conducting peer assessment in a group on students' emotions due to the interpersonal processes and feelings of comfort. Students assessed the contributions of their group's members in two ways. First, they assessed them individually, and second, they assessed them collectively, through an open discussion involving all group members. They also collectively assessed other groups' performances. Students' feelings of discomfort and worry were significantly lower when doing the collective assessment than when doing the individual one, which was also confirmed by the qualitative results.

In their study, Praver et al. (2011) found that having to write comments in addition to providing scores led to negative emotions on the part of students. The students had to facilitate a reading circle and were assessed by their peers, either with scores only (in the first three rounds) or with scores and written comments (in the last three rounds). At the end of the course, students had to indicate which peer assessment output they found the more useful, and which caused more nervousness and embarrassment. The vast majority of students found it more useful to provide and receive a combination of scores and written comments. However, most students also indicated that providing and receiving comments in addition to scores made them more nervous, and that providing comments in addition to scores caused more embarrassment, while there was no difference in embarrassment for the receiving part.

Finally, van Heerden and Bharuthram (2021) found a mixed impact in terms of familiarity on interpersonal processes. In their study, after exchanging feedback with a peer, students indicated how well they knew this peer and how they felt during the peer feedback activity.

Based on descriptive analyses, it seems that students who knew their peers well found the experience more comfortable, enjoyable, and less challenging than those who did not have such a close relationship. The students explained that knowing their peers made it easier to take feedback as constructive criticism instead of viewing it as being harsh and critical and that it was associated with an implied level of honesty and trust. However, some students thought that not knowing their peers was beneficial because it made the process more objective.

4. Discussion

Previous literature reviews have revealed that interpersonal processes can influence peer assessment outcomes and processes (e.g., Panadero et al., 2023; Topping, 2017). Therefore, when designing peer assessment activities, the impact of design elements on interpersonal processes should be considered. Otherwise, these activities may not achieve their goal of promoting student learning. This systematic literature review innovates by linking the "agents" and "context" elements of the MISCA model (Panadero & Lipnevich, 2022) to explore how peer assessment can be designed to avoid the negative impacts of interpersonal processes and maximise its effectiveness.

Twenty-seven studies that investigated the impact of the instructional design elements of peer assessment on interpersonal processes were reviewed. The included studies were limited to peer-reviewed papers. Even though we did not restrict the included papers to a specific year of publication, none of them were published prior to 2009. Moreover, almost half of the papers were published in the last three years, which makes this literature review particularly timely. It seems that calls for more attention to the interpersonal processes of peer assessment (Strijbos et al., 2009; van Gennip et al., 2009) have been heard, and that the research on the subject is rapidly developing. If this research attention persists, an update of the literature review presented in this paper will be necessary in a few years.

Our review has uncovered that privacy was the instructional design element whose impact on interpersonal processes has been investigated the most, with more than half of the included studies focusing on it. It seems that anonymity is seen as the easiest solution to minimising the negative impacts of interpersonal processes. There is consistent evidence from several studies that privacy increases feelings of comfort (S. Kim, 2019; Lim et al., 2021; Raes et al., 2015; Su, 2023), decreases fear of disapproval (Rotsaert et al., 2018; Vanderhoven et al., 2015), and alleviates the tensions students can feel (Kaya, 2021; F. Yu & Liu, 2009). It also seems that it decreases peer pressure (Açkay et al., 2021; Vanderhoven et al., 2015), even though some studies did not find any significant effect (L. Li, 2017; Raes et al., 2015). Yet, as already suggested by Panadero and Alqassab (2019), the results confirmed that, even if anonymity can be beneficial, it is not a magical solution to erase student's concerns. On the contrary, it could even have detrimental effects on some processes: evidence suggests that privacy can convey negative emotions (Kaya, 2021) and has a detrimental impact on perceived fairness (Açkay et al., 2021; Lin, 2018), perceptions toward the assessors (F. Yu & Wu, 2011), and value congruency (Rotsaert et al., 2018). However, some studies did not

find that privacy had any significant effect on these processes (Güler, 2017; Vanderhoven et al., 2015; F. Yu, 2012). Finally, some studies aimed to examine the effect of privacy, but this was confounded with the effect of face-to-face contact. The detrimental effect they found in the case of trust (S. Kim, 2019) and face-saving (S. Kim & Lan, 2021) seems to come from the absence of contact, not from anonymity itself.

The second most common instructional design element investigated in the included studies was the format of peer assessment. The five studies focusing on the impact of the format of peer assessment used different interventions, all of which aimed to provide increased amounts of guidance to the students. Not surprisingly, it seems that offering more guidance has a positive impact: it could help students "save face" (Xue et al., 2023; M. Zhang et al., 2022) and increase students' perceptions of comfort (Ching & Hsu, 2016; Tzeng et al., 2021), trust (L. Cheng et al., 2023; Ching & Hsu, 2016), and psychological safety (Ching & Hsu, 2016). However, these results must be interpreted with caution due to the limitations in the studies' research designs, especially as the study with the stronger research design (Panadero et al., 2013) did not find any significant impact in terms of format on comfort and perceived fairness. Panadero et al. (2013) also identified a negative impact of a rubric on friendship marking. However, this negative impact should be put into perspective: it did not imply a less valid peer assessment for friends in the rubric condition, as students using a rubric only overscored their friends, while students without a rubric overscored all their peers.

The impact of the presence of training on interpersonal processes has been investigated in four studies and seems to have only a positive effect. Two studies have found that general training increased students' feelings of comfort (Ocampo et al., 2023; Parikh & Sheehan, 2016), while one study has found that a training session specifically created to decrease students' feelings of peer pressure achieved its target (L. Li, 2017). The absence of a significant effect related to training in the last study (Senden et al., 2023) could be explained by the fact that students had high levels of trust and psychological safety prior to the training. This would be consistent with the results obtained by Ocampo et al. (2023) that training only increased trust in the case of students who started with low levels of trust. In sum, while previous literature reviews have shown that providing training prior to peer assessment has a positive impact on the efficacy of peer assessment (H. Li et al., 2020), students' peer assessment skills (van Zundert et al., 2010), and students' attitudes toward peer assessment (van Zundert et al., 2010), the present review adds that training is also beneficial when it comes to minimising the negative effects of interpersonal processes.

Finally, it is impossible to arrive at any conclusion with regard to the other three design elements investigated in the included studies: the unit of assessment (Thanh & Gillies, 2010), the product of assessment (Praver et al., 2011), and the matching between assessors and assessees (van Heerden & Bharuthram, 2021). Indeed, each of these elements are only investigated in one study, and these studies have weak research designs. More research is needed to understand the impact that these three design elements can have on the interpersonal processes of peer assessment.

Table 3. Instructional design elements and interpersonal processes investigated in the included studies

Element	Impact	Emotions due to the interpersonal processes	Face-saving	Comfort	Perceived fairness	Friendship marking	Psychological safety	Trust	Peer pressure
Format (more guidance)	Positive	Xue et al. (2023); Zhang et al. (2022)	Xue et al. (2023) ; Zhang et al. (2022)	Ching & Hsu (2016); Tzeng et al. (2021)			Ching & Hsu (2016)	**Cheng et al. (2023)**; Ching & Hsu (2016)	
	Negative	Xue et al. (2023); Zhang et al. (2022)				**Panadero et al. (2013)**			
	Absent			**Panadero et al. (2013)**	**Panadero et al. (2013)**				
Presence of a general training	Positive			**Ocampo et al. (2023)**				**Ocampo et al. (2023)**, Parikh & Sheehan (2016)	
Presence of a specific training	Positive								**Li (2017)**
	Absent						**Senden et al. (2023)**	**Senden et al. (2023)**	
Unit of assessment (group)	Positive	Thanh & Gillies (2010)		Thanh & Gillies (2010)					
Product (presence of comments)	Negative	Praver et al. (2011)							
Matching (familiarity)	Positive	van Heerden & Bharuthram (2021)		van Heerden & Bharuthram (2021)				van Heerden & Bharuthram (2021)	
	Negative				van Heerden & Bharuthram (2021)				

Note. Studies with an experimental or quasi-experimental design are in bold. Studies are highlighted in green if the researchers found that the design element had a desirable effect on interpersonal processes and in red if they found an undesirable effect.

4.1 Gaps unveiled by the literature review and perspectives for future research

Our review found that the impact of only seven of the 21 instructional design elements identified by Alqassab et al. (2023) had been investigated. The impact of other elements on interpersonal processes should also be investigated. In the review conducted by Alqassab et al. (2023) it was found that peer assessment in research is not as diverse as it could theoretically be. With important similarities in the design of these activities, it may not be necessary to investigate the impact of each of the 21 instructional design elements. In fact, these authors found that – in the research published up to date – some instructional design elements with regard to peer assessment often occur together, and found five main design patterns of peer assessment. For example, when peer assessment takes place face-to-face, it tends to be synchronous, formative, non-anonymous, bi-directional, and individual, with a matching of assessors and assessees based on self-selection, and with oral and written comments as an output (Alqassab et al., 2023). Rather than examining the impact of each element individually, it would be interesting to examine the impact of these patterns on the interpersonal processes of peer assessment. Moreover, these patterns are determined by 10 main instructional design elements (Alqassab et al., 2023), which could therefore be investigated as a priority. The present review found that the impact of only four of these elements (product, matching, contact, and privacy) on interpersonal processes has been investigated. The other elements are the purpose (formative or summative), the scope of involvement (e.g., giving and/or receiving feedback), the possibility of revision after the peer assessment, the unit of assessment assessor (is the peer assessment conducted individually or in a group?), the unit of assessment assessee (is the peer assessment conducted by an individual peer or by a group?), and the group constellation (is the peer assessment conducted between or within groups?). These elements are likely to influence interpersonal processes. For example, the question of perceived fairness should be more prominent in the case of summative assessment than in that of formative assessment.

The present review is innovative in terms of considering the interconnections between the "agents" and "context" elements of the MISCA model (Panadero & Lipnevich, 2022) when examining the impact of instructional design elements (context) on interpersonal processes (agents). However, instructional design elements are only one aspect of the "context" element in the MISCA model, other important aspects being the instructional context (e.g., does the peer assessment activity take place at the beginning, middle, or end of the course?), the classroom climate, and the general context (e.g., educational level). Investigating the influence of classroom climate might be particularly relevant as it seems to have an important impact on students' perceptions of peer assessment (Rotsaert, 2017). Another perspective is to investigate the three other elements of the MISCA model (Panadero & Lipnevich, 2022). As student characteristics are the central element of the model, it might be relevant to focus first on this element and explore how these characteristics interact with interpersonal processes. Student characteristics include the intra-personal elements of the conceptualisation developed by Panadero et al. (2022), which were not included in the review presented in this paper. As these elements (e.g., self-efficacy) also play an important role in

peer assessment (Panadero et al., 2023), it would be useful to conduct a literature review to investigate how they are affected by the instructional design elements of peer assessment.

Another important student characteristic to investigate is culture. The studies included in this review suggest that student culture is important as some interpersonal processes may be more relevant in an Asian context than in a European one (S. Kim & Lan, 2021; Xue et al., 2023; M. Zhang et al., 2022). These studies were conducted in homogeneous cultural contexts, but their findings raise questions about peer assessment in intercultural contexts such as which interpersonal factors are salient and how peer assessment activities should be designed in these contexts. Developing knowledge on this topic is important because the higher education population is increasingly multicultural. Six per cent of all higher education students in OECD countries are international students, with a much higher proportion in some countries such as 22 per cent in Australia (OECD, 2023). This multicultural context offers opportunities to prepare students to work effectively in an increasingly intercultural workplace, but it is not without challenges (Popov et al., 2012). Research is lagging behind in providing educators with ideas on how to design inclusive and culturally sensitive peer assessment activities.

4.2 Limitations and directions for further research

The first limitation we encountered during the process was linked to the subject of the review itself. As already noted by Panadero et al. (2023), the different interpersonal processes of peer assessment are ill-defined. In the included studies, the processes investigated are rarely defined, and it was hard to tell if they covered the same reality from one study to another. On the one hand, some processes were given the same name but measured with regard to entirely different items (e.g., psychological safety in Ching & Hsu's 2016 study and Rotsaert et al.'s 2018 study). On the other hand, some processes were given different names but appeared to measure close processes by using very similar items (e.g., perception toward the assessor in F. Yu's 2012 study, and trust in peers' evaluative abilities in Rotsaert et al.'s 2018 study). This lack of clear definitions also influences the way these processes are measured. Most studies relied on non-validated scales which affects the results obtained and hampers comparisons between studies. In their literature review on the intra and interpersonal factors of peer assessment, Panadero et al. (2023) proposed definitions of the main processes, which can serve as a basis for developing research on the subject. We argue that a second important step to advance the research is to create and validate scales for each interpersonal process. In the meantime, we recommend researchers provide all the items used to measure interpersonal processes. Indeed, while conducting the review, we noticed that the scales used were not always available, which prevented us from verifying what researchers meant by the different interpersonal processes. Therefore, in this review, we found ourselves constrained to using the terms used by the authors of the studies.

The second limitation is the lack of variety in the contexts in which the studies were conducted. The vast majority of studies were conducted in higher education, most of them in the domain of education or EFL. This focus on higher education was also noted in other reviews on peer assessment (e.g., Alqassab et al., 2023; H. Li et al., 2020). It is problematic

because nothing guarantees that results found with higher education students apply to students in primary and secondary education. For example, the only study conducted in primary education (F. Yu & Wu, 2011) showed that privacy led students to write impulsive comments evidencing irrational emotions (e.g., "POOR! POOR! POOR!", F. Yu & Wu, 2011, p. 2175). We can imagine that privacy will not have the same consequences for more mature students and therefore that privacy will not have the same impact on interpersonal processes either. For the studies conducted in higher education, the vast majority were conducted within the context of "soft disciplines", which present some important differences from "hard disciplines" regarding teaching and learning (Neumann et al., 2002). Hence, more studies conducted in primary and secondary education are needed, alongside more studies in "hard disciplines" of higher education.

Moreover, the included studies were conducted in a limited number of countries. It may be because we restricted our search to papers published in English, which may have led us to miss out on relevant studies published in other languages. Search in other languages should be conducted to verify if studies have been conducted in other parts of the world (e.g., studies conducted in South America and published in Spanish). If the absence of studies is confirmed, it would be necessary to conduct studies in a larger variety of countries and other parts of the world. Indeed, it has been argued that culture plays a role in explaining the relationships between instructional design elements and interpersonal processes (e.g., Fithriani, 2018). For example, students coming from collectivist cultures, which are more common in Asian countries, tend to work toward maintaining group harmony and mutual face-saving, which can affect their willingness to criticize their peers (Fithriani, 2018).

The third limitation is linked to the quality of the included studies. While performing the quality appraisal, we identified important limitations with regard to the included studies, such as a lack of information on the reliability and validity of the instruments used for quantitative and mixed-method studies, or a lack of detail on how the data were triangulated in the case of qualitative studies. There is a dearth of studies using strong research designs, which hampers the research on the subject. For example, concerning the study of the impact of privacy, studies with similar procedures have drawn different conclusions. Rotsaert et al. (2018) have investigated the impact of a transitional approach to anonymity by asking the same students to first conduct an anonymous peer assessment, and then a non-anonymous peer assessment. The same procedure was used by S. Kim (2019) and S. Kim and Lan (2021), but this time to compare the impact of anonymous and non-anonymous peer assessment, without taking into consideration the fact that the experience of the first activity could influence the second one. This illustrates the importance of using control groups composed of different students (Creswell, 2012). The predominance of studies with weak research designs is a problem that has been pointed out in the literature on peer assessment for some time now, but it seems that despite calls to take action (e.g., Strijbos and Sluijsmans, 2010; Van Zundert et al., 2010), this observation still holds for recently published studies. In addition to a lack of experimental and quasi-experimental designs, a frequent problem is that researchers do not determine if their samples are large enough to detect the potential effects. A power analysis was conducted in only one study (Ocampo et al., 2023), which was one of

the studies that included the most participants. An explanation for these methodological limitations could be that studies on peer assessment are almost always conducted in real educational contexts and that the emphasis is put on ecological validity. It is however possible to combine ecological and internal validity as was showcased by a number of the included studies (e.g., Panadero et al., 2013; F. Yu, 2012).

4.3 Implications for practice

Based on the present review, we can offer four recommendations for practitioners implementing peer assessment. The first one is that a training session prior to peer assessment is preferable. Previous literature reviews have already shown that the most critical factor in ensuring peer assessment effectiveness is the provision of training (e.g., H. Li et al., 2020). Our review evidenced that training also has a positive impact on interpersonal processes such as students' feeling of comfort (Ocampo et al., 2023; Parikh & Sheehan, 2016). Combined, these findings suggest that training should be considered an indispensable part of a peer assessment exercise.

The second recommendation is to carefully consider the use of anonymity. Indeed, evidence suggests that anonymity can help make students more comfortable, yet it could also have some detrimental effects such as conveying negative emotions (Kaya, 2021) or creating negative perceptions of the peer as an assessor (F. Yu & Wu, 2011). Linked to this, we recommend gauging students' perceptions before deciding how to design peer assessment activities. Indeed, it seems from the data that the different interpersonal processes are not all impacted in the same ways by the instructional design elements. Privacy, for example, could be a good idea for students who are very uncomfortable with the idea of conducting a peer assessment (S. Kim, 2019; Lim et al., 2021; Raes et al., 2015; Su, 2023), but not for students who fear that peer assessment will be unfair (Açkay et al. 2021; Lin, 2018). It is therefore important to understand students' primary concerns, before deciding on designing the peer assessment activity.

Finally, it is important to bear in mind that interpersonal processes are not the only factor to consider when designing peer assessment. For example, it would be problematic to conclude from the study of Praver et al. (2011) that we should ask students to provide only scores, and not written comments, to prevent negative emotions from arising. Indeed it has been established that written comments are more beneficial for learning than scores (Lipnevich & Smith, 2009; Shute, 2008), so a wiser conclusion would be to ask students to provide comments while manipulating other design elements to mitigate the possible increase in negative emotions.

4.4 Conclusion

The present study reviewed 27 empirical studies on the impact of the instructional design elements of peer assessment on interpersonal processes. Privacy is the most commonly-investigated design element. It has a positive impact on some interpersonal processes and a negative impact on others. The second most commonly-investigated design element was the

format of peer assessment: it seems that increased amounts of guidance have a positive impact on interpersonal processes, but studies with stronger research designs are needed to confirm this. The presence of training, the third most commonly-investigated design element, seems to also have a positive impact on interpersonal processes, which confirms that training is an important step with regard to peer assessment. The review revealed that the impact of other design elements has not been sufficiently studied to conclude anything, and that further research is needed. It also revealed some limitations in the field, such as a lack of clear definitions and adapted instruments to measure interpersonal processes or a dearth of studies using strong research designs. These gaps and needs in the research should inform further research. Nevertheless, the present review still provides some implications for practice, such as the importance of training students prior to peer assessment, and the importance of gauging students' perceptions to decide how best to design peer assessment activities.

5. Appendices

5.1 Appendix A. Search string for ERIC

((MAINSUBJECT.EXACT("Friendship") OR MAINSUBJECT.EXACT("Interpersonal Relationship") OR MAINSUBJECT.EXACT("Trust (Psychology)") OR MAINSUBJECT.EXACT("Motivation") OR MAINSUBJECT.EXACT("Peer Relationship")) OR (motivation* OR emotion* OR interpersonal* OR "psychological safety" OR friend* OR "value diversit*" OR "value congruen*" OR interdependen* OR trust* OR anonymity OR relation* OR social OR "fear of disapproval" OR anxiety OR sham*)) AND ((MAINSUBJECT.EXACT("Higher Education") OR MAINSUBJECT.EXACT("College Students") OR MAINSUBJECT.EXACT("Elementary Schools") OR MAINSUBJECT.EXACT("Elementary Education") OR MAINSUBJECT.EXACT("Junior High Schools") OR MAINSUBJECT.EXACT("Secondary Education") OR MAINSUBJECT.EXACT("Elementary School Students") OR MAINSUBJECT.EXACT("High Schools") OR MAINSUBJECT.EXACT("Middle Schools") OR MAINSUBJECT.EXACT("Junior High School Students") OR MAINSUBJECT.EXACT("Middle School Students") OR MAINSUBJECT.EXACT("Compulsory Education")) OR ti(student* OR school* OR "elementary education" OR "secondary education" OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher education" OR college* OR "high school*" OR "highschool*" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*) OR ab(student* OR school* OR "elementary education" OR "secondary education" OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher education" OR college* OR "high school*" OR "highschool*" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*)) AND (MAINSUBJECT.EXACT("Peer Evaluation") OR ti("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*") OR ab("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*"))

5.2 Appendix B. Search string for PsycInfo

((MAINSUBJECT.EXACT("Elementary Education") OR MAINSUBJECT.EXACT("Elementary Schools") OR MAINSUBJECT.EXACT("High Schools") OR MAINSUBJECT.EXACT("College Students") OR MAINSUBJECT.EXACT("Higher Education") OR MAINSUBJECT.EXACT("High School Education") OR MAINSUBJECT.EXACT("Middle Schools") OR MAINSUBJECT.EXACT("Secondary Education") OR MAINSUBJECT.EXACT("Middle School Education") OR MAINSUBJECT.EXACT("Junior High School Students") OR MAINSUBJECT.EXACT("Middle School Students") OR MAINSUBJECT.EXACT("Elementary School Students")) OR ab(student* OR school* OR "elementary education" OR "secondary education" OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher education" OR "college*" OR "high school*" OR "highschool" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*) OR ti(student* OR school* OR "elementary education" OR "secondary education" OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR "higher education" OR "college*" OR "high school*" OR "highschool" OR "grade 1" OR "grade 2" OR "grade 3" OR "grade 4" OR "grade 5" OR "grade 6" OR "grade 7" OR "grade 8" OR "grade 9" OR "grade 10" OR "grade 11" OR "grade 12" OR "primary education" OR "public school*" OR "secondary school*" OR "junior high" OR "K-12*" OR k12 OR preuniversit* OR "pre-universit*" OR universit*)) AND (MAINSUBJECT.EXACT("Peer Evaluation") OR ab("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*") OR ti("peer evaluation*" OR "peer assessment*" OR "peer feedback" OR "peer grading" OR "peer marking" OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR "peer-feedback" OR "peer-grading" OR "peer-marking" OR "peer-review*")) AND ((MAINSUBJECT.EXACT("Trust (Social Behavior)") OR MAINSUBJECT.EXACT("Peer Relations") OR MAINSUBJECT.EXACT("Motivation") OR MAINSUBJECT.EXACT("Interpersonal Relationships") OR MAINSUBJECT.EXACT("Friendship") OR MAINSUBJECT.EXACT("Anonymity")) OR (motivation* OR emotion* OR interpersonal* OR "psychological safety" OR friend* OR "value diversit*" OR "value congruen*" OR interdependen* OR trust* OR anonymity OR relation* OR social OR "fear of disapproval" OR anxiety OR sham*))

5.3 Appendix C. Search string for Scopus

(TITLE-ABS-KEY ("peer evaluation*" OR "peer assessment*" OR {peer feedback} OR {peer grading} OR {peer marking} OR "peer review*" OR "peer-evaluation*" OR "peer-assessment*" OR {peer-feedback} OR {peer-grading} OR {peer-marking} OR "peer-review*")) AND (TITLE-ABS-KEY (student* OR school* OR {elementary education} OR {secondary education} OR "elementary school*" OR "middle school*" OR learn OR learner* OR learning OR {higher education} OR college* OR "high school*" OR highschool* OR {grade 1} OR {grade 2} OR {grade 3} OR {grade 4}) OR TITLE-ABS-KEY ({grade 5} OR {grade 6} OR {grade 7} OR {grade 8} OR {grade 9} OR {grade 10} OR {grade 11} OR {grade 12} OR {primary education} OR "public school*" OR "secondary school*" OR "junior high*" OR "K-12*" OR k12 OR preuniversit* OR "pre universit*" OR universit*)) AND (ALL (motivation* OR emotion* OR interpersonal OR {psychological safety} OR friend* OR "value diversit*" OR "value congruen*" OR interdependen* OR trust* OR anonymity OR relation* OR social OR {fear of disapproval} OR anxiety OR sham*))

5.4 Appendix D. Critical appraisal for quantitative studies

		Li (2017)	Lin (2018)	Ocampo et al. (2023)	Panadero et al. (2013)	Parikh & Sheehan	Praver et al. (2011)	Senden et al. (2023)	Yu (2012)	Yu & Wu (2011)
General	Description should be specific and detailed	2	2	2	2	2	2	2	2	2
	Whether the study was experimental or nonexperimental is noted	2	2	2	2	2	2	2	2	2
Participants	There is a clear and detailed description of the participants and their characteristics	2	2	2	2	1	1	2	2	1
	There is an explanation of how the sample was selected	2	2	2	2	2	2	2	2	2
Data collection tools	Data collection tools are clearly described	2	2	0	0	2	0	2	2	2
	Information is provided as to whether existing, modified, or newly developed data collection tools are used	2	2	2	2	1	2	2	2	2
	There is information about the reliability and validity of the data collection tools	2	0	0	1	0	0	1	2	1
	When appropriate, sample items or questions are included	2	2	2	2	2	2	2	0	2
Procedures	The procedure for conducting the study are clearly and adequately described	2	2	2	2	2	2	2	2	2
	There is a description of the process, duration, and schedule of the study	2	2	2	2	2	2	2	2	2
	The data collection process is described	2	2	2	2	2	2	2	2	2
Data analysis	There is an explanation of the study's design and data analysis	2	2	2	2	2	2	2	2	2
	Statistical procedures used to test the study's hypotheses are listed when appropriate	2	2	2	2	2	2	2	2	2
Ethics	Ethical guidelines are followed	/	/	/	/	/	/	2	/	/
Results	The main findings are reported	2	2	2	2	2	2	2	2	2
	The results are clearly presented and easy to follow	2	2	2	2	2	2	2	2	2
	There are tables, graphs, or charts that help explain the findings	2	2	2	0	0	2	2	2	2

		Li (2017)	Lin (2018)	Ocampo et al. (2023)	Panadero et al. (2013)	Parikh & Sheehan	Praver et al. (2011)	Senden et al. (2023)	Yu (2012)	Yu & Wu (2011)
Discussion	The results are examined, explained and evaluated	2	2	2	2	2	2	2	2	2
	There is a discussion of the findings in relation to the original questions and/or hypotheses	2	2	2	2	2	2	2	2	2
	Findings are discussed in relation to prior research on the topic	1	2	2	2	2	0	2	2	2
	The study's conclusions are justified based on the results of the study	2	2	2	2	2	1	2	2	2
	The authors provided possible explanations for unexpected results	2	2	2	1	/	1	2	2	1
	Weaknesses and limitations of the study are discussed	2	2	2	2	2	2	2	1	1
	Implications of the findings for practitioners are presented	1	0	0	2	2	1	0	1	2
	There are suggestions for further research	2	2	2	2	1	1	2	1	2
Validity and reliability	There is evidence of the study's internal validity, especially in experimental studies	2	2	2	2	1	1	2	2	2
	There is evidence of external validity	2	2	2	2	2	2	2	2	2
	The validity of the data collection instrument(s) is presented	2	0	0	1	0	0	0	1	0
	The reliability of the data collection instrument(s) is presented	2	0	0	1	0	0	2	2	2
	The level of reliability of the measure is appropriate for its use in the study	2	/	/	1	/	/	1	2	2
	The reliability levels of the data collection instrument(s) are reported when validity is discussed	2	0	0	1	0	0	2	2	2

Note. 0 = not respected, 1 = partly respected, 2 = respected, / = no information

5.5 Appendix E. Critical appraisal for qualitative studies

		Ackay et al. (2021)	Kaya et al. (2023)	Xue et al. (2023)	Zhang et al. (2022)
General	Description should be specific and detailed	1	1	1	1
Researcher's role	Researcher's subjectivity and preconceived notions about the topic are discussed	0	0	0	0
Setting	Detailed information about the setting and rationale for selecting it is provided	1	1	1	1
Participants	Background information about the participants is offered	1	1	1	1
	Participant selection is based on their knowledge of the topic and their ability and willingness to share it	2	2	2	2
Data collection tools	The process of data collection is clearly and comprehensively described	1	2	2	2
	Methods used to generate data are clearly described	1	1	2	2
	Multiple sources of data collection were used to enable triangulation	2	1	2	2
	Data collection strategies are appropriate for the research question(s)	2	2	2	2
Procedures	The procedure for conducting the study are clearly and adequately described	1	1	1	1
	A record of the process, including interviews, observations, protocols, field notes, and transcripts, is available in the appendix	0	0	0	0
Data analysis	The method of data analysis is logically developed and clearly presented	2	1	2	2
	The process for coding the data and for identifying categories, themes, and pattern is outlined	1	1	1	1
	The relationship among the pattern are supported by the data	2	2	2	2
Ethics	The identity of the participants and the settings is kept confidential	2	2	2	2
	Formal ethical permissions to conduct the study were obtained	/	/	2	2
	The participants were informed about the nature and research and their confidentiality was assured	/	/	2	2
Results	There is a clear explanation of how the findings are drawn from the data analysis	2	1	2	2
	The results are reported in a "thick" narrative	2	1	2	2
	The findings are supported by illustrative quotations and specific instances	2	1	2	2
	The findings clearly and adequately address the research question(s)	2	2	2	2
	Contradictory findings and outlier data are included in the research findings	2	1	1	1

		Ackay et al. (2021)	Kaya et al. (2023)	Xue et al. (2023)	Zhang et al. (2022)
Discussion	The meaning and implications of the study are discussed in relation to the research questions	2	2	2	2
	The discussion relates the study's findings to existing literature on the topic	2	2	2	2
	The findings are discussed within the broader social, cultural, and/or political contexts	0	0	1	1
	The researcher reflects on the experiences of conducting the study	0	0	0	0
Validity and trustwothiness	The credibility, transferability, dependability, and conformability of the study are discussed	0	0	0	0
	The research process and data collection phases entailed a prolonged engagement	/	/	/	/
	The researcher discussed personal biases and preconceived assumptions	0	0	0	0
	The research process, procedures, and documentation are recorded	2	/	2	2
	There is a rich "tick" description of the participants, the setting(s), and the events	1	1	1	1
	There is an explanation of how the data results were triangulated	1	0	1	1
	Processes for alternative possible interpretation and discrepant information are presented	0	0	0	0
	Procedures for checking the research accuracy from the perspective of participants and peer researchers are used	0	0	0	0

5.6 Appendix F. Critical appraisal for mixed-method studies

		Cheng et al. (2023)	Ching & Hsu (2016)	Güler (2017)	Kim (2019)	Kim & Lan (2021)	Lim et al. (2021)	Raes et al. (2015)	Rotsaert et al. (2018)	Su (2023)	Thanh & Gillies (2010)	Tzeng et al. (2022)	van Heerden & Bhauruthram (2021)	Vanderhoven et al. (2015)	Yu & Liu (2009)
General	The reasons for choosing a mixed-methods approach are clearly explained	0	0	1	0	0	1	1	0	1	0	0	0	1	0
	The methods used match the research questions, and each method is designed to answer at least one research question	2	2	2	1	1	2	2	2	2	1	2	1	2	2
Researcher's role	The researcher role – an objective or subjective stance – is assumed to match the research question being investigated	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Site and participants	The site and participants are fully described and the reasons for selecting them are presented	1	1	1	1	1	1	1	1	1	0	1	0	0	0
Data collection tools	Data collection tools, both quantitative and qualitative, are described in detail	1	2	2	1	2	2	2	2	2	2	2	1	1	2
	Sample items from data collection tools are included	2	2	0	1	2	2	2	2	2	2	2	0	2	2
Procedures	The steps used to collect both quantitative and qualitative data are explicitly described	2	2	2	2	2	2	2	2	2	2	1	1	2	2
	The procedures for data analysis reflect the type of mixed data designs (embedded, two-phased or triangulation)	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Data analysis	It is indicated whether the analysis occurred sequentially or concurrently	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	It is indicated when the integration of quantitative and qualitative data occurred	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	It is clear whether the quantitative or qualitative approach is given priority	1	1	1	1	1	1	1	1	1	1	1	0	1	1

		Cheng et al. (2023)	Ching & Hsu (2016)	Güler (2017)	Kim (2019)	Kim & Lan (2021)	Lim et al. (2021)	Raes et al. (2015)	Rotsaert et al. (2018)	Su (2023)	Thanh & Gillies (2010)	Tzeng et al. (2022)	van Heerden & Bhauruthram (2021)	Vanderhoven et al. (2015)	Yu & Liu (2009)
Ethics	The ethical considerations listed for quantitative and qualitative studies are followed	/	/	/	2	/	2	/	/	/	/	/	/	/	/
Results	Findings are presented in different formats based on the method and data source used to generate the data	2	1	1	1	1	1	2	2	1	1	2	2	2	/
	An explanation is offered of how findings from the quantitative and qualitative methods are integrated	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Discussion	The major findings are reiterated in relation to the study's questions	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Conflicting findings are examined and properly explained	2	/	/	1	2	/	/	2	/	/	2	1	1	/
	The study's findings are evaluated in relation to prior research in the field	2	2	2	2	2	2	2	2	2	0	2	2	2	2
Validity, reliability, and trustworthiness	The procedure for ensuring the validity, reliability, and trustworthiness of mixed-methods research follow the specific requirements of each approach	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Chapter 3

How to make students feel safe and confident?

Designing an online training targeting the social nature of peer feedback

Peer feedback has proven to be very beneficial for student learning; however, by its social nature, peer feedback raises concerns for many students. To diminish these, and maximize the benefits of a peer feedback activity, we created an online training targeting psychological safety and trust. The objective of this chapter is to describe the design process and to detail the composition of the training. The training was delivered to higher education students and included five stages: discovery of students' representation, lecture on how to provide effective feedback, peer feedback practice, role-play and discussion in small groups and summary of key learning points. A questionnaire one month after the training and interviews with five students revealed that students' engagement in and perceptions of the training were highly variable. Implications for research and practice are discussed.

Keywords

Peer feedback, training, social interactions, higher education, psychological safety, trust

Senden, M., De Jaeger, D., Rotsaert, T., Leroy, F., & Coertjens, L. (2023). How to Make Students Feel Safe and Confident? Designing an Online Training Targeting the Social Nature of Peer Feedback. In Noorozi, O., & De Wever, B. (Eds), *The Power of Peer Learning: Fostering Students' Learning Processes and Outcomes* (pp. 325-346). Cham: Springer International Publishing.

How to make students feel safe and confident?

Designing an online training targeting the social nature of peer feedback

---

1. Peer feedback and its social nature

1.1 The benefits of peer feedback

How to foster student learning is a crucial question in educational research. Feedback – defined as "a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies" (Carless & Boud, 2018, p. 1315) – has been proposed as an important tool for student learning and numerous studies have confirmed this (e.g. P. Black & Wiliam, 1998; Hattie & Timperley, 2007). Indeed, in a recent meta-analysis of 435 studies, Wisniewski and colleagues (2020) found that feedback has a moderate size effect (d=0.48) on students' learning.

A particularly effective kind of feedback is feedback from peers. Indeed, empirical evidence supports the value of peer feedback and suggests it can even be a more useful tool for learning than teacher feedback. Wisniewski et al. (2020) found student-to-student feedback to be more efficient than teacher-to-student feedback. In another recent meta-analysis, Double and colleagues (2020) addressed the effect of peer assessment (which sometimes, but not always include peer feedback). Based on 54 (quasi-) experimental studies, this meta-analysis concludes that peer assessment impacts student learning more positively than teacher assessment.

The value of peer feedback does not only lies in the fact it helps students to developed specific subject-related learning, it also pushes them to develop more general feedback skills (Carless et al., 2011). By giving opportunities for students to practice making judgements, peer feedback contributes, for example, to the development of evaluative judgment, which is defined as "the capacity to make decisions about the quality of work of self and others" (Tai et al., 2018, p. 471). Evaluative judgement is a necessary skill for students to become independent lifelong learners, which should be a goal of higher education (Tai et al., 2018).

1.2 Students' concerns and how to take them into account

Given that peer feedback can be a very beneficial activity for students' learning, we could expect students to be eager to participate in peer assessment activities. However, this is only partly the case. Indeed, the majority of students report to like peer assessment and to find it

useful (Hanrahan & Isaacs, 2001; Mulder et al., 2014). However, they also express a series of concerns. These concerns are various but have a common element: they emerge from the fact that peer assessment is a social experience (e.g. Hanrahan & Isaacs, 2001; Mulder et al., 2014; Wilson et al., 2015). Some students fear, for example, that their peers will be biased or will not put enough effort into their assessment, feel they lack the skills to evaluate their peers, find it difficult to be objective and feel uncomfortable evaluating their peers and being evaluated by them (Hanrahan & Isaacs, 2001; Mulder et al., 2014; Wilson et al., 2015). These concerns are not anecdotic: in Stanier (1997)'s study, 40% of the students found peer assessment to be an uncomfortable experience and in Mostert and Snowball (2013)'s study, 29% felt that their peers did not engage enough in the activity and 19% did not trust their peers as assessors.

The main suggestion in literature to overcome students' concerns linked to the social nature of peer feedback is anonymity. In F. Yu & Liu (2009)'s study, for example, students preferred using a surname rather than their real name in a peer assessment activity and, in Vanderhoven et al. (2015)'s study, students experienced less peer pressure and fear of disapproval in an anonymous peer assessment activity compared to a non-anonymous one. Rotsaert et al. (2018) have found that, when peer feedback is used multiple times in a course, fading anonymity can be used as an instructional scaffold. When students first had the opportunity to experience peer feedback anonymously, they continue to provide feedback of the same quality and to feel safe when anonymity is removed and the importance they place on anonymity decreases.

However, expecting anonymity to relieve every tension created by peer assessment is unrealistic (Panadero & Alqassab, 2019). Panadero and Alqassab (2019)'s literature review on the effects of anonymity in peer feedback shows mixed results, with only a slight positive tendency toward anonymity. Their main conclusions are that more research on the effects of anonymity is needed and that the instructional context and goals need to be considered (Panadero & Alqassab, 2019).

Moreover, anonymity is not always possible in peer feedback (e.g. feedback on an oral presentation), and, even when possible, it is not always desirable. Indeed, anonymity necessarily means that students can not interact with one another and discuss the feedback, which removed the richest part of feedback if we see it as a dialogical process (Ajjawi & Boud, 2017). Additionally, not all potential undesirable effects of the social nature of peer feedback can be cancelled out by the use of anonymity (e.g. the fear of not being able to provide valuable feedback). Therefore, it is necessary to find other ways to create an environment in which students feel safe and comfortable to participate in peer feedback activities.

To ease tensions linked to the social aspects of peer feedback, students could be trained on these aspects. In L.Li (2017)'s study, for example, students were assigned to three groups: identity group (the identity of assessors and assesses was revealed to each other), anonymity group (both assessors and assesses were anonymous), and the training group (the identities were known, but students followed a training, aimed at controlling the possible negative

impact of having their identities revealed). Results indicated that both the training and the anonymity groups showed a larger improvement in their performance than the identity group. Moreover, regarding their perception, students in the training group valued peer assessment activities more and experienced less pressure from peers than the students in the two other groups. Thus, it seems that when anonymity is not feasible, training could counteract the negative impact of having students' identities revealed.

1.3 Trust and psychological safety

L. Li (2017)'s study suggests that the provision of peer feedback training could be useful, but in her study, she focused on peer pressure, which is not the only important variable linked to the social nature of peer feedback. Based on the literature on collaborative learning and group work, van Gennip and colleagues (2009) identified four variables that could be of importance in a peer assessment activity, namely:

(1) Psychological safety: the belief shared by members of a group that they can take interpersonal risks in this group
(2) Value congruency: the similarities of team members' opinions about what the tasks, missions and goals of their team should be
(3) Interdependence: the fact that everyone needs to participate actively in the assessment task because, if some students do not provide feedback it will have an impact on the students whose work they assessed
(4) Trust: (1) the confidence that a student has in his/her own ability to assess their peers' work – i.e., trust in oneself – and (2) the trust in their peers' capacity to assess his/her work – i.e., trust in peer

Of these four variables, trust and psychological safety appear to be key to consider when implementing a peer feedback activity. Indeed, van Gennip, Segers, and Tillema (2010) found that, in the context of secondary-vocational education, a high level of psychological safety and trust (in the self and the peer) had a positive impact on perceived learning. For value congruency and interdependence, the relationship with perceived learning was less clear. Panadero (2016) confirms this: the relevance of trust and psychological safety is more evident than that of value congruency and interdependence, as these two latter variables are relevant in contexts where students have shared goals, which is not necessarily the case in the context of peer feedback (e.g. with online anonymous peer feedback). Consequently, for the training, we focused upon trust and psychological safety.

Regarding the term 'trust", it is important to highlight its two facets: trust in oneself and trust in peers (van Gennip, Gijbels et al., 2010). A student can trust his or her ability but not the ability of his or her peers, or vice versa. In a study by K. Cheng & Tsai (2012), for example, the majority of students (74%) trusted their abilities to assess their peers. A smaller percentage of students (57%) also trusted their peers' ability.

The notion of psychological safety originally came from organisational psychology where it can be defined as the "perceptions of the consequences of taking interpersonal risks in a particular context" (Edmondson & Lei, 2014, p. 24). Linked to its origin, the majority of research on psychological safety has been conducted in the working environment (e.g.

Edmondson et al., 2007); however, even in the working environment, an emphasis was placed on learning behavior. When people feel psychologically safe, they are less afraid to take interpersonal risks, which means that they are more willing to express themselves without worrying about possible negative reactions from other members of their team. Therefore, in a psychologically safe environment, team members are more willing to carry out learning behavior like seeking feedback, asking for help or talking about errors (Edmondson, 1999). De Stobbeleir and colleagues (2020) confirmed this: when employees perceived their environment as psychologically safe, they seek more feedback from their peers.

In an educational context, Soares & Lopes (2020) have shown that psychological safety has a positive influence on academic performance. 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. Hence, psychological safety is a requirement for peer feedback. In the context of peer assessment, psychological safety is defined by Panadero (2016, p. 251) as "the extent to which students feel safe to give sincere feedback as an assessor and do not fear inappropriate negative feedback as an assessee".

1.4 Online training

Online courses, such as MOOCs, have gained popularity over the last decade (Shah, 2020). For such MOOCs, it is challenging to exceed knowledge transfer and provide a real educational experience for students (Suen, 2014). To achieve this and keep the workload for teaching staff feasible, peer feedback is frequently used (Suen, 2014): it allows students to obtain feedback to enhance their learning.

Yet, students' concerns linked to the social nature of peer feedback may be amplified in MOOC settings. As Suen (2014) explains, peer feedback in MOOCs takes place in a context where there is, at best, few instructor mediation, supervision or guidance and where students have little incentive to take peer feedback activities seriously. In this context, students are often dissatisfied with the use of peer feedback and complain that their peers give them superficial or inconsistent feedback (Hew, 2018). Therefore, taking students' concerns into account and training them before a peer feedback activity could be even more important in MOOCs than in traditional on-campus courses.

1.5 The present study

Although some leads exist on how to take into account the interpersonal context when designing peer feedback activities, only a few studies have been conducted (e.g. Vanderhoven et al., 2015; Rotsaert et al., 2018). It remains veiled how trust and psychological safety can be stimulated in an online setting. To fill this gap, we set out to design an online training targeting these two aspects, to optimize students' learning from peer feedback.

The purpose of this chapter is to present the training and the rationale behind its different components. In the section "Training design", we will explain how the literature was explored to find effective tools for training. Subsequently, in the section "Training

procedure", a detailed outline of the different stages of the training will be presented. Moreover, even though an evaluation of the training is beyond the scope of this article, we will give some elements on how the training was received by students in the section "Students' perceptions of the training session". Finally, we will discuss some limitations and perspectives.

It is important to specify that the training is conceived for peer feedback activities where the feedback are performance feedback, i.e. feedback on students' performance, and not process feedback, i.e. feedback on the way students performed a task (Gabelica et al., 2012). Indeed the training is based on research done on performance feedback and its specific challenges, whose results cannot necessarily be generalized to process feedback (Gabelica et al., 2012).

2. Training design

2.1 Training objectives

Our purpose is to design a training to tackle students' feelings of distrust and psychological unsafety before participating in a peer feedback activity. Given that it is recommended to train students on how to provide peer feedback (van Zundert et al., 2010), we also included a more general part to our training to help students provide effective feedback. Therefore, the training has six objectives:

(1) to clarify for the students what the objectives and advantages of peer feedback are
(2) to increase students' skills for providing effective feedback
(3) to address students' concerns about peer feedback
(4) to increase students' feeling of psychological safety
(5) to increase students' trust in their ability to assess others
(6) to increase students' trust in their peers' ability to assess them

For our two first objectives (objectives 1 and 2) the learning outcomes are knowledge and skills. The learning outcomes of the last four objectives (objectives 3 to 6) can be considered as attitudes, which are defined as "beliefs and opinions that support or inhibit behavior" (Oskamp, 1991, cited by Blanchard & Thacker, 2012, p. 37). Because the training mostly targets students' attitudes, it must allow the active participation of students. Therefore, the training consisted mainly of role-plays and discussions, two effective methods to transform attitudes (Blanchard & Thacker, 2012).

2.2 Inspirations from existing training

As it is common to train students before a peer feedback activity and some training in other contexts than peer feedback may have targeted trust and psychological safety, we additionally explored this literature.

It has been shown that training students before a peer assessment activity improved the reliability of peer assessment, peer assessment skills and students' attitudes toward peer assessment (van Zundert et al., 2010). Training can focus on various aspects, like how to

decide what is important to assess, how to judge a performance or how to provide feedback for future learning (Sluijsmans et al., 2004). Their length may also vary, some of them being very comprehensive (e.g. Sluijsmans et al., 2002), while others are shorter but can still be effective, as shown by Alqassab and colleagues (2018a). It is from the latter that we drew to design our training, which we want to keep short enough to fit it into already busy schedules: it seems unrealistic that more than one session will be dedicated to peer feedback training in a course where peer feedback is only a way to help students learn and not an objective in itself.

The training created by Alqassab et al. (2018a) consisted of a general discussion on peer assessment, a lecture on Hattie and Timperley (2007)'s framework, group and individual exercises to integrate the theory and practice sessions. Alqassab and colleagues (2018a) found that, for medium or high-achieving students, the training increased the proportion of self-regulation feedback, which are considered higher-level feedback and are more effective (Hattie & Timperley, 2007). The training did not affect low-achieving students.

In addition, Dusenberry and Robinson (2020) created a training session to increase students' feeling of psychological safety before working in small groups. Their training lasted 50 minutes and was composed of a video lecture, a short discussion and a hands-on exercise. Contrary to their hypothesis, the level of psychological safety was not higher for the students who followed the training than for the students in the control group. They identified several limits in their training, which could explain this absence of a significant effect. A first limitation is that their training was not context-specific. The same training was given to a variety of students, working on various projects, and there was no direct link between the training and the teams' projects. A second limitation is that the video lecture took up about half of the training time, which did not leave much time for more active learning methods. Consequently, if we want to make sure our training is effective, it seems important to avoid these pitfalls by making our training context-specific and by favouring active learning methods, as also recommended by Blanchard & Thacker (2012).

3. Training procedure

We provided the peer feedback training to third-year university students in physical education following a seminar on acrobatic sports didactics. Forty-one students were enrolled in this mandatory seminar, but six students did not participate in either the training or the peer feedback activity (even though both were mandatory). During this seminar, students had to create an instruction sheet illustrating a gymnastic exercise, assess the instruction sheet of seven of their peers and improve their work based on the feedback they received. The training took place during the second seminar session. It was provided by a researcher (first author) in collaboration with the course's professor (fourth author). The online training was composed of five stages (see table 1): discovery of student's representation, lecture on how to provide effective feedback, peer feedback practice, role-play and discussion in small groups, and summary of key learning points.

We conducted the training online, through the Microsoft-Teams platform. Like other videoconferencing applications (e.g. Zoom), Teams allows us to divide participants into break-out rooms, which was essential as most of the training time was spent in subgroups.

Doing the training online incited us to carefully plan it. As Bolinger and Stanton (2020) explained, it is possible to run synchronous online role-plays but the logistics are more difficult to manage. It is not possible, for example, to pass sheets around or to easily identify which students have questions once they are in the break-out rooms. To address these difficulties, we tried to make the instructions as explicit as possible (e.g. what students were expected to do, how much time they had…) and we gave them a very detailed roadmap for the role-plays. We also made sure that there were two instructors present online, which made it possible to visit all sub-groups to answer questions.

As mentioned above, it was challenging to find a balance between the importance of having comprehensive training and the necessity of keeping it time-efficient, so it can be relatively easily inserted into courses. Table 1 presents an overview of the training, with the timing devoted to each activity of this two-hour session. As recommended by Blanchard and Thacker (2012), we started by targeting the knowledge and skills before moving on to attitudes.

Table 1. Overview of the training session

Stage	Students' activities	Objectives	Timing
Discovery of students' representation	In the large group, students answer to a Wooclap (active)	To adapt the rest of the training to address students' concerns about peer feedback (objective 3)	10 min
Lecture on how to provide effective feedback	In the large group, students listen to the lecture (passive)	To increase students' skills for providing effective feedback (objective 2)	10 min
Peer feedback practice	Individually and then in pairs, students practice giving feedback and discuss it (active)	To increase students' skills for providing effective feedback (objective 2)	30 min
Role-play and discussion in small groups	In small groups, students perform role-play and discuss in small-groups (active)	To increase students' feeling of psychological safety (objective 4) To increase students' trust in their ability to assess others (objective 5) To increase students' trust in their peers' ability to assess them (objective 6)	50 min
Summary of key learning points	In the large group, students listen to and contribute to the synthesis (passive)	To clarify the objectives and advantages of peer feedback (objective 1) To address students' concerns about peer feedback (objective 3)	20 min

Note. The first part (discovery of students' representations) took place a week before the training session.

3.1 Stage 1. Discovery of students' representations

A week before the training session, we asked students to answer some questions through an interactive platform (Wooclap). We asked them to write down in one word what peer feedback means to them, to write down the advantages and disadvantages of peer feedback and to tell us if they had any concerns or fear linked to the use of peer feedback. For the three first questions, students saw the responses of other students appear live and could *like* them. As an example, Figure 1 is the word cloud generated by students' answers regarding the disadvantages of peer feedback.

Figure 1. Word cloud generated by students' answers regarding the disadvantages of peer feedback

Discovering students' representations allowed us to tailor the training to this specific group of students, which could enhance training efficiency (Dusenberry & Robinson, 2020). This group of students were concerned that their peers are not qualified enough to assess them, that their peers are not objective enough to assess them (more precisely they fear that they will be "too nice") and that they themselves are not qualified enough to assess their peers. These results confirmed the relevance of providing a training targeting the notion of trust. Some aspects linked to psychological safety also emerged, although to a lesser extent (e.g. the fear of being judged as stupid).

3.2 Stage 2. Lecture on how to provide effective feedback

At the start of the two-hour training, we explained the six objectives (see Table 1) and linked them to students' concerns based on the Wooclap responses. Then we gave a short lecture on how to provide effective feedback based on the framework of Hattie and Timperley (2007). We explained to students that the purpose of feedback is to reduce the gap between actual and desired performance and it should therefore contain an answer to the three following questions: Where am I going? How am I going? And Where to next? (Hattie &

Timperley, 2007). We also described the four feedback levels (self, task, process and auto-regulation), explained why the former two were less effective and gave examples of feedback at each level. The lecture format allowed us to convey essential knowledge to students, but we kept it under 20 minutes as to not lose students' attention (Blanchard & Thacker, 2012).

3.3 Stage 3. Peer feedback practice

For stage 3, students had the opportunity to practice giving feedback and to familiarize themselves with the rubric they will use afterwards for the real peer feedback activity. The day before the training session, students had to hand in an assignment (similar but not identical to the main assignment). During the training, we paired them in breakout rooms and randomly assigned them two assignments. Students had to individually assess the assignments with a rubric and then, in pairs, compare their assessments and discuss possible disagreements. After returning to the large group, time was set aside for them to ask questions. We also asked them to give examples of feedback they would provide and think together about how to make them as effective as possible.

3.4 Stage 4. Role-play and discussion in small groups

As explained by De Ketele et al. (2007), role-play is a training method in which participants interpret the role of different characters in a specific situation, to allow an analysis of the representations, feelings and attitudes related to this situation. What distinguishes role-play from other simulations is its emphasis on interpersonal interactions (Bolinger & Stanton, 2020), which makes it particularly relevant for training on trust (in others) and psychological safety.

To the best of our knowledge, there are no existing role-plays on trust and psychological safety in peer feedback described in the literature. Consequently, we designed them ourselves, following general guidelines provided by Bolinger and Stanton (2020). The role-plays' scenarios were conceived to bring students to project themselves into peer feedback situations and to identify problems that may emerge in these situations. Based on the Wooclap responses (see stage 1), we selected the two most appropriate role-plays among several that we had created. The first role-play consisted of three friends who participated in a peer feedback activity but were all dissatisfied with the received written comments for different reasons (e.g. the feedback were only positive, without any suggestion for improvement). In the second role-play, students had to put themselves in the shoes of three students who had to decide what grade and feedback to give to peers who did a poor oral presentation.

Students were randomly divided into small groups, with each group performing the same role-play simultaneously (i.e. multiple role-plays, Blanchard & Thacker, 2012). This format allowed us to involve all students and to let various elements emerge for the discussion afterwards (given that the scenario will be played slightly different in every group) while taking much less time than if each group had played one after the other (Blanchard & Thacker, 2012).

Once split into groups of six, students received a detailed roadmap with two role-play scenarios and instructions on how to play and discuss them. For each role-play three students acted out the roles while the three others observed the role-play and took notes to inform the following discussion. Having two role-plays allowed each student to play one role, either in the first or second role-play.

After performing and discussing the two role-plays, they stayed in sub-groups to synthesize their discussions. More precisely, we asked them to identify the benefits and interpersonal risks of peer feedback, and what the professor and themselves as students can do to ensure that a peer feedback activity works well.

Participating in role-play simulations can bring discomfort to some students, especially if they are not used to role-playing in class (Bolinger & Stanton, 2020). Therefore, we tried to make the situation as comfortable as possible. Playing in small groups instead of in front of everyone should help students feel at ease. Moreover, by having two role-plays, students more reluctant to participate could observe the first one before actively participating in the second one. And finally, students could choose which role they wanted to play (some roles being more demanding than others).

3.5 Stage 5. Summary of key learning points

The last training stage was an open discussion to synthesize all the sub-groups' ideas in the large group. This method is used to generate participation, find out what participants think or have learned and stimulate recall of relevant knowledge (Blanchard et Thacker, 2012). We asked a student from each group to report the key points of their discussions and took live notes on our slideshow so students could see how the discussion progressed. We used this moment to explain and justify the choices made by the professor for the organisation of the peer feedback activity and to link these choices to the elements discussed by students and with the literature on peer feedback. For example, when a student said that they would feel more confident if they were assessed by more than one peer, we explained that this feeling was coherent with the literature (e.g. Sung et al., 2010) according to which when the number of assessors is large enough, peer feedback is as reliable as teacher feedback and that is why they had to provide feedback to seven of their peers for this course (unlike what they did during the training). We also used this moment to address any remaining concerns.

Based on the discussion, we made a mind map (see figure 2) that we sent to students a few days after the training session. This mind map allows students to keep a record of the key ideas identified together during the training under a visual and accessible format. We also provided them with the slideshow used during the training, for further detail.

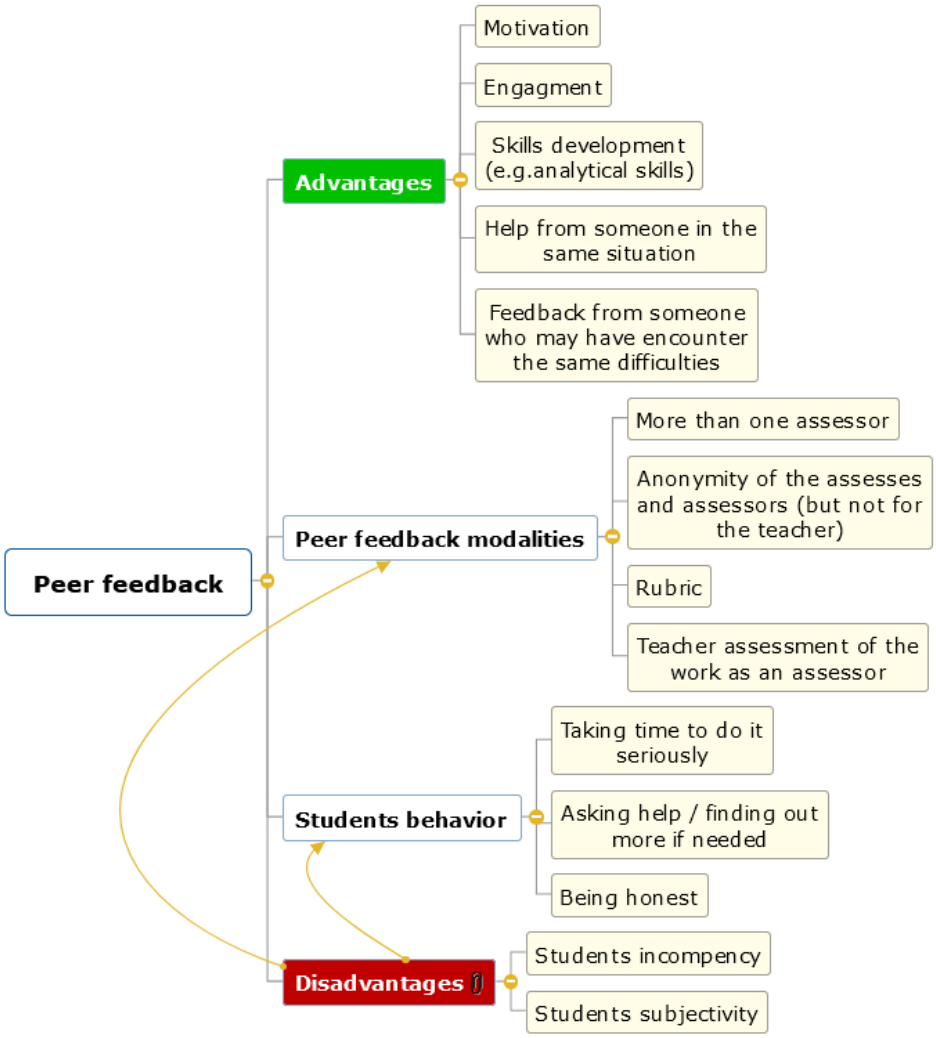


Figure 2. Mind map summarizing the discussions

4. Students' perceptions of the training

A month after the training session, students answered a short questionnaire to assess it. At this point, students had had the opportunity to transfer knowledge into practice because they had already done the peer feedback activity. Of the 35 students who participated in the training and the peer feedback activity, 27 answered the questionnaire (response rate: 77%). At the end of the questionnaire, students could leave their contact information if they agreed to participate in an interview. We conducted semi-directed interviews with the five students who agreed to participate.

4.1 Questionnaire conception and interview process

We constructed our questionnaire based on Grohmann and Kauffeld (2013)'s Questionnaire for professional training evaluation. This questionnaire is based on Kirkpatrick's framework (Kirkpatrick & Kirkpatrick, 2006, cited by Grohmann & Kauffeld, 2013) which distinguishes four levels: reaction, learning, behavior and organisational impact. Grohmann and Kauffeld (2013) have divided the reaction and organisational impact levels into two sub-levels which gives them six subscales (each one composed of two items): satisfaction (reaction level), utility (reaction level), knowledge (learning level), application to practice (behavior level), individual organisational results (organisational level) and global organisational results (organisational level). As the two last subscales are not relevant to our context, we limited our questionnaire to the first four subscales. The eight items of these subscales were subsequently adapted to the higher education context (e.g. the item "In my everyday work, I often use the knowledge I gained in the training" was replaced by "In the peer feedback activity, I used the knowledge I gained in the training") and translated to French.

In addition, we included four items from Holgado-Tello et al. (2006)'s training satisfaction rating scale, which measures participants' general impression of the training.

Our questionnaire is therefore composed of 12 items divided into five subscales (see Table 2 for details). In line with Grohmann and Kauffeld (2013), we used an 11-points response scale. The responses range from 0 per cent to 100 per cent, with steps of 10 per cent. The general impression scale is reliable with a Cronbach's alpha of .854. For the five other subscales, we calculate Spearman-Brown Coefficient, as it is recommended for two-item scales (Eisinga et al., 2013). All scales are reliable, with Spearman-Brown coefficients ranging between .752 and .929. At the end of the questionnaire, we allowed students to add a written comment.

Table 2. Item examples, number of items and cronbachs' α of each subscales

Subscale	Adapted from	Example item	Number of items	Cronbach's α	Spearman-Brown coefficient
General impression	Holgado-Tello et al. (2006)	The training merits a good overall rating	4	.854	/
Satisfaction	Grohmann & Kauffeld (2013)	I enjoyed the training session very much	2	.874	.884
Perceived utility		Participation in this kind of training is very useful for me	2	.850	.861
Perceived learning		I learned a lot of new things in the training	2	.747	.752
Behavioral changes		I successfully manage to apply the training contents in the peer feedback activity	2	.929	.929

The interviews were held using Teams and lasted approximately 30 minutes. We transcribed them and used N-Vivo (version 20.5) to code the data.

4.2 Insights from questionnaire and interview data

As you can see in Figure 3, students' general impression is generally positive (M=66.5, SD=16.4), with the majority of students rating the training around 70%. The satisfaction is a bit lower (see figure 4), with very high variability (M=59.3, SD=20). The same pattern is present for perceived utility (MD=49.6, SD=22.5), perceived learning (MD=53.7, SD=18.3) and behavioral changes (MD=55.4, SD=24) as shown in figures 5, 6 and 7.

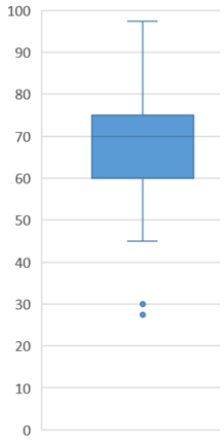


Figure 3. Boxplot for the subscale General impression

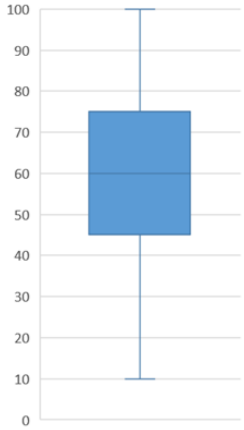


Figure 4. Boxplot for the subscale Satisfaction

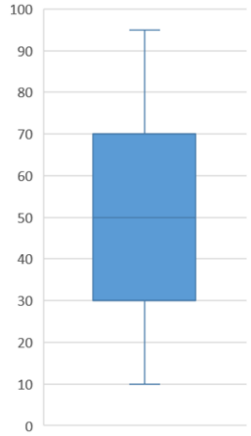


Figure 5. Boxplot for the subscale Perceived utility

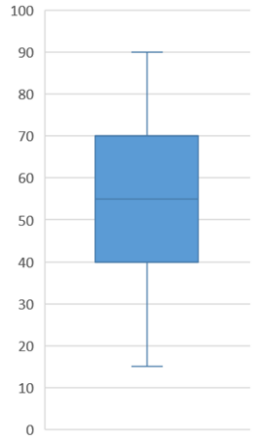


Figure 6. Boxplot for the subscale Perceived learning

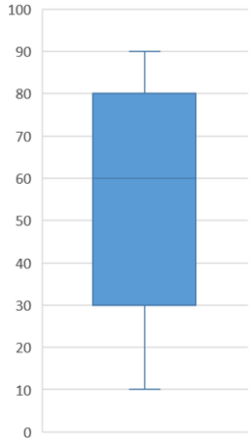


Figure 7. Boxplot for for the subscale Behavioral changes

The most striking result is the high variability of students' perceptions. For each scale (although to a lesser extent for the general impression), the standard deviation and range are wide, with some students who saw little value in the training (with some aspects evaluated at only 10%) and others who seem to have very positive perceptions of the training (with a score of 90 or 100%).

We see two main reasons for the diversity of students' perceptions linked to the training. The first concerns variability in students' needs as illustrated by quotes from David and Robin[1].

> I want to point out that I was already familiar with the principle of peer feedback because I saw it in Research methodology in Master1. That is why I haven't learned as much as others, I think. (David, referring to a course in which the peer review process in research is explained; perceived learning: 55, perceived utility: 70)

> Before we had the course I clearly thought that I wasn't going to be able to… to provide feedback […] and then, with your course, I put this more into perspective, a lot more and I'll say that I was… I wanted to try and assess my peers, and I could see where I had to go. (Robin, perceived learning: 90, perceived utility: 95)

The variability in students' needs seems to stem from their previous experiences. While most students following this course were in their third year and only followed bachelor courses, some students, like *David*, were also taking some master courses (based on the number of ECTS they have acquired). Moreover, students also had different extracurricular experiences. Several physical education students had student jobs as sports instructors or coaches, for example, which enable them to develop assessment and feedback skills. Additionally, given that students practice various sports outside their courses, some students have a much higher level than others in acrobatic sports. This high level of expertise could lead them to overestimate their ability to easily provide feedback in this specific discipline. These factors could explain why some students expressed a strong need for guidance, while others felt they already had the necessary knowledge and skills before the training.

Another possible reason concerns the variability in students' implications in the training session and the seminar more generally. While some students were genuinely interested in the seminar contents, others only followed it because it was mandatory. Given that it was an online session and that students had their cameras off (as not to saturate their wifi), it was more difficult to discern if they were truly paying attention, or if they were even there. In an interview, for example, a student explained that for another session of the course he let his computer with Teams turned on to appear present and went for a run. We tried to make the session as interactive as possible to avoid this, but it is still possible that we lost some students at times.

Moreover, an important part of the training took place in small groups and we observed that some groups worked better than others did in this online context. Indeed, it is well-known that the physical presence of an educator is important for student engagement (e.g. Hunter,

[1] Names have been changed to protect the participants' confidentiality.

cited by Bolinger & Stanton, 2020). The set-up made it difficult for us to know whether the students were taking the role-play seriously and to quickly identify which groups needed help. Although there were two instructors to visit each break-out room, students were just among themselves most of the time and, while most groups seemed to work efficiently, we felt others needed a little push to keep working seriously. This feeling was confirmed by some of the comments in the interviews. For example, Raphaël, who liked the training (satisfaction score of 75), but did not feel that it was useful (perceived utility score of 30) or that he learned from it (perceived learning score of 20), explained: "Well, in the group I was in […] There was some misunderstanding in the group, and we botched the part where we were supposed to take the role.".

This variation in students' engagement could explain why some students felt they learned less from the training or said they did not really apply the training content while doing the peer feedback activity.

4.3 Conclusions and implications

Our goal was to create an online training to tackle students' feelings of distrust and psychological unsafety before participating in a peer feedback activity. To this aim, we explored the literature to find effective learning methods. The training that we created was composed of five stages (discovery of students' representations, theory, practice, role-plays and summary), which allowed students' active participation. This training was implemented in the context of a physical education university seminar and we collected data on how students perceived it. Based on this, we can draw some conclusions, and stemming from them, implications for practice and perspectives for research.

When given the opportunity students express concerns linked to the social nature of peer feedback. Indeed, the gathered responses showed that, even though students saw various advantages to peer feedback, they also raised a series of concerns, like the fear of not being qualified enough or the fear that their peers will be "too nice" when assessing them. This confirms previous findings (e.g. Mostert & Snowball, 2013; Mulder et al., 2014; Wilson et al., 2015) and suggests that, when planning a peer feedback activity, it is essential to take time to let students express these concerns and to address them.

The literature on training methods (e.g. Blanchard & Thacker, 2012) and the studies we drew upon to create the training (e.g. Dusenberry & Robinson, 2020) converged on the idea that (inter)active learning methods are necessary. This training with active learning methods was delivered online, thus making it a potentially useful part of MOOCs (Suen, 2014). For other courses, while normally on campus, online alternatives had to be sought during the COVID-19 outbreak. Indeed, according to UNESCO (2021), more than 220 million tertiary students worldwide have been confronted with university closures and online courses. It is therefore important to conceive learning activities such as role-play that can take place online. In the present case, the group was small enough so we could quickly visit each online break-out room during the sub-groups activities, however, students are generally far more numerous

in MOOCs. Future studies should investigate if the training is feasible with larger groups of students.

We obtained encouraging insights when asking students' opinions about the training. About half the students were positive: they said that they learned from it, found it useful and that it influenced their behavior during the peer feedback activity. Other students valued the training less. Differences in students' perceptions may be explained by factors like their prior knowledge or experience with peer feedback or by their varying engagement with the training (due to the online context). These factors could be investigated in future studies. We expect students' perceptions to be less variable if the training is delivered on campus, as physical presence incites engagement (Hunter, 2004, as cited in Bolinger & Stanton, 2020).

The overall positive impact of the training will have to be confirmed in future studies. Indeed, an evaluation study was beyond the scope of this chapter, in which we focused on students' perceptions of the training. A quasi-experimental study, with a large sample and pre- and post-test should be conducted to verify whether the training has a positive impact on students' perceived level of trust and psychological safety. In addition, given that the goal of peer feedback is to develop student learning and to incite them to be proactive recipients of feedback, such a study could investigate the impact of the training on students' learning due to the peer feedback activity, as well as regarding their feedback literacy skills (Boud et al., 2022).

Now the training focuses predominantly on peer feedback provision; its objectives are to teach students how to provide effective feedback and to ensure they feel safe and confident while doing so. A perspective could be to redesign it so it also encompasses peer feedback processing. Indeed, no matter how good the received feedback is, students still need the support of an adequate learning context to efficiently use it to revise their work (Panadero & Lipnevich, 2022; Wichmann, 2018). Given that interpersonal factors – such as trust and psychological safety – play a role in feedback provision, but also in peer feedback processing (Aben et al., 2019), it would be interesting to create an intervention that explicitly considers the social aspects that play a role in peer feedback processing.

All in all, it seems that an online training with (inter)active methods such as role-plays is a promising way to address students' concerns raised by the social nature of peer feedback.

Chapter 4

Safe and sound

Examining the effect of a training targeting psychological safety and trust in peer assessment

---

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-hour 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, relying on recommendations by Blanchard and Thacker (2012). 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). 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. 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

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

Safe and sound

Examining the effect of a training targeting psychological safety and trust in peer assessment

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 & 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; H. 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 & 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, Segers & Tillema, 2010).

1.1.1 Psychological safety

The notion of psychological safety originally came from organisational psychology (Edmondson & 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 organisations to learn (Edmondson & 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 & 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. 1057).

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, Segers & Tillema, 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 (K. Cheng & 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.1057). 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 (K. Cheng & 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, Segers & Tillema, 2010) and are associated with deeper learning approaches (K. Cheng & 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, 2023), and they would rather be anonymous or use a nickname than use their real name (F. Yu & 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 (F. Yu & 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, Segers & Tillema, 2010 or 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 (L. 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-minute training session that aimed to increase students' perceptions of psychological safety before working in small groups (Dusenberry & Robinson, 2020). This absence of effect could be explained by two factors: the training was not context-specific and 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 L. 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-hour 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 (L. 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 method

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 materials). 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 materials. 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-hour 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-hour training aimed to increase students' perceptions of psychological safety, trust in the self as an assessor, and trust in the other as an assessor.

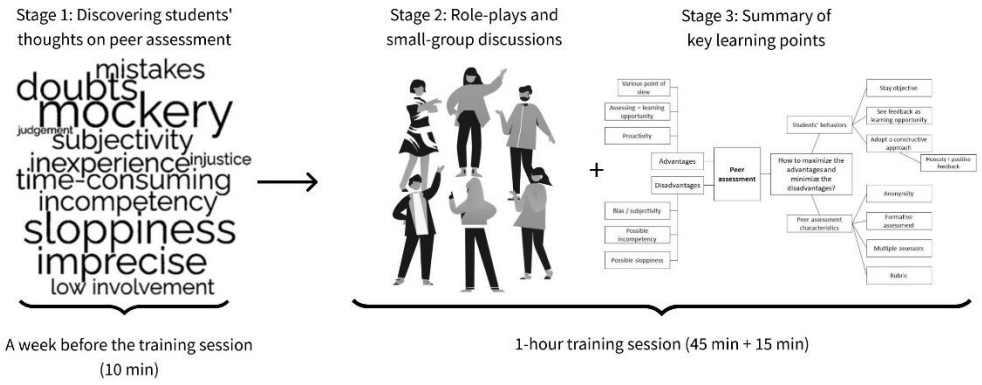
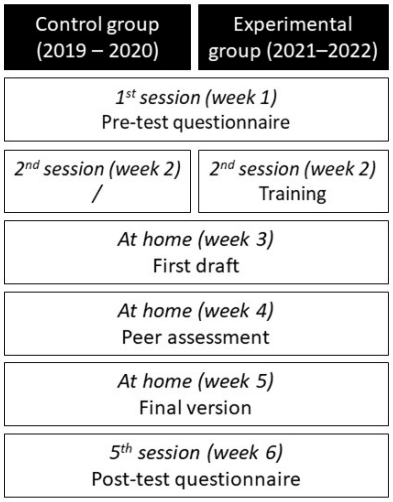


Figure 1. Training session

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 & 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.

Figure 2. Study timeline



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).

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

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 2874 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. (2018b) 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".

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 and 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".*

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 (N. Black et al., 2016). The Gwet's AC_1 indicated a moderate to substantial agreement, Gwet' AC_1 = .61 (95% CI, .546 to .674), p <. 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 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 (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).

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.

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	.76**														
3. Trust self pre-test	.40**	.37**													
4. Trust self post-test	.53**	.68**	.63**												
5. Trust other pre-test	.12	.05	-.06	-0.2											
6. Trust other post-test	.29*	.18	-.10	.13	.65**										
7. Improvement global grade	-.09	-.18	-.02	-.12	.06	-.12									
8. Improvement grade “diagram”	-.18	-.04	-.08	.00	-.08	-.15	.67**								
9. Improvement grade “relevance of the situation”	.04	.03	.12	.09	.40**	.20	.19	.01							
10. Improvement grade “instructions”	-.05	-.16	-.12	-.16	-.04	-.26	.51**	.15	-.11						
11. Improvement grade “success criteria”	.13	-.12	.13	-.10	.06	.01	.53**	.01	.04	.11					
12. Verification feedback	.00	-.05	-.05	-.08	-.03	.04	-.08	-.04	-.11	.02	-.14				
13. Correction feedback	-.03	-.07	-.11	-.15	-.10	.00	-.06	-.01	-.19	-.01	-.03	.57**			
14. Justification feedback	.05	.03	.00	-.02	.18	.24*	-.02	.03	-.02	-.06	.04	.36**	.29**		
15. Question feedback	-.12	.03	.02	.10	-.11	.13	-.24*	-.14	-.15	-.07	-.20	.10	.21*	.11	
16. Suggestion feedback	.04	.00	-.14	-.12	-.02	.18	.03	.09	-.17	.09	.02	.30**	.19	.37**	.19

Note. * = $p < .05$; ** = $p < .01$

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 display 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.

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)

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 ($t(1432.829) = 1.181$, $p = .238$). Moreover, the two groups were comparable ($t(14187.907) = -0.957$, $p = .338$) and, in both groups, the levels of psychological safety were not altered significantly from pre-test to post-test ($t(362.903) = 1.161$, $p = .246$).

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 ($t(1377.802) = -1.164$, $p = .245$). 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 ($t(8621.171) = -2.054$, $p = .040$). In both groups, levels of trust in the self as an assessor did not change statistically from pre-test to post-test ($t(359.655) = 1.327$, $p = .185$).

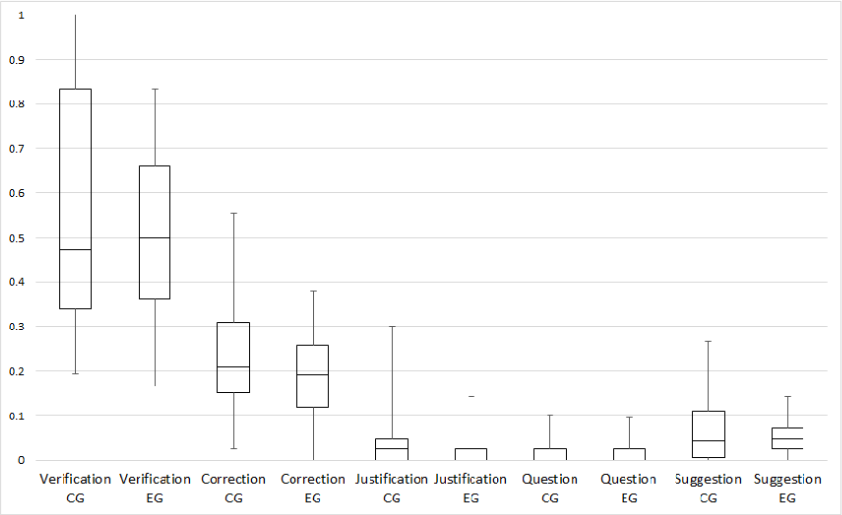
3.1.3 Trust in the other as an assessor

There was no significant interaction effect for trust in the other as an assessor ($t(980.787) = 1.282$, $p = .200$). 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 ($t(867.910) = -5.997$, $p < .001$). In both groups, levels of trust in the other as an assessor did not change statistically from pre-test to post-test ($t(798.583) = -0.422$, $p = .673$).

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 ($U = 1030$, $p = .472$), question feedback ($U = 1072$, $p = .632$), and suggestion feedback ($U = 963$, $p = .218$). However, there was a significant effect for justification feedback ($U = 840$, $p = .020$), 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 ($t(93) = 2.13$; $p = .036$).



Note. CG = Control Group, EG = Experimental Group

Figure 3. Box-plots of students' scores for each feedback type in the control group and the experimental group

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.

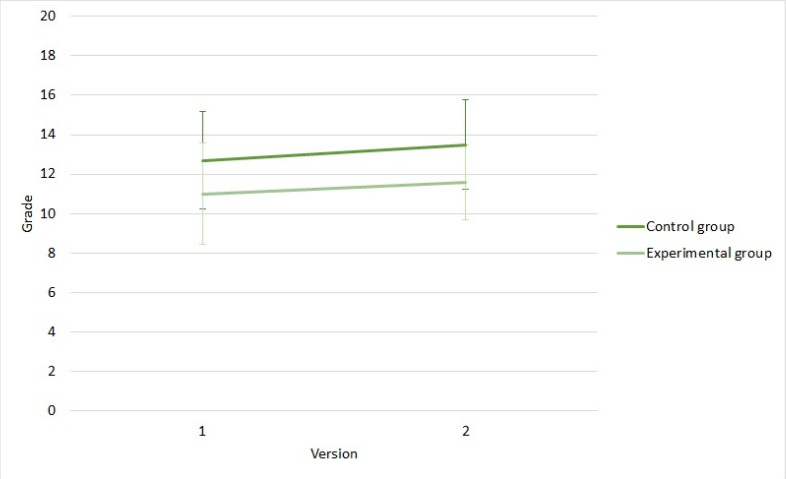


Figure 4. Evolution of students' grades (on /20)

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 < .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 = .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 < .001) and for their final version (U = 559, p < .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 = .003) and "relevance of the situation" (W = 70, p = .006), but not for the criteria "instructions" (W = 365, p = .167) and "success criteria" (W = 354, p = .599). The 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.

Table 6. Means, standard deviations and Mann-Whitney U tests' results for difference scores

	M (*SD*) in the control group	*M* (*SD*) in the experimental group	Mann-Whitney U	Sig.	Effect size (r)
Difference score for the global grade	0.80 (2.07)	0.57 (1.48)	915	.548	.07
Difference score for "diagram"	0.42 (1.00)	0.24 (0.98)	873	.324	.12
Difference score for "relevance of the situation"	0.31 (0.66)	0.02 (0.35)	708	.004	.28

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-hour 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 study that also aimed at making students feel at ease before peer assessment but targeting perceptions of peer pressure (L. 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 & 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 & 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 two 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 & 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, either a correction, a justification, a question, or a 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 & 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, Segers & Tillema, 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 & 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 & 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 & 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 & 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 felt 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 & 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 & 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 & 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 fifty 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-hour 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 & Thacker, 2012). Moreover, with a 1-hour 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 (L. 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 & 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 (L. 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-hour training session prior to the peer assessment exercice. Prior research found that guidance can enhance the peer assessment process (e.g. Misiejuk & Wasson, 2021; Bloxham & Campbell, 2010; Harland et al., 2017). 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 & 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 & Campbell, 2010). The third is the use of a rebuttal, a text that assessees must write to explain why they found the received feedback relevant or not and justify how they acted upon it (Harland et al., 2018). 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, Segers & Tillema, 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. Aben et al., 2022; Berndt et al., 2018; Bouwer & Dirkx, 2023; Van Meenen et al., 2023; Winstone, Nash, Parker, Rowntree, 2017). 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

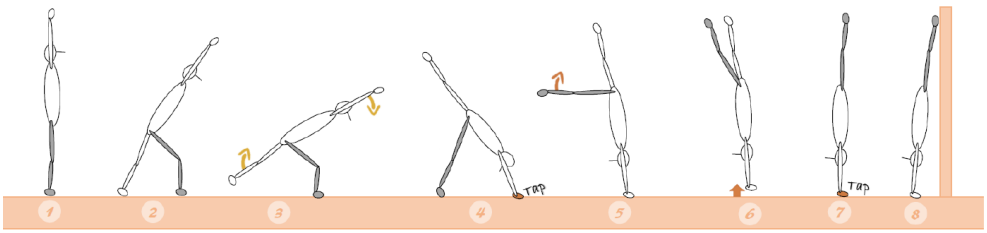
This study aimed to investigate the efficacy of a training session created to enhance students' perceptions of psychological safety and trust before a peer assessment activity. Findings suggested that students' perceptions are related to students' performance improvement, which confirms that it is important 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.

5. Supplementary materials

5.1 Example of a student's work

Learning situation: Front handspring

Handspring against a wall



Instruction : Keep your whole body aligned (except the dominant leg) throughout the movement

Sucess criteria : Hands lift off the ground as the dominant leg rises to the vertical (see 5)

5.2 Students' rubric

Criteria	Unsatisfactory	Satisfactory	Good
Video	The angle of view and/or the poor resolution make it difficult to understand the situation or the movement OR the movement is very badly executed	1) The angle of view and the resolution allow a global understanding of the situation AND of the movement AND 2) the movement is correctly executed	1) The angle of view, the resolution and the quality of technical execution allow an immediate and perfect understanding of the situation and the movement AND 2) The movement is very well executed AND 3) the video is ideally cut: neither too short nor too long.
Diagram	The understanding of the situation or of the movement is not immediate because the material is not clearly identifiable OR the successive positions of the gymnast are not clear (back? front? L or R profile?) OR are unrealistic	At first glance, the situation is clearly identifiable and the positions of the gymnast are globally 1) clear (back, face, profile L/R ?) AND 2) realistic (Dimensions of the segments? Position of the head in relation to the arms?)	The understanding of the situation and the movement is immediate, precise and perfect: -the material is represented in its details - the articular positions are physiological and precise- the dimensions of the segments/material do not vary- the movements are realistic.
Instructions	The instruction is missing OR is incorrectly formulated (French/spelling) OR is impersonal OR the instruction describes the movement but does not specify how to perform it.	The instruction is 1) personal AND 2) qualitative ("when or how to perform the movement to succeed...") AND 3) correctly formulated BUT complex OR not very precise	The instruction is 1) personal AND 2) qualitative ("when or how to perform the movement to succeed...") AND 3) formulated very clearly, simple AND precise
Success criteria	1) The success criterion is missing OR 2) It is incorrectly formulated (French/spelling) OR 3) There is no link between the criterion and success OR 4) The criterion is difficult to observe	1) The success criterion is correctly formulated 2) It is relevant (if this criterion is not met the movement is incorrect) 3) It can be easily observed by a non-expert external observer.	1) Its formulation is correct, clear, simple and precise AND 2) It is relevant (if this criterion is not met, the movement is incorrect) AND 3) It is easily observable by an external observer non-expert and even by the performer himself.
Overall coherence	1) The technique shown on the diagram is clearly incorrect (timing, placement, ... do not match the do not correspond to the video) AND/OR 2) The instructions and/or the criteria for success do not corresponds to the diagram/video	1) The technique shown on the diagram is correct AND 2) The diagram, the instructions and the success criteria are consistent with each other AND 3) Important start OR end positions are missing from the diagram	1) The technique shown in the diagram is correct AND 2) The diagram, the instructions and the criteria for success are consistent with each other AND 3) Important start OAND end positions are represented

5.3 Instrument to report the characteristics of peer assessment design

Instrument to report the characteristics of peer assessment designs

Designed by Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education.* **Use that citation if you include the instrument in your publication**.

TITLE OF THE ARTICLE
Safe and sound: Examining the effect of a training targeting psychological safety and trust in peer assessment
Authors
Morgane Senden, Dominique De Jaeger, & Liesje Coertjens

Our study investigates:

☒ The impact of peer assessment on interpersonal/intrapersonal factors
☐ The impact of interpersonal/intrapersonal factors on peer assessment
☐ The impact of peer assessment on performance/behaviour
☐ The impact of moderator/mediator variables on peer assessment outcomes

Describe the characteristics of your peer assessment study in the table below.

	Category	Description [i]	Our study
Context			
1	**Subject domain**	Subject domain the study was done in (e.g., mathematics, instructional sciences, accounting, etc.)	Physical education (didactics of acrobatic sports)
2	**Place/Time**	Where was the PA conducted? (In class or out of class?)	☒ In class/class time ☒ Out of class/free time
3	**Setting**	Formal or informal education setting?	☒ Formal ☐ Informal
4	**Requirement**	Was PA compulsory or voluntary for assessor/assessee?	☒ Compulsory ☐ Voluntary
5	**Alignment**	Was the PA activity aligned to curriculum, learning goals or teaching?	☒ Yes ☐ No
Instructional design			
6	**Purpose**	What was the assessment purpose of the PA activity? (Formative, summative or both?)	☐ Formative ☐ Summative ☒ Both
7	**Object**	What was assessed? (e.g., written assignment, oral presentation, contribution to group work)	Instruction sheet illustrating a gymnastic exercice
8	**Product/Output**	What was the output of the PA? (e.g., score, written feedback, oral feedback, or a combination)	Score (level in a rubric) and written feedback
9	**Relation to staff assessment**	Was PA done without staff assessment (substitutional) or in addition to staff assessment (supplementary)?	☐ Substitutional ☒ Supplementary
10	**Official weight**	Did participation in the PA activity or the grade given by peer(s) contribute to learners' final grades?	☐ No ☐ Yes-for participation in PA ☐ Yes-for PA grade ☒ Yes-both (PA & participation) ☐ Other: Click here to add text
11	**Reward**	Was there a reward for participation in PA?	☒ No ☐ Yes-course credit ☐ Yes-incentives (e.g., free time, money, etc.) ☐ Other: Click here to add text
12	**Directionality**	Was the learner assessing another without being assessed (unidirectional) or acting as both assessor and assessee (bidirectional)?	☐ Unidirectional ☒ Bidirectional
13	**Degree of interactivity**	How did the assessee demonstrate engagement and response to PA?	☐ Reactive: assessee responds to assessor ☐ Reciprocal: same learners assess each other on same task ☐ Negotiated: PA was done more than once on the same task and both parties negotiated it
14	**Frequency**	How often was the PA of the same task done? (Once, twice, etc.)	Once
15	**Group constellation**	Did members of the same group assess each other (intragroup) or peers from another group (intergroup) or both?	☐ Intragroup ☐ Intergroup ☐ Both
16	**Constellation assessor**	The number of assessors assigned to each assessee	7

17	**Constellation assessee**	The number of assessees per assessor	7
18	**Unit of assessment (assessor)**	At what level did the assessor(s) perform PA? Individual, group, or both?	☒ Individual ☐ Group ☐ Both
19	**Unit of assessment (assessee)**	At what level did the assessee(s) experience the PA? Individual, group, or both?	☒ Individual ☐ Group ☐ Both
20	**Privacy**	Did assessor and assessee know the identity of each other? (Was PA public, single-blind, or double-blind?)	☐ Public ☐ Single-blind (for assessor) ☐ Single-blind (for assessee) ☒ Double blind (anonymous) ☐ Other: Click here to add text
21	**Contact**	Was PA done face-to-face or online? How was the contact between assessor and assessee?	☐ **Face-to-face synchronous:** same time same place (ST SP) ☐ **Online synchronous:** same time different place (ST DP), e.g., skype, chat, etc. ☒ **Online asynchronous:** different time different place (DT DP), e.g., e-mail, SWoRD, etc. ☐ **Other:** Click here to add text
22	**Matching**	How were assessor and assessee matched for the PA activity? (e.g., random, skill, self-select, gender, friendship, performance, etc.)	Random
23	**Format**	How was the PA guided?	☐ Freestyle (no instruction provided to learers) ☐ Guided instruction ☐ Guided prompts ☒ Guided criteria/rubrics ☐ Guided criteria/rubrics and prompts ☐ Other: Click here to add text
24	**Training**	Did the learners receive PA training at any time? If yes, describe the moment in which they received it and the type of training.	They received a 1-hour session to prepare them before the peer assessment (they learn how to give effective feedback according th Hattie & Timperley (2007)' framework and they practice peer assessment)
25	**Revision**	Did learners revise their work after receiving or providing PA?	☐ No ☒ Yes
26	**Scope of involvement**	How were learners involved in the PA activity?	☐ Development of assessment criteria ☐ Gave PA only ☐ Received PA only ☒ Gave and received PA ☐ Additional self-assessment ☐ Other: Click here to add text
Outcomes			
27	**PA Outcomes**	These variables are directly measured as outcomes of the PA activity (i.e., why was the PA activity conducted?) Select all the options that apply to your study from the right column.	☐ **Beliefs & perceptions:** including perceptions of learning capacity to perform PA or any perceptions related to the PA processes (e.g., fairness, usefulness), metacognition and self-regulation, attitudes and beliefs (e.g., self-efficacy), teachers' perceptions/conceptions. ☐ **Emotions and motivation:** emotions experienced by learners (e.g., achievement emotions, social emotions, etc.) & all motivational beliefs (e.g., learning motivation). ☒ **Performance:** academic/domain specific performance, achievement, improved draft/work (i.e., revision). ☐ **Skills:** quality of contribution to the group, professional behaviour, problem solving skills, work habits, interpersonal skills, metacognitive & self-regulatory skills. ☐ **Reliability of PA:** consistency of PA scores compared to other peer assessors, or PA over several rounds. ☐ **Validity of PA:** accuracy of PA compared to teachers/tutor/expert's assessment. ☒ **PF content:** characteristics of the feedback messages including type, focus, quality, frequency of comments/posts (i.e., participation), etc. ☐ **PF processing:** Includes implementation, reactions to PF, seeking PF, coping with PF, etc. ☐ **Other:** Click here to add text
Moderators/mediators			
28	**PA Moderators/mediators**	Variables that are not usually manipulated but are taken into account when investigating PA. Select the variables that have been explored in your study from the right column.	☐ **Gender:** of assessor/assessee. ☐ **Ability & Skills:** includes prior knowledge, prior performance, achievement level, GPA, finished high school, previous level of education, year of enrolment, etc. ☐ **Skills:** reviewing ability, computer skills, etc. ☐ **Age:** of assessor/assessee. ☐ **Culture, ethnicity, nationality or race** ☐ **Other:** Click here to add text
This tool is based on: Alqassab, M., Strijbos, J., Panadero, E., Fernández Ruiz, J., Warren, M., & To, J. (2023). A systematic review of peer assessment design elements. *Educational Psychology Review*.			

Note. PA = peer assessment; PF = peer feedback

5.4 Training description

The training lasted one hour and was composed of three steps: (1) discovery of students' thoughts about peer assessment, (2) role-plays and discussions in small groups, and (3) summary of key learning points.

1. Discovery of students' thoughts about peer assessment (5 min)

The week before the training session, students were asked to answer four questions through an interactive platform (Wooclap): (1) In one word, what does peer assessment mean to you? (2) What are the advantages of peer assessment? (3) What are the disadvantages of peer assessment? (4) Do you have any concerns related to peer assessment? For the first three questions, students' answers were shown live as a words cloud and students could *like* them. Figure 1 is an example of a words-cloud generated for the third question (disadvantages of peer assessment).

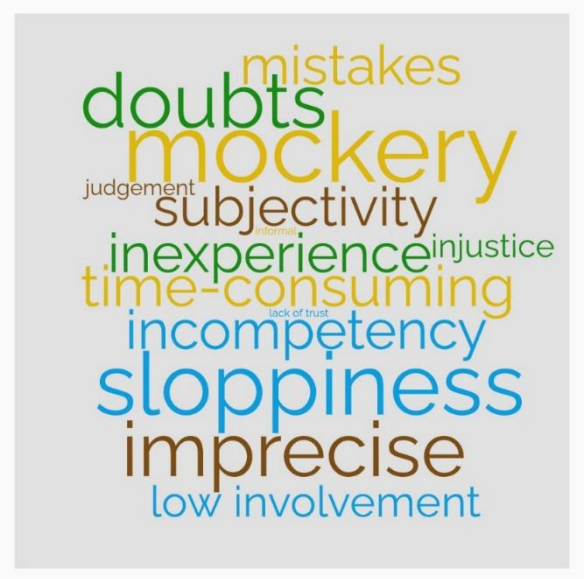


Figure 1. Words cloud for disadvantages of peer assessment

2. Role-plays and discussions in small groups (50 min)

Based on students' answers to the Wooclap, we designed two 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 and discuss them. The first role-play consisted of three friends who participated in a peer feedback activity but were all dissatisfied with the received written comments for different reasons (e.g. the feedback was only positive, without any suggestion for improvement). In the second role-play, students had to put themselves in the shoes of three students who had to decide what grade and feedback to give to peers who did a poor oral presentation.

For each role-play three students acted out the roles while the three others observed the role-play and took notes to inform the following discussion. Having two role-plays allowed each student to play one role, either in the first or second role-play. After performing and

discussing the two role-plays, they stayed in sub-groups to synthesize their discussions. More precisely, we asked them to identify the benefits and interpersonal risks of peer feedback, and what the professor and themselves as students can do to ensure that a peer feedback activity works well.

Role-play example

Léa, David, and Chloé are third-year psychology students. For one of their courses, they had to write a research paper. They had to develop a problematic, research question and hypotheses, then imagine a methodology to test them. They then participated in an anonymous peer assessment which was intended to give them some ideas for improvement before they handed in the final version of their work (evaluated by their teacher). Lea, David, and Chloe meet after classes for a drink and take the opportunity to complain about the feedback they received. Indeed, none of them is satisfied with it. Léa is the first to complain: "I only received very superficial feedback, always positive. How am I supposed to learn anything from this?" David continues: "It's the opposite for me, the guy who assessed me destroyed me. If I consider his feedback, it means that I have to start from scratch". Chloe is no more satisfied than he is: "All my assessor did was spot my spelling mistakes. That's not all that counts!"

Roles:

Lea: You are very disappointed. You feel that the student who gave you feedback was not qualified enough to assess your work properly and accurately. You only received feedback like, "You have a lot of references in your bibliography. Great!" which is not very helpful to improve your work. Fortunately, you were able to think of ways to improve yourself by realizing that some of the suggestions made to the student you assessed also applied to your work. Otherwise, you would have wasted your time participating in this peer review. You envy David who received some real suggestions for improvement. You initiate the role-play by explaining why you don't like the feedback you received.

David: You are upset and a little offended. Most of the comments you receive seem useful (e.g., "Be careful, when you put references at the end of a paragraph, it's not always clear which idea in the paragraph came from which reference."). But you have a hard time digesting this one: "It seems to me that the methodology you suggest does not allow you to answer your research question. To answer it, you should observe your participants instead of giving them questionnaires". This comment calls into question the core of your work, of which you were quite proud. You envy Chloe, who has received only surface comments that do not question the value of all her work and are easy to act upon.

Chloe: You are annoyed by the comments you received. The student who read your work just corrected the spelling mistakes (there were quite a few because you were so busy not proofreading). The comments you received are not very helpful and you think that the student who assessed you must have thought you were completely stupid. You envy Lea, who received a lot of compliments.

3. Summary of key learning point (10 min)

The last training stage was an open discussion to synthesize all the ideas in the large group. We asked a student from each group to report the key points of their discussions and took live notes on our slideshow so students could see how the discussion progressed. We used this moment to explain and justify the choices made by the professor for the organisation of the peer feedback activity and to link these choices to the elements discussed by students and with the literature on peer feedback. We also used this moment to address any remaining concerns. Based on the discussion, we created a mind map (see figure 2) that students received a few days later. This mind map allows students to keep a record of the key ideas identified together during the training in a visual and accessible format.

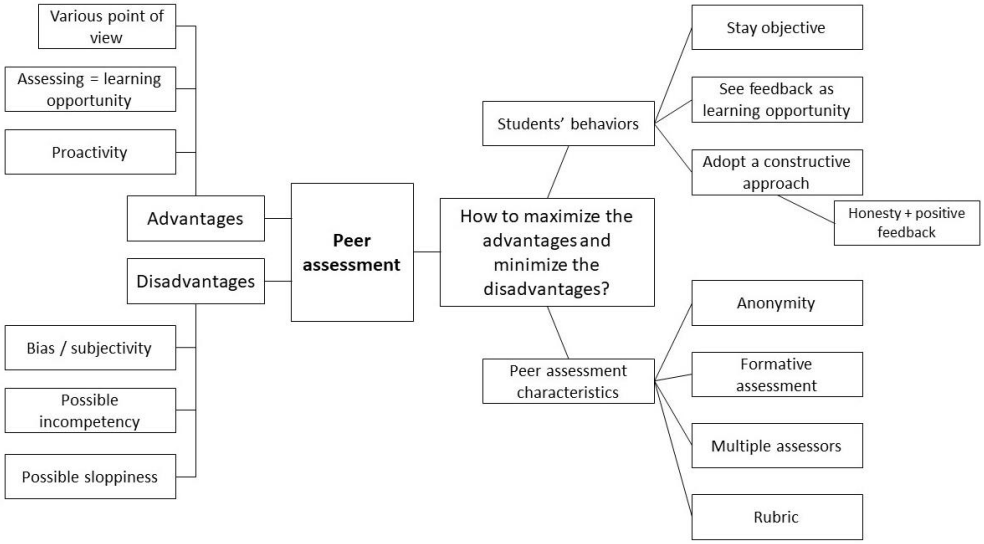


Figure 2. Mind map

Chapter 5

Fostering higher education students' perceptions of trust and participation in peer feedback

Investigating the efficacy of an intervention combining training and backward evaluation

Peer feedback is an effective way to foster higher education students' learning, but only if students are actively engaged in it. Yet, students often express resistance to peer feedback. One reason for this resistance is that students may not trust that they and their peers are able to provide useful feedback. To enhance these feelings of trust in the self and in the peer as an assessor, we conducted a quasi-experimental mixed method study with an explanatory sequential design in a first-year master course in educational sciences. An intervention was created composed of two elements: (1) a training session targeting trust in peer feedback and (2) a backward evaluation. Students in the experimental group (N=95) received the intervention, while students in the control condition (N=94) participated in the peer feedback activity without receiving the intervention. Quantitative results revealed that students positively reacted to the training and that the intervention positively affected their rate of participation in peer feedback. However, contrary to our hypotheses, no significant impact of the intervention was found on students' perception of trust in the self and in the peer as an assessor. Follow-up interviews were conducted with sixteen students to better understand these quantitative results. The qualitative results suggest that the intervention may have affected the meaning that students attributed to peer feedback and their feeling of legitimacy to provide feedback to peers.

Keywords

Peer feedback, trust in the self, trust in the other, intervention, training, backward evaluation, higher education, mixed-method

Fostering higher education students' perceptions of trust and participation in peer feedback

Investigating the efficacy of an intervention combining training and backward evaluation

1. Introduction

The important role that feedback plays in the development of students' learning has long been recognized (e.g., P. Black & Wiliam, 1998; Sadler, 1989). Given that in today's higher education teaching staff rarely have the time to provide feedback to students, alternatives are needed so that students can still benefit from feedback (Huisman et al., 2019). Peer feedback is a promising alternative with multiple advantages. First, peer feedback is even more efficient than teacher feedback to foster students' learning (Double et al., 2020; Wisniewski et al., 2020). Second, peer feedback can contribute to the development of students' feedback literacy skills, which are essential skills for students to benefit from any feedback (Carless & Boud, 2018). Third, through peer feedback, students can also learn skills needed to assess themselves, and thus to become lifelong independent learners (Tai et al., 2018). Therefore, peer feedback is a useful tool to enhance both immediate and future learning.

To fully reap the benefits of peer feedback, students' active involvement is indispensable. Yet, students are often initially reluctant to participate in peer feedback and show little active engagement (e.g., Al Qunayeer, 2020; Poverjuc et al., 2012). According to students themselves, one explanation for their resistance to peer feedback is that they lack the confidence to provide feedback to their peers and doubt the relevance of their peers' feedback (Al Qunayeer, 2020; Poverjuc et al., 2012). Therefore, it seems that to be engaged in peer feedback students need to have sufficient levels of trust in their abilities and their peers' abilities to provide feedback, which is not always the case (e.g., L. Li & Steckelberg, 2013; Wilson et al., 2015). Trust in the self as an assessor can be defined as a "belief about the ability to perform peer assessment as an assessor" and trust in the other as an assessor as the "confidence in a peer's capability to perform a fair and/or accurate peer assessment" (Panadero et al., 2023, p. 1057).

Although it is known that trust in the self and trust in the other as an assessor play important roles in peer feedback (Panadero et al., 2023), there is a lack of research investigating how to ensure high levels of trust so that students can benefit as much as possible from peer feedback. The current study aims to fill this knowledge gap by developing an intervention consisting of a training session targeting trust and a backward evaluation. Students' reactions to the training session were measured, as well as the impact of the intervention on students' attitudes and behaviours The measurement of students' reactions,

attitudinal change, and behavioural change is a strength of the current study as they are all necessary to draw a complete picture of the impact of the intervention (Blanchard & Thacker, 2012). Participants' reactions to an intervention are important to measure as they will affect their motivation to learn from it and can sometimes explain the ineffectiveness of an intervention. However, an intervention may be well received without having a positive impact on the knowledge, skills or attitudes it was designed to address, so these are also important to measure. Finally, since the ultimate goal of the intervention is to have a significant impact on student behaviours, these are also crucial to measure (Blanchard & Thacker, 2012).

In the next sections, we will review findings on trust in the self and in the other as an assessor in peer feedback, and argue that a combination of training and backward evaluation could be an effective way to increase trust.

1.1 Trust in peer feedback

The role that both trust in the self and trust in the other plays in peer feedback has been investigated in three studies. Trust in the self and in the other have been found to be positively correlated with students' conceptions of peer assessment and perceived learning (van Gennip, Segers & Tillema, 2010), as well as with the perceived educational value of peer assessment (Rotsaert et al., 2017). It was also found that students with high levels of trust have positive approaches to and conceptions of learning during peer feedback (K. Cheng & Tsai, 2012). High levels of trust in the self and in the other were associated with a tendency to approach peer assessment with internal motivations and sophisticated learning strategies (e.g., reflecting on the meaning of what is learned), while levels of trust in the self, but not in the other, were associated with an understanding of how peer assessment can help foster learning. Even though these relationships are only correlational, they suggest that trust plays an important role in peer feedback. This raises the question of how to ensure that students trust their own and their peers' abilities to provide feedback.

Several studies have addressed the question of how to ensure trust in peer feedback activities. Anonymity has been proposed to ensure trust, but either did not seem to have an impact on it (Rotsaert et al., 2018), or, when an absence of contact accompanied anonymity, was associated with lower levels of trust (S. Kim, 2019). The use of role-play during peer feedback has also been tested as a way to ensure trust (Ching & Hsu, 2016). When providing feedback to their peers' assignments (a case analysis of a complex instructional design problem), students had to play the role of a self-selected stakeholder and provide feedback from the stakeholder perspective instead of their own perspective. Students reported high levels of trust, but in the absence of a pre-test and a control group, it is not possible to know if these levels of trust are due to the intervention. Moreover, this intervention has the limitation of not being suited for every assignment (e.g., it would not make sense to adopt a stakeholder perspective to provide feedback on mathematical problem-solving). Therefore, while role-plays could have merits, other interventions more suitable for practice are necessary to develop to increase students' trust in the self and in the other as an assessor.

1.2 Training and backward evaluation as an intervention

It is well-established that training students before peer assessment is an essential step. The presence of training was found to be the most critical moderator of the impact of peer assessment on learning in a meta-analysis on the subject (H. Li et al., 2020). Moreover, according to a literature review, the presence of a training session before peer assessment is associated with a higher inter-rater agreement between students, feedback of higher quality, greater performance improvement, and more positive attitudes toward peer assessment (van Zundert et al., 2010).

The impact of training on higher education students' perceptions of trust more specifically has been investigated in two studies (Ocampo et al., 2023; Parikh & Sheehan, 2016), with some positive results. In Parikh and Sheehan's (2016) study, students reported high levels of trust in their and their peers' abilities after receiving a 30-minute training on how to provide feedback. However, it is not possible to conclude that the training had a positive impact on trust given the weak research design used: there was no control group nor a pre-test. In a more rigorous study, Ocampo et al. (2023) investigated the impact of a 180-minute training on how to provide feedback with the help of a rubric on students' trust in the self as an assessor. They found that male trained assessors trusted themselves more than male untrained assessors, but only when they were providing feedback to women (the effect was not found with female assessors, or when male assessors provided feedback to men). Therefore, despite all the benefits of training (van Zundert et al., 2010), it is not possible to affirm that a training session on how to provide peer feedback is enough to strongly affect students' perceptions of trust.

Training can also be designed to target specific attitudes linked to the social nature of peer feedback, which, to our knowledge, has only been done by two studies. While not specifically on trust, the first study (L. Li, 2017) investigated the impact of a training session specifically designed to address the pressure students can feel from their peers during peer feedback. Students who received the training felt significantly less peer pressure and saw significantly more value in peer feedback than students who did not receive it. It seems therefore that specific training can be an effective way to address students' concerns linked with peer feedback. However, this positive effect was not found in the second study (Senden et al., 2023), which focused on trust and psychological safety. In this study, no effect of the training was found, neither on students' attitudes (their perceptions of trust in the self and in the other and of psychological safety) nor on students' behaviours (their performance and the quality of the feedback they provided). However, an important limitation of this study, which adopted a two-cohort pre-and post-test design, was that the two cohorts of students were not comparable at the pre-test, probably due to the COVID-19 pandemic (the control group took place in 2019-2020, and the experimental group in 2021-2022).

Even though a training session seems to be a promising way to address students' concerns (L. Li, 2017), it may not be sufficient by itself to ensure trust (Senden et al., 2023). An important limitation of training is that it only equips students before the peer feedback but

cannot provide support during the actual tasks (Lee, 2015). Moreover, as higher education schedules tend to be very busy, training sessions are often limited to a few hours (e.g., L. Li, 2017; Ocampo et al., 2023; Senden et al., 2023). This means that there is no time to combine various methods, a key element to enhance training effectiveness (Martin et al., 2014). A way to overcome these limitations could be to combine a training session with other activities that share the same objectives but can be organised during peer feedback and outside class time.

Another possible way to increase students' perceptions of trust could be by increasing the possibility of interactions between students. A study found that students tend to have more trust in their peers' abilities to provide feedback when they have the opportunity to dialogue with them (S. Kim & Lan, 2021). Dialogue around feedback is recognised as a way to increase the effectiveness of feedback, whether the feedback comes from teachers (Nicol & Macfarlane-Dick, 2006) or peers (Zhu & Carless, 2018). Organising oral exchanges between students is impossible in the case of online anonymous peer feedback. However, other forms of interactions are possible. Offering students the possibility to react to the received feedback through backward evaluation is a first step towards dialogic feedback (Zhao et al., 2024).

Backward evaluation is a way to enhance peer feedback by adding a supplementary step: the feedback receiver provides feedback to the feedback provider on the quality of the received feedback, for example by indicating how useful the feedback was (Misiejuk & Wasson, 2021). To the best of our knowledge, there is an absence of studies investigating its impact on trust specifically, and only two studies investigating its impact on other attitudes (M. Kim, 2009; Ng & Yu, 2023). In a qualitative study, students reported that backward evaluation incited them to be more respectful and serious during peer feedback (Ng & Yu, 2023). Moreover, a quasi-experimental found backward evaluation to be associated with a larger performance improvement, more motivation to participate in peer feedback, and more metacognitive awareness of the learning process (M. Kim, 2009).

Backward evaluations could positively affect both trust in the other and in the self. Concerning trust in the self, backward evaluation could help students adequately assess their skills as feedback providers and help them know how to improve if needed. If they received positive feedback on their feedback, they could realize that they already have good abilities to provide feedback, which should increase their trust in the self. On the contrary, if they received feedback suggesting they still have to improve, they should also receive suggestions from their peers, which could help them to improve and, therefore, increase their trust in the self.

Concerning trust in the other, providing backward evaluation forces students to actively process the feedback they receive to be able to critically appraise it, which is important considering that some students do not even read the feedback they receive (Orsmond et al., 2005; Sinclair & Cleland, 2007). This could push students to see the value of feedback from peers for their learning, and therefore reinforce the trust they have in their peers. Moreover, if the possible positive impact of backward evaluation on students' respectful and serious behaviour (Ng & Yu, 2023) is reflected in feedback of higher quality, this could also positively affect trust in the other.

1.3 The present study

This study aims to examine if trust in the self and in the peer as an assessor can be fostered by an intervention composed of two elements: (1) the provision of a training session that aims to improve trust and (2) the presence of a backward evaluation. To this aim, the current study adopts a quasi-experimental mixed-method study with an explanatory sequential design (Creswell, 2012): the quantitative data were analysed first, and the qualitative analyses were conducted afterwards to better understand the quantitative results.

In the first phase, to gain a comprehensive understanding of the intervention's overall impact, we quantitively investigated several outcomes of the intervention. The students' reactions to the training were measured, as were the attitudinal and behavioural changes caused by the intervention as a whole. As it is established that providing students with a general training session is a requirement for the efficacy of the peer feedback-process (van Zundert et al., 2010), the students in the control group also received a 1-hour training session on how to provide peer feedback. Therefore, experimental students' reactions to the 2-hour comprehensive training session (1-hour on how to provide peer feedback and 1-hour targeting trust) were compared to control students' reactions to the 1-hour training session on how to provide peer feedback. Four research questions and six hypotheses were formulated:

RQ1 (reaction to the training): Is the 2-hour training session (how to provide peer feedback and trust) provided to students in the experimental condition perceived more positively than a general 1-hour training (how to provide peer feedback) provided to students in the control condition?

Hypothesis 1a: Students in the experimental condition have more positive impressions of the training than students in the control condition.

Hypothesis 1b: Students in the experimental condition report more utility and application to practice of the training than students in the control condition.

RQ2 (attitudinal changes): Does the intervention (training and backward evaluation) have a positive impact on students' perceptions of trust in the self and in the other as an assessor?

Hypothesis 2a: Perceptions of trust in the self as an assessor increase more in the experimental condition than in the control condition.

Hypothesis 2b: Perceptions of trust in the other as an assessor increase more in the experimental condition than in the control condition.

RQ3 (behavioural changes): Does the intervention (training and backward evaluation) have a positive impact on students' participation in subsequent peer feedback activities?

Hypothesis 3a: Students in the experimental condition participate more in the third peer feedback activity than students in the control condition.

Hypothesis 3b: Students in the experimental condition participate more in the fourth peer feedback activity than students in the control condition.

In the second phase, the follow-up interviews were analysed to deepen our understanding of specific results from the quantitative analyses and, therefore, to answer the following research question:

RQ4: What explanations do the follow-up interviews provide for the quantitative results?

2. Method

2.1 Participants

Participants were students following a course on the psychology of learning (5 ECTS) at a French-speaking Belgian university. This course is mandatory for students enrolled in a master in educational sciences, a master designed to allow students to combine their studies with a part-time or full-time job (often as teachers or educators). One hundred eighty-nine students agreed to participate in the study (72% of students enrolled in the course). Most participants were women (71%) and they were between 21 and 58 years old (M = 31). In one of their prior courses, students participated in peer assessment, so they already had experience with it. However, they did not received any training prior to this peer assessment.

To investigate the effect of the intervention, students were assigned to a control condition (N = 94) and an experimental condition (N = 95). We took advantage of the fact that students were divided into eight smaller groups (of approximately 25 students) for two small-group practice sessions of 180 minutes provided by the teaching assistants. Hence, we randomly divided these smaller groups between the two conditions. The allocation to the control or experimental conditions was conducted at the group level to reduce the risks of contamination between the groups (students self-register to the groups, usually in the same group as their friends). Students in both conditions were comparable in terms of age ($t(70) = -1.48$, $p > .05$) and gender ($\chi^2(1) = 2.28$, $p > .05$).

2.2 Description of the course's activities

After each chapter of the psychology of learning course, students had to submit an assignment: a case study based on the theoretical framework they studied in this chapter (e.g., socio-constructivism). Throughout the course, students submitted four assignments that were neither graded nor read by the course professors, but that were collectively discussed with the professor during the following course. They aimed to provide opportunities for students to prepare for the exam, which also consisted of case analyses, and therefore to ensure the constructive alignment of the course (Biggs, 1996).

A peer feedback activity followed each assignment. These activities were organised through the workshop module on Moodle. Students submitted their assignments as a Word document so their peers could directly add comments to it, alongside more global feedback

written in Moodle. Each student provided feedback to three randomly assigned peers. Both the identities of the assessor and assessee were concealed. To guide them, the rubric that the professors will use to grade the exams was available. This rubric was discussed with the students during the course preceding the first peer feedback activity. A description of all design elements of the peer feedback activities using Panadero et al.'s (2023) reporting instrument is available in Appendix A.

To prepare them for the peer feedback, all students had to participate in a 1-hour training session, which aimed to increase their abilities to provide high-quality peer feedback. This session was integrated into a more comprehensive training session for students in the experimental condition, while it was provided as a standalone training session for students in the control condition. The training was provided by teaching assistants when students were in smaller groups and was composed of two parts. First, a lecture was given to explain to students how to provide effective feedback based on Hattie and Timperley's (2007) feedback model. Second, students had the opportunity to practice: in pairs, they had to provide feedback on a peer's assignment. This practice was followed by a discussion that allowed students to check with the teaching assistants if the feedback they wrote was relevant, to see how they could still improve it, and to discuss any remaining questions.

2.3 Intervention

To tailor the intervention to students' needs, in order to enhance its effectiveness (Dusenberry & Robinson, 2020), we asked students about their thoughts on peer feedback. During the first course, we asked all students, through an interactive platform, what they saw as the advantages and disadvantages of peer feedback, and if they had any concerns related to the use of peer feedback. This revealed that these students seem to be very resistant to participating in peer feedback. The more frequent answer concerning the advantages of peer feedback was "There are no advantages". Moreover, students saw a lot of disadvantages and expressed numerous concerns such as a heavy workload, a lack of skills from their peers that could cause misleading feedback, or a lack of engagement of their peers in the activity.

The intervention was composed of two parts: (1) a second hour of training focused on trust in peer feedback, and (2) a backward evaluation. The second hour of training was designed to address students' concerns related to the social nature of peer feedback, with a focus on the question of trust. More specifically, the goals of this training were to (1) raise students' awareness of the aims and benefits of peer feedback, (2) increase students' trust in their abilities to provide peer feedback, and (3) increase students' trust in their peers' abilities to provide peer feedback. The training mainly used active methods, which are recommended to transform attitudes (Blanchard & Thacker, 2012). First, the training started with a short interactive lecture provided by the teaching assistants. They explained why peer feedback activities were implemented in this course, linking this with the research on the benefits of peer feedback, they recognized students' concerns with the use of peer feedback, and they explained the aim of the training. Then, in groups of five or six, students participated in two role-plays. Students received a detailed document with two role-play situations, the description of the role they were invited to play, and three questions to launch discussions

after each role-play. During the role-play, they moved between the groups to answer any questions. The role-plays were created following recommendations by Bolinger and Stanton (2020). The first situation was about three students who could choose to participate or not in peer feedback, while the second was about a teaching team that had to decide if they would implement peer feedback in their course. We felt confident asking students to adopt these two perspectives because most of them are teachers. For each role-play, one person was in favour of peer feedback, one person was not, and one person was hesitant. The two or three other students observed the role-play and took notes. Finally, after the role-plays and small-group discussions, students synthesized their discussion with the other students and the teaching assistant. Together they synthesized what themselves as students and the teaching team could do to maximize the advantages and minimize the risks of peer feedback. The discussion was also an occasion to explain the choices made in the design of the peer feedback activities (e.g., the fact that it was purely formative, that it was anonymous, that students received feedback from three peers, etc.).

After the second peer feedback activity, students from the experimental condition participated in a backward evaluation. The procedure was based on the backward evaluation conducted in previous studies (Misiejuk et al., 2021; Wu & Schunn, 2020). The students back-evaluated each of the peer feedback they received by (1) rating the helpfulness of the feedback on a 5-point scale (ranging from *Not useful at all* to *Extremely useful*), (2) providing a comment explaining their helpfulness rating, and (3) selecting improvement suggestions from a list (*Feedback could be formulated more sympathetically*, *Feedback could be more justified/argued*, *Feedback could include more suggestions for improvement*, *Feedback could be more focused on the most important/relevant aspects*, *Feedback could be more specific, pointing out exactly what needs to be changed*) and/or writing their own suggestion.

2.4 Study procedure

Students in both conditions participated in the same peer feedback activities, at the same time (see Figure 1). To be able to investigate the intervention's effect while remaining ethical, the training was introduced at different times for students in the control and experimental conditions. When students in the experimental condition received the comprehensive 2-hour training session, students in the control condition only received the 1-hour training session on how to provide effective peer feedback. They received the second hour of the training after the last questionnaire. Another difference between the two groups was that only students from the experimental condition participated in a backward evaluation after the second peer feedback activity. To investigate the impact of the intervention, students from both conditions answered a questionnaire on three occasions. Moreover, sixteen semi-structured interviews were conducted at the end of the course.

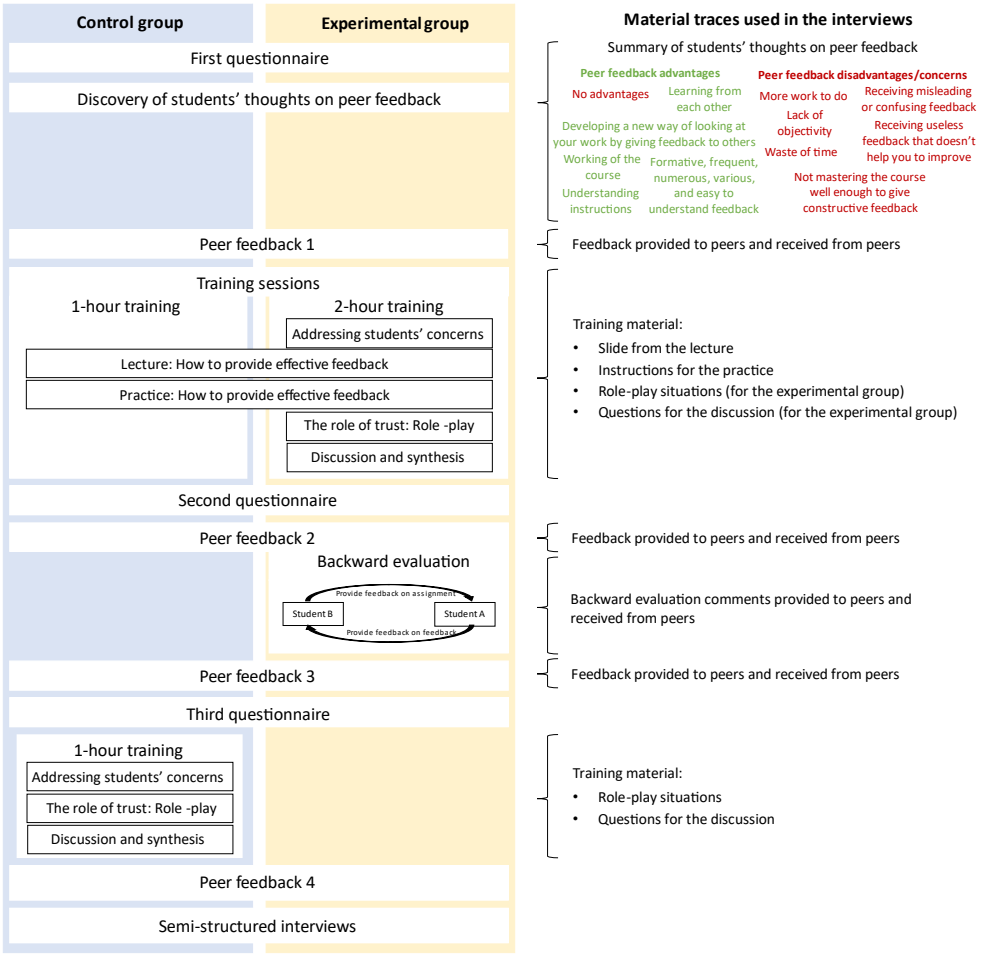


Figure 1. Study timeline and material traces used during the semi-directed interviews

2.5 Quantitative phase

2.5.1 Measurements

To investigate the impact of the intervention on students' perceptions of trust, these perceptions were measured at the first, second, and third questionnaires (see Appendix B). Trust in the self as an assessor and trust in the other as an assessor were measured with items developed by Rotsaert et al. (2018) translated into French. Items were measured using a seven-point Likert scale ranging from 1 (totally disagree) to 7 (totally agree). We only used the items referring to peer feedback, excluding those referring to peer grading. Moreover, given that the reliability of the scale measuring trust in the self as an assessor was low when including the three items, we excluded a problematic item ("I trust that the feedback my peers provide me is based on my work"). Macdonald's ω were used to check the reliability of the

3-item scale measuring trust in the self and Spearman-Brown coefficients were used to check the reliability of the 2-item scale measuring trust in the other. The reliability of trust in the self was excellent, while that of trust in the other was still slightly insufficient at two measurement times (see Table 1).

To verify whether the students positively reacted to the training, their perceptions of the training session were also measured. Items from Grohmann and Kauffeld's (2013) and Holgado-Tello et al.'s (2006) were used. The items were adapted to the context of this study (e.g. "In my everyday work" was replaced by "In the peer feedback activity") and translated into French. The second questionnaire that students filled in right after the training included items measuring students' impressions of the training (factor 1). These items were adapted from Holgado-Tello et al.'s (2006) training satisfaction rating scale, which measures participants' general impression of a training session (e.g., "The method used enabled us to participate actively in the training session") and from two subscales of Grohmann and Kauffeld's (2013) Questionnaire for professional training evaluation measuring satisfaction with a training session (e.g., "I really enjoyed the training session") and perceived learning (e.g., "I learn a lot of new things during this training session"). The third questionnaire, which was filled out by students after they had the opportunity to apply the training to two peer feedback activities, included items measuring the training's perceived utility and application to practice (factor 2). These items were adapted from two other sub-scales of Grohmann and Kauffeld's (2013) questionnaire for professional training evaluation: perceived utility (e.g., "The training session was very useful in helping me carry out peer feedback") and application to practice (e.g., "During peer feedback activities, I used the knowledge acquired during this training session"). A confirmatory factor analysis confirmed this structure into two factors. Although the chi-square is significant ($\chi^2(43) = 107$, $p < .001$), the other indices suggest an acceptable fit (CFI = .95, TLI = .92, RMSEA = .06, SRMR = .10). Macdonald's ω suggest good reliability for both subscales (see Table 1). At the end of the third questionnaire, students had the opportunity to leave a comment, which 41 students did.

2.5.2 Quantitative analyses

Although participation in the training sessions and the peer feedback activities were mandatory, an important part of the students did not participate in them and, consequently, could not be included in the analyses. Students were included in the quantitative analyses if they (1) participated in the training sessions and (2) participated in the second peer feedback activity (and the backward evaluation for students in the experimental condition). Forty-eight participants from the control condition and 46 participants from the experimental condition were included in the analyses. There was no difference in terms of age ($t(136) = 0.95$, $p > .05$) and gender ($\chi^2(1) = 0.69$, $p > .05$) between the participants included in the analyses and those excluded from them.

Table 1. Reliability of survey scales

Scale	Adapted from	Example item	Number of items	McDonald's ω or Spearman-Brown coefficient					
				Control group			Experimental group		
				T1	T2	T3	T1	T2	T3
Trust in the self as an assessor	Rotsaert et al. (2018)	I feel I am able to assess my peers	4	.97	.94	.94	.96	.94	.96
Trust in the other as an assessor	Rotsaert et al. (2018)	I think my peers are able to provide me objective feedback	2	.71	.79	.77	.61	.58	.79
Impressions of the training	Holgado-Tello et al. (2006) and Grohmann & Kauffeld (2013)	The training merits a good overall rating	8	/	.92	/	/	.94	/
Perceived impact of the training	Holgado-Tello et al. (2006) and Grohmann & Kauffeld (2013)	I successfully managed to apply the training contents	4	/	/	.95	/	/	.94

There were missing data for trust in the self and trust in the other: 22% of missing data on the first questionnaire, 11% on the second, and 27% on the third. The missing data were handled through the use of maximum likelihood, a recommended method when data are missing at random (i.e., the probability of missingness is related to the variables in the study), as this method produces less biased parameter estimates than traditional methods such as listwise deletion and fosters statistical power (Enders, 2010).

Two Mann-Whitney's U tests were conducted to investigate if the 1-hour and 2-hour training sessions were differently perceived by students (impression of the training and perceived utility and application to practice of the training). Two linear mixed models were conducted to investigate whether the evolution of students' levels of trust (in the self and in the other) differed between the conditions. To evaluate whether there was a link between the conditions and participation in the third and fourth peer feedback activities, two χ^2 tests were conducted. All analyses were conducted on Jamovi (version 2.3.28).

2.6 Qualitative phase

2.6.1 Participants selection

Sixteen students were selected from the whole sample of students (*N*=189) to participate in the semi-structured interviews based on two purposeful sampling methods. First, we adopted a maximal variation sampling strategy, which means that participants differing in key characteristics were sampled to present multiple perspectives (Creswell, 2012). Thirteen participants were selected to represent students' variation on gender, age, level of trust in the self and trust in the other at the pre-test, and, for students in the experimental condition, on the usefulness of the provided feedback as judged in the backward evaluation (see Table 2). Second, three participants (Colas, Aaron, and Elise) were selected through an extreme case sampling strategy, which means that participants with extreme characteristics were selected (Creswell, 2012). In the open question of the last questionnaire, these students wrote comments that reflected strong feelings about peer feedback (e.g., Colas wrote "In my opinion, I'm incapable of giving quality feedback. For me, it's an exercise that only brings me anxiety and stress").

Students' characteristics are presented in Table 2. Pseudonyms were used to ensure students' confidentiality. The first letter of the pseudonym indicates if students come from the experimental condition ("E"), the control condition ("C"), or were supposed to be in the experimental condition, but were absent during the training session ("A"). We mostly selected students from the experimental condition because the interviews aimed to better understand students' reactions to the intervention. The selected students were contacted by email, and all sixteen agreed to participate.

Table 2. Participants in the semi-structured interviews

Pseudo (gender)	Age	Trust in the self (TS) at T1[1]	Trust in the other (TO) at T1[1]	Range of the peers' judgment of feedback usefulness according to the backward evaluation[2]
Participants selected with the maximal variation sampling strategy				
Achille (M)	23	1.25 (low)	2.33 (low)	Not useful at all – not very useful (2)
Amelia (F)	40	2.5 (low)	5.67 (high)	Not very useful – rather useful (2)
Callie (F)	22	3.5 (medium)	5.67 (high)	/
Célia (F)	32	1 (low)	3.33 (medium)	/
Clémentine (F)	39	7 (high)	5 (medium)	/
Eden (M)	23	2.75 (low)	6 (high)	No backward evaluation received (0)
Eleanor (F)	35	1.25 (low)	2.67 (low)	Very useful (1)
Elena (F)	22	2.25 (low)	4.67 (medium)	Not very useful – very useful (2)
Eliot (M)	41	1.5 (low)	4.33 (medium)	Rather useful – extremely useful (3)
Elmir (M)	37	6.25 (high)	6.33 (high)	Rather useful – very useful (3)
Emiliano (M)	53	5.75 (high)	4.67 (medium)	Very useful – extremely useful (2)
Erin (F)	43	/	/	Not useful at all – not very useful (2)
Eva (F)	21	5 (medium)	5.67 (high)	Not useful at all – not very useful (2)
Participants selected with the extreme case sampling strategy				
Aaron (M)	30	2.75 (low)	2.25 (low)	Did not participate in the backward evaluation
Colas (M)	27	/	/	/
Elise (F)	27	2.75 (low)	4.67 (medium)	Not very useful (1)

[1] Low if the value is less than 3/7 (29% of students for TS, 11% for TO), medium if the value is between 3 and 5 (49% of students for TS, 64% for TO), and high if the value is more than 5/7 (22% of students for TS, 25% for TO).
[2] Each student was supposed to receive a backward evaluation from three students, however, not all their peers participated in it. In brackets is the number of students who provided backward evaluation.

2.6.2 Data collection and analyses

Although the quantitative results confirmed the positive impact of the intervention, they did not allow us to understand how the intervention affected students' participation in subsequent peer feedback activities. Therefore, the aim of the follow-up interviews was twofold: (1) to discover how participants explained their participation or lack of participation in peer feedback and (2) to understand how the intervention impacted this participation.

The interviews were carried out by the first author, either at her office (for five students) or online through the Microsoft Teams platform (for eleven students), depending on the participants' preference. Most of the interviews were conducted just before the winter break, after all course activities, but before the exam session. Three were conducted during the

January exam session (Erin, Emiliano, and Eva) because students were not available before. The interviews lasted between 27 and 110 minutes, with an average of 61 minutes.

The interviews were semi-directed and used material traces of the activities, which consisted of traces left by the participants' behavior (Theureau, 2010). The material traces were (1) the assignments submitted by students and the feedback received on them for each activity, (2) their peers' assignments and the feedback they provided to their peers for each activity, (3) the material used during the training sessions (4) a summary of students' views on the advantages and disadvantages of peer feedback according to what they submitted through an interactive platform during the first course, and, for students in the experimental group, the comments they provided and received during the backward evaluation (see Figure 1). These material traces allowed students to better remember their experiences (Theureau, 2010).

Three categories of questions were asked: (1) questions about themselves and their previous experiences with peer feedback, (2) questions about their experiences of the different activities (each peer feedback activity, the training session, and the backward evaluation) to understand what was important to them and how their perspectives evolved, and (3) questions about their current perceptions of peer feedback. Moreover, if students did not talk about certain aspects spontaneously (e.g., the anonymity of peer feedback), we asked questions about these aspects at the end of the interview. Interviews were recorded and transcribed. The interview guide and the material traces used with one of the participants are available in the Appendix C.

To provide a deeper understanding of the quantitative results, the qualitative data were analysed through thematic analysis (Braun & Clarke, 2006). The first two steps of the analysis (familiarizing with the data and generating initial codes) were conducted in parallel by the first author and a master's student doing a research internship. The coding strategy was mostly inductive. The two researchers then met to discuss the codes, sort out any differences, and conduct the third step (searching for themes). The last three steps (reviewing themes, defining and naming themes, and writing the analysis) were conducted by the first researcher.

3. Results

3.1 Quantitative results

3.1.1 RQ1: Students' reactions to the training

As shown in Figure 3, students seem to have reacted positively to the training, but there is an important intragroup variability, especially in the control group. Mann-Whitney *U* tests were conducted to investigate if students' reactions were more positive in the experimental (2-hour training) than in the control condition (1-hour training). The effect of the condition was investigated on students' impressions of the training (satisfaction and perceived learning) and students' perceived impact of the training (perceived utility and application to practice).

For students' impressions of the training, the effect is significant (U = 529, p = .003). Students in the experimental condition (M = 81.5, SD = 14.3) had a more positive impression of the training than students in the control condition (M = 70.5, SD = 17.4). This effect is of medium size (r_{rb} = .386).

The difference between the two groups is also significant for students' perceived impact of the training (U = 300, p < .001). Students in the experimental condition (M = 76.1, SD = 16.7) perceived the training as having a more positive impact than students in the control condition (M = 59, SD = 22.3). This effect is of medium size (r_{rb} = .470), close to a large effect (>0.5).

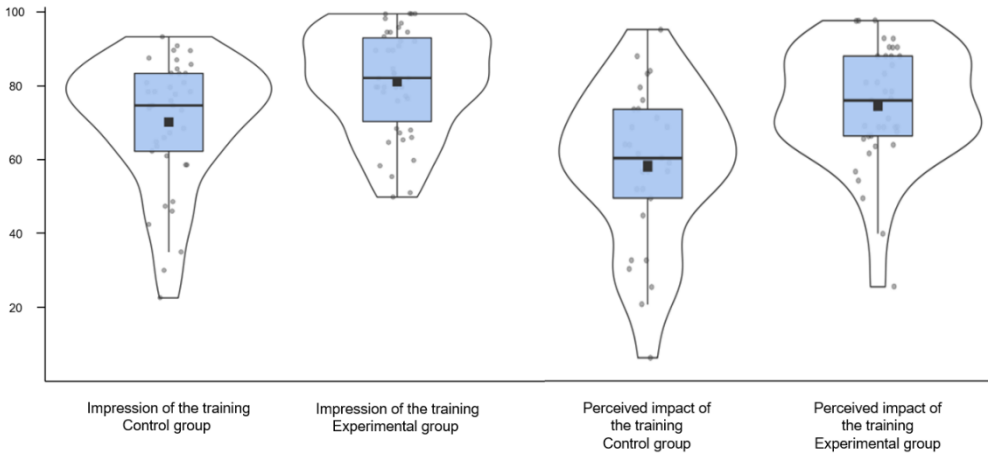


Figure 3. Boxplot of students' reactions to the training

3.1.2 RQ2: Impact of the intervention on students' perceptions of trust

The mean scores and standard deviation of students' perceptions of trust at each time are displayed in Table 3, while the results of the mixed method linear models are displayed in Tables 4 (for trust in the self) and 5 (for trust in the other). The results show an absence of significant effects. Contrary to our hypothesis, there was no significant interaction effect between time (linear) and group, neither for students' perceptions of trust in the self as an assessor (t(138.7) = 1.440, p = 0.152) nor for students' perception of trust in the other as an assessor (t(137) = 0.561, p = 0.576).

Table 3. Means and standard deviations of students' perceptions of trust

	Trust in the self as an assessor			Trust in the other as an assessor		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
Control group	4.09 (1.60)	4.02 (1.34)	4.17 (1.54)	3.91 (1.15)	3.99 (1.14)	3.92 (1.16)
Experimental group	3.75 (1.59)	3.93 (1.52)	4.30 (1.54)	4.07 (0.99)	4.2 (0.88)3	4.32 (1.20)

Table 4. Fixed effects parameter estimates of the linear mixed models for trust in the self

Effect	Estimate	Standard error	95% confidence interval		df	t	p
			Lower	Upper			
Intercept	3.911	0.196	3.526	4.296	98.5	19.920	<.001
Time (linear)	0.001	0.188	-0.367	0.369	142.4	0.005	0.996
Time (quadratic)	0.009	0.175	-0.334	0.352	149.1	0.053	0.958
Group	0.112	0.276	-0.430	0.653	93.5	0.404	0.687
Time (linear) * Group	0.360	0.250	-0.130	0.850	138.7	1.440	0.152
Time (quadratic) * Group	0.023	0.236	-0.441	0.487	143.4	0.097	0.923

Table 5. Fixed effects parameter estimates of the linear mixed models for trust in the other

Effect	Estimate	Standard error	95% confidence interval		df	t	p
			Lower	Upper			
Intercept	3.860	0.139	3.588	4.132	93.6	27.797	<.001
Time (linear)	0.335	0.195	-0.048	0.718	88.5	1.716	0.090
Time (quadratic)	0.077	0.144	-0.204	0.358	141.2	0.536	0.593
Group	-0.093	0.136	-0.360	0.173	148.4	-0.685	0.494
Time (linear) * Group	0.108	0.193	-0.270	0.486	137.0	0.561	0.576
Time (quadratic) * Group	0.073	0.184	-0.288	0.433	142.0	0.395	0.694

3.1.3 RQ3: Impact of the intervention on students' participation in peer feedback

As can be seen in Figure 3, participation in the peer feedback activities dropped remarkably after the second activity. To investigate if students in the experimental group participated more in the peer feedback activities that followed the intervention (peer feedback 3 and peer feedback 4), we conducted two chi-square tests. The observed and expected frequencies for the third and the fourth peer feedback activities are displayed in Tables 6 and 7.

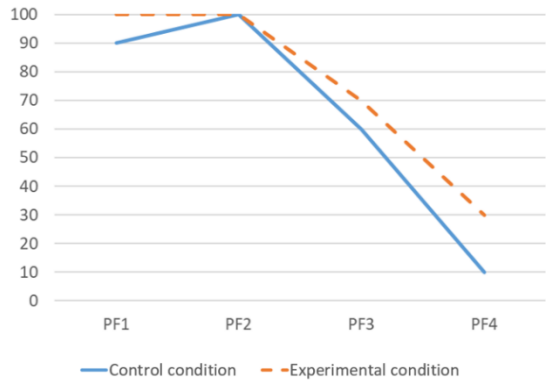


Figure 3. Participation (in %) in each peer feedback activity for participants included in the quantitative analyses (right, *N*=94)

For the third peer feedback activity, the results were not significant ($\chi^2(1) = 0.86$, $p = .353$). Contrary to our hypothesis, students in the experimental condition did not participate more in the activity than students in the control condition. Yet, the results were significant for the fourth peer feedback activity ($\chi^2(1) = 5.84$, $p = .016$, $\varphi = .249$). There is a moderate association between condition and participation in the fourth peer feedback activity. The proportion of students participating in this last activity was larger in the experimental condition than in the control condition.

Table 6. Contingency table for the third peer feedback activity.

		Participation in the activity		
Group		No-participation	Participation	Total
Control	Observed	19	29	48
	Expected	16.9	31	48
Experimental	Observed	14	32	46
	Expected	16.1	29	46
Total		33	61	94

Table 7. Contingency table for the fourth peer feedback activity.

		Participation in the activity		
Group		No-participation	Participation	Total
Control	Observed	43	5	48
	Expected	38.3	9.7	48
Experimental	Observed	32	14	46
	Expected	36.7	9.3	46
Total		75	19	94

3.2 Qualitative results

We developed three major themes (legitimacy to provide feedback, meaning given to the activities, and fear of lack of benevolence from peers) and three minor themes (time, mandatory nature of the activities, and trust in the learning process). Table 8 indicates for which participants each theme was of importance. We considered that a theme was not important if participants did not spontaneously mention it, of medium importance if participants spontaneously talked about it, and of high importance if participants came back to this theme throughout the interview. For each theme, Table 8 also indicates if the intervention seems to have pushed students' perspectives to evolve. The absence of evolution is not necessarily negative, some students already had positive perspectives from the outset (e.g., they always saw the meaning of peer feedback for their learning).

Table 8. Importance and evolution of the three major themes for each participant

	Meaning		Legitimacy		Fear of lack of benevolence	
	Importance	Evolution	Importance	Evolution	Importance	Evolution
Aaron	Medium	No	High	No	Medium	No
Achille	High	No	High	No	-	/
Amélia	High	No	High	No	-	/
Callie	Medium	No	Medium	No	Medium	No
Célia	High	Yes	High	No	Medium	No
Clémentine	High	No	Medium	Yes	Medium	No
Colas	Medium	No	High	No	High	No
Eden	Medium	No	High	No	-	/
Eleanor	High	Yes	Medium	No	-	/
Elena	Medium	No	High	Yes	High	Yes
Eliot	Medium	Yes	High	Yes	Medium	No
Elmir	Medium	No	Medium	No	Medium	No
Elise	Medium	Yes	Medium	No	Medium	No
Emiliano	High	No	High	Yes	-	/
Erin	High	Yes	High	Yes	Medium	No
Eva	High	No	High	Yes	-	/

3.2.1 The meaning given to peer feedback

The meaning given by students to peer feedback activities seems to be a decisive reason for their choice to participate or not in peer feedback. Some students considered providing and receiving feedback to be meaningful activities for their learning, while other students saw these activities as meaningless. This can be illustrated with the example of four students with very contrasted perspectives and for whom the meaning (or lack of) given to peer feedback did not evolve. Achille and Amélia, two students from the experimental condition who were both absent during the training session, explained that they did not see meaning in the peer feedback activities, while Clémentine and Emiliano considered from the outset these activities as beneficial to their learning. According to Achille and Amélia, the sole purpose of providing feedback is to help the student who will receive it. In contrast, Clémentine and Emiliano thought that providing feedback is as useful, if not more, as receiving it. They explained that providing feedback enables them to learn through the discovery of other students' work and through the reflection needed to provide feedback. They both talk about providing feedback in very positive terms (such as the pleasure of discovery, an intellectual game, etc.). The difference between these four students also lies in their perspectives on the received feedback. Achille and Amélia felt that the feedback they received during the first peer feedback activities was of insufficient quality to help them progress and, therefore, that the whole activity was a waste of their time. Even though they received feedback of similar quality, the perspective of Clémentine and Emiliano was different. They both considered that what is important is not to receive the "right" answer, but to receive feedback that encourages reflection and, therefore, fosters their learning.

> Most of these comments [the advantages and disadvantages written by students during the first course] assume that there's some kind of enlightened being who's going to come up with a truth. But as we've seen, that's not how learning happens. Learning occurs when something questions us, when it challenges our convictions. [...]. On the contrary, I think that even if you get feedback from someone who knows absolutely nothing, you can still learn from it. (Emiliano, meaning importance: high, meaning evolution: no)

The intervention seems to have enabled several students to give meaning to peer feedback. More specifically, these students identified the discussions during the second part of the training session as triggering changes in their perspective and, therefore, enabling them to give meaning to the activity. Students related two changes in their perspectives occurring during the training session. The first one was to move from seeing giving feedback as primarily useful for others to primarily useful for oneself.

> I didn't really see the point of giving feedback in that way, so it was difficult for me to engage in the activities. [...] And I understood it at that point. I said to myself oh in fact, I'm not doing it to give..., at first I was doing it thinking I was doing it to do a favour to others... and I said to myself well come on, at some point, we're in over our heads, it's a bit every man for himself because we can't cope any more [laughs]. And then I said to myself, but in fact, I'm doing it for myself when I give feedback to others. (Eleanor, meaning importance: high, meaning evolution: yes)

The second change was to move from seeing feedback provided by peers as information that allows them to correct their assignment to elements that can push their reflection and, thus, foster their learning. It seems that students viewed implications of this both as feedback providers and receivers. As feedback providers, it means that they can provide useful feedback even though they are still learning themselves and sometimes still in doubt. Eliot, for example, explained that having this training session helped him to feel reassured because he understood that his feedback could help his peers even if he did not find the "miracle formula" to write. As feedback receivers, it means that even if students do not give as much credit to their peers as to their teacher, they can still learn from their peers.

> But it was really when we were having this discussion about the difference between teacher feedback and peer feedback that I said to myself, but yes, that's really what it's all about. The aim is really to get us thinking. [...] I think we all necessarily trust our teacher more than our classmates, at least I think so. So, yes, it's experienced quite differently. But I think that peer assessment brings more to the table in that it's questioned, in fact. (Célia, meaning importance: high, meaning evolution: yes)

3.2.2 The feeling of legitimacy to provide peer feedback

The question of whether they were legitimate to provide feedback to their peers is a question that was discussed by all participants, even though they did not all use the word "legitimate" (for example, Eden wondered who he was to provide feedback). This is remarkable given that it is something that we never introduced ourselves during the interviews. Moreover, for most participants, it was a central question to which they returned on multiple occasions. In all cases, students approached this question from the angle of lack

of legitimacy and questioned their own legitimacy, except for Achille who questioned both his own and his peers' legitimacy.

The most common explanation given by participants for their feeling of lack of legitimacy is that they do not have a higher level of skills than their peers. Several participants point out that the psychology of learning is not their area of expertise, and therefore they do not feel legitimate to provide feedback on this subject. Other recurring reasons that students identified to explain a lack of legitimacy are (1) a perceived lack of clarity about teachers' expectations, (2) a lack of experience with peer feedback activities, and (3) personal characteristics, such as their experience, background, or general level of self-confidence.

When talking about whether they felt legitimate to provide feedback, students added nuances: they could feel legitimate to provide certain types of feedback and not others or to provide feedback on certain aspects and not others. For example, Emiliano explained that, for the first peer feedback activity, he felt legitimate to provide comments on his peers' assignments, but not to provide suggestions for improvement. Several students also explained that they felt legitimate to provide feedback on presentation aspects (e.g., layout, punctuation, grammatical errors, etc.), but not on the content of the assignments.

> I just don't feel right giving feedback on the content. On the presentation, yes, 100%, because I think that, really, everyone is capable of following a form and with objective criteria you can't go wrong. [...] But the content, it's huge. It's too much. (Colas, legitimacy importance: high, legitimacy evolution: no)

Students mentioned several consequences of the lack of legitimacy they felt. First, most students explained that they felt uncomfortable while providing feedback to their peers without feeling legitimate to do it and thus did not appreciate these activities. Second, some students reported that the type of feedback they provided and what they focused on were limited by their feeling of legitimacy, if they did not feel legitimate to provide a certain type of feedback, they did not do it. Third, not feeling legitimate seems to have led some students to not participate in some of the peer feedback activities. For example, Aaron, a student for whom the question of legitimacy was very important and whose perspective did not evolve, explained: " That's why I didn't do it too... that feedback, because if I had to give feedback myself on a subject I hadn't mastered at all yet, I felt I had very little legitimacy".

The intervention seems to have had a positive impact on this central issue of legitimacy. Some students reported feeling more legitimate to provide feedback thanks to either the training session or the backward evaluation. Eliot, Eva, Erin, Elena, and Clémentine all explained that they felt more legitimate after participating in the 2-hour training session. The training allowed them to feel equipped to provide feedback to their peers and to understand how they could help their peers, and, therefore, to feel more legitimate to provide feedback.

> I feel a little more legitimate for the simple reason that I feel a little better equipped. I had a workshop that equipped me, that brought me exactly the tools I needed, or at least a lot of the tools I needed to be able to provide feedback. (Eliot, legitimacy importance: high, legitimacy evolution: yes)

Emiliano was the only one to explain that the backward evaluation reinforced his feeling of legitimacy, however, it seems to be a key element to him. In the backward evaluation, his peers rated his feedback as being very useful but explained that they would have liked to receive more suggestions for improvement. Emiliano explained that, at first, he did not feel legitimate to provide suggestions, but thanks to the backward evaluation he realized that suggestions were what his peers expected of him.

> I thought I could only comment on what existed and not on what could exist or how it could be improved. But some of the students sent me a request saying, "Okay, tell me how I can improve this? Give me more advice". [...] I saw it as an element of legitimacy. Not only do I have the right to do it, but they ask me to. It's useful for them. So, it reinforced... yes... in any case, it reassured my need for legitimacy in relation to feedback, especially in relation to suggestions for improvement. (Emiliano, legitimacy importance: high, legitimacy evolution: yes)

3.2.3 The fear of a lack of benevolence from peers

The last major theme developed is the fear of a lack of benevolence from peers. This theme concerned fewer students than the two first themes, but we nevertheless considered it to be a major theme because it was central in Colas and Elena's interviews. They both express a fear of receiving unkind peer feedback and it seems to be a major obstacle to their involvement in the activities.

> [I experienced receiving feedback] like a fear [laughs]! Like a fear of judgment. I think in writing, people tend to be more aggressive than orally. [...] I think in itself, a benevolent person will always remain benevolent and a person without tact will never have tact. So I don't think there's any way of training a person. (Colas, fear of lack of benevolence importance: high, fear of lack of benevolence evolution: no)

While Colas explained that according to him it is not possible to encourage students to be more benevolent, Elena, whose perspective did evolve, did not share his opinion. She explained that one of the reasons she appreciated the training session was because the importance of benevolence was highlighted: "For me, the most important thing in feedback is benevolence, and to think that, with this training, the other people who evaluate us, who give us feedback, will, in turn, be benevolent too."

3.2.4 Minor themes: time, mandatory nature of the activity, and trust in the learning process

Three other themes, less central in the interviews, were developed. The intervention did not seem to affect these themes. However, they are still worth analysing because they can help understand why students did or did not participate in the peer feedback activities.

The first minor theme is time: the time peer feedback took, and the lack of time students had to do it. Time was mentioned by all students, probably because most of them combine their studies with a job and have busy schedules. Lack of time was given by some students as an explanation as to why they did not participate in some peer feedback activities. However, it seems that what distinguishes students who participated in the peer feedback activities from those who did not was not the time they had to devote to the activities, but the

perception that these activities were worth their time. Therefore, the determinant factor seems to be the meaning students give to these activities, even though their busy schedules may push them to question the relevance of these activities more than students who have less busy schedules.

> Of course, it's extra work, we're not going to lie. It takes time. If you want to do it correctly, well, that's it. But, it's extra work, which for me, before I had the small training courses, was perhaps a waste of time. I don't feel that way now. Because I realize how much reflection it brings. (Célia)

The second minor theme is the mandatory nature of the activities. When asked why they participated in peer feedback, most students explained that the main reason was because they were mandatory. For some students, it was even the only reason for their participation. It could explain why fewer students participated in it when they realized that there were no negative consequences linked to an absence of participation. Thus, it seems that making these activities mandatory is the most effective way to ensure students' participation. However, it does not seem to be an effective way to engage students in the activities. Students who only participated in peer feedback because it was mandatory did not report to reap any benefits from it. For example, Colas explained: "I felt I'd done a lot of work for nothing. I thought it was going to be included in the process and that it was necessary to validate the course credits, which wasn't actually the case".

The last minor theme is the trust that students have in the learning process developed by their professors. This theme was particularly present among two students, Célia and Eleanor. They explained that, even if at first, they did not perceive the value of peer feedback, they participated in it anyway because it was the learning process developed by their professors. They did not only participate in it because it was mandatory, but also because they trusted that if their professor implemented peer feedback, it was because it could be useful for their learning. This trust allowed them to keep participating in peer feedback until they perceived its value thanks to the intervention. However, they both regretted seeing the value of peer feedback so late (especially Célia who was in the control condition and only received the second part of the training at the end). According to them, they would have been even more involved in peer feedback if they had seen the value of these activities from the start.

> If I attend a class, it's to do things properly, and because I feel that the professors who teach us, if they suggest something, it's very likely that there's a value in it. Basically, I did it because it was an activity that was proposed, and it was my job as a student to take part in it. [...] To be honest with you, I didn't realize the benefits until very late. I think I might have taken it even more seriously... I've always taken my work seriously, but I think I might have considered it differently if I had had that perception beforehand. (Célia)

4. Discussion

The current study aimed to investigate if the combination of a training session targeting trust and a backward evaluation could increase students' perception of trust in the self and in the other as an assessor and, in consequence, increase their willingness to participate in peer

feedback activities. A mixed-method approach with an explanatory sequential design (Creswell, 2012) was used to investigate the impact of the intervention, which was provided to students in the experimental condition (*N*=95) while students in the control condition participated in the peer feedback without receiving the intervention (*N*=94). First, quantitative analyses were conducted to investigate the intervention's outcomes: students' perceptions of the training, the evolution of students' perceptions of trust in the self and in the other, and students' participation in the peer feedback activities. Then, in the second phase, sixteen follow-up interviews were conducted to better understand the quantitative results.

The first two hypotheses concerned students' reactions to the training: we hypothesised that students who received the more comprehensive 2-hour training would have more positive impressions of the training session (H1a) and would report more utility and application to practice of the training (H1b) than students who only received the 1-hour training on how to provide peer feedback. Both hypotheses were confirmed. This means that students positively reacted to the training, which should have positively affected their motivation to learn from it (Blanchard & Thacker, 2012).

Concerning attitudinal changes, we hypothesised that students' perceptions of trust in the self as an assessor (H2a) and in the other as an assessor (H2b) would increase more in the experimental than in the control condition. The results did not confirm these hypotheses: the intervention neither had a significant effect on students' perception of trust in the self nor on trust in the other as an assessor. This result is consistent with the absence of effects of a training session targeting trust and psychological safety found by Senden et al. (2023) but contradicts L. Li's (2017) study that found a positive impact of training targeting peer pressure on students' perceptions of peer pressure.

For trust in the other, it is possible that the intervention did indeed affect trust, but this impact was not significant in the results due to the insufficient reliability of the scale used to measure it. Trust in the self as an assessor and in the peer as an assessor were both measured with French translations of scales used by Rotsaert et al. (2018). These scales were chosen because in Rotsaert et al.'s (2018) study they have shown good reliability with participants similar to those of the current study: Belgian higher education students majoring in education. Even though a rigorous process was adopted to translate the items from Dutch to French with forward and backward translation as recommended (e.g., Tsang et al., 2017), a pilot test was not conducted to verify the psychometric proprieties of the translated scale. Moreover, given that peer feedback only consisted of written comments in the current study, we excluded items from Rotsaert et al.'s (2018) scale that mentioned scores, which may have led to an insufficient number of items to ensure reliability (Tsang et al., 2017). These elements could explain why the reliability of the scale measuring trust in the other was low and could explain the absence of a significant impact of the training on students' perceptions of trust. Future studies should be conducted to rigorously translate and validate scales measuring trust in the self and trust in the other as an assessor in French, by respecting all the steps recommended in the literature (e.g. Boateng et al., 2018; Tsang et al., 2017).

The last two hypotheses concerned behavioural changes: we hypothesised that students in the experimental condition would participate more in the third (H3a) and in the fourth (H3b) peer feedback activities than students in the control condition. The results partially supported our hypotheses: students from the experimental group participated more in the fourth peer feedback activity than students from the control group, but there was no difference between the two conditions for the third peer feedback activity. This result can seem counterintuitive given that between the third and the fourth peer feedback activity, for ethical reasons, students from the control condition also had received the training focused on trust (second hour of the training session). It may be that for the control condition, the training came too late to have a meaningful impact on students' participation, or that the backward evaluation is an essential component for the intervention to be successful. Indeed, a previous study already found a positive impact of a backward evaluation on students' motivation to participate in peer feedback (M. Kim, 2009). The absence of impact of the intervention on the third activity could be explained by the participation rates. Approximately half of the students participated in the third activity, while only a smaller percentage participated in the fourth (approximately 30% for the experimental condition and 10% for the control condition). It is possible that the mandatory nature of the peer feedback was still the main reason to participate in the third activity, while by the fourth activity, all the students had realized that not participating in the peer feedback activities would not negatively affect their course grades, and only the more motivated students participated in it.

Our original assumption was that students receiving the intervention would positively react to the training session, which, combined with the backward evaluation, would positively affect students' perceptions of trust, which would positively affect their participation in peer feedback. Results confirmed that the training was well-received and that the intervention was successful in increasing participation in peer feedback. Based on previous studies suggesting a link between trust and participation in peer feedback (Al Qunayeer, 2020; Poverjuc et al., 2012), we hypothesised that with the intervention students would have more trust in the self and in the other and, therefore, would be more inclined to participate in peer feedback. However, the intervention did not have a significant impact on students' perceptions of trust in the self and in the other as an assessor. In addition to the explanation for the absence of effects on self-reported trust, the interviews indicated an additional explanation for the success of the intervention in the absence of a significant impact on students' perceptions of trust.

The possible explanation is that the intervention missed its target, students' perception of trust, but still ended up positively affecting students' willingness to participate in peer feedback through two other mechanisms: the meaning given to peer feedback and students' feeling of legitimacy. According to the qualitative results, an important determinant of students' participation in peer feedback is the meaning they give to these activities. This concept is sometimes also referred to in the literature as the educational value of peer feedback (e.g., McGarr & Clifford, 2013) and was found to be associated with students' perceptions of trust in the only study investigating their relationship (Rotsaert et al., 2017). Without surprise, students who saw peer feedback as meaningless for their learning were not

inclined to participate in it. Even though it was not the main goal of the training, it seems to have enabled some students to see more meaning in peer feedback. This is consistent with findings from L. Li's (2017) study according to which students who received a training session designed to address their concerns saw more value in peer feedback than students who did not receive the training session. The discussions that followed the role-plays seem to have brought two insights to some students: (1) when they provide feedback, it is for them that it can be the most useful, and (2) feedback from peers does not have to be totally accurate to incite them to reflect on it and, therefore, to learn from it. Therefore, it is possible that instead of ensuring trust in the self and in the other as an assessor, the intervention allowed students to see the value of peer feedback even if they do not trust their own or their peers' abilities to provide feedback.

The intervention may also have affected students' feeling of legitimacy to provide feedback to their peers. During the interviews, students seldom mentioned the word "trust", while the concept of legitimacy was brought up by every participant and seemed central for most of them. Students often explained that they questioned their legitimacy to provide feedback, which seemed to have negative consequences, such as making them uncomfortable, pushing them to only provide more superficial feedback, or even pushing them to not participate in peer feedback. Six of the sixteen interviewed students reported that their feeling of legitimacy increased during the semester. Although it could simply be because they developed their knowledge on the subject of the course, the intervention seemed to have played an important role: students reported they felt more legitimate thanks to the 2-hour training session because they felt more equipped to provide feedback. Given the intervention aimed to increase trust, a concept that seems to be close to legitimacy, it is plausible that it positively affected students' sense of legitimacy. The concept of legitimacy, which seemed central for students participating in this study, is never included in literature reviews on students' concerns linked to peer feedback (e.g., Panadero et al., 2023; van Gennip et al., 2009) or, to our knowledge, in any other study on peer feedback. Therefore, further research is necessary to confirm the importance of legitimacy in peer feedback.

In sum, even though the intervention did not have a significant impact on students' perceptions of trust in the self and in the other, it was successful in fostering students' participation in peer feedback. This confirms that training designed to target specific students' concerns can be effective, as already suggested by a previous study (L. Li, 2017) and suggests that the absence of effects in Senden et al.'s (2023) study was either due to the methodological limitation of the study or to the absence of a backward evaluation to support the training efficacy. Future studies should investigate if the intervention is also effective in other contexts, for example with students enrolled in other majors than education, and if the positive effects of the intervention on students' participation in peer feedback activities are maintained over time.

4.1 Limitations and pathways for future research

The first two limitations were due to the study procedure adopted. First, the intervention was tested as a whole, which prevents us from disentangling the effect of the training session targeting trust from those of the backward evaluation. Given that these activities take time to organise for the teaching team and to participate in for students, it is important to know the role that each component plays in the effectiveness of the training (e.g., one is not effective without the other, they are both effective alone but less than taken together, etc.). The qualitative results suggested that the training element is essential to the intervention and that the backward evaluation is a useful add-on but this should be confirmed by future research.

Second, given that we wanted to allow all students to benefit from the training targeting trust for ethical reasons, students in the control condition participated in it after the last questionnaire but before the last peer feedback activity to have the opportunity to apply the training to at least one peer feedback activity. This means that students' rate of participation in the fourth peer feedback activity and the qualitative data were collected when both groups had received the training session targeting trust, which complicated the interpretation.

To overcome these two limitations linked to the study procedure, a study could be conducted with a greater number of students, in a course that implements multiple peer feedback activities throughout the year. Four groups of students could be compared during a semester: (1) a group receiving the complete intervention, (2) a group receiving only the training session targeting trust, (3) a group receiving only the backward evaluation, and (4) a control group receiving neither the training session nor the backward evaluation. During the following semester, the elements proven effective could be provided to students who did not receive them.

Conducting future studies with a greater number of participants is all the more relevant as another limitation of the current study is that approximately half of the students were excluded from the quantitative analyses, which had a negative impact on the statistical power of the analyses (Howell, 2007). The impact of the intervention could not be investigated on these students, either because they were absent during the training session or because they did not participate in the backward evaluation or the peer feedback activities.

Finally, a fourth limitation of the study is that students in the experimental group only participated in one backward evaluation. This choice was made due to organisational constraints: automated backward evaluation was not possible on the peer feedback tool available (the workshop module in Moodle) and that made it time-consuming to organise. However, even though a single instance of backward evaluation has been found beneficial to enhance the peer feedback process (e.g., M. Kim, 2009; Ng & Yu, 2023), multiple backward evaluations could be even more effective. Indeed, it would allow students to have feedback on their feedback on multiple occasions, and, therefore, to see if their feedback is judged as more useful when they try to improve it. Given that other peer feedback tools, such as SWoRD, Peergrade, PeerWise, and PEARS have options for backward evaluations (Misiejuk & Wasson, 2021), future research should investigate if multiple backward evaluations have more positive impacts than a single backward evaluation.

4.2 Practical implications

The current study has practical implications for higher education educators who implement peer feedback or would like to implement it. The results suggest that the intervention could be used when organising peer feedback activities. The mechanisms of the intervention still need to be better understood and replication studies are needed to confirm the results and verify if the intervention stays effective in a variety of contexts. However, given that the training session was well-received by students and that the intervention increased their willingness to participate in peer feedback, we feel confident to recommend its use, especially for groups of students who are reluctant to participate in peer feedback because they do not trust their and their peers' abilities. Therefore, all the training material is available on request.

4.3 Conclusion

The current study aimed to investigate the efficacy of an intervention composed of a training session targeting trust in peer feedback and a backward evaluation. Results revealed a positive impact of the intervention on students' participation in a peer feedback activity. However, contrary to our hypotheses, the intervention did not have a significant impact on students' perceptions of trust in the self and in the peer as an assessor. Two explanations were proposed to explain these results: (1) limitations in the scales used may have prevented us from seeing the impact of the intervention, (2) the intervention may have affected the meaning that students attributed to peer feedback or their feeling of legitimacy to provide feedback instead of their perceptions of trust. Future studies are needed to empirically test these two explanations.

5. Appendices

5.1 Appendix A. Instrument to report the characteristics of peer assessment designs

Instrument to report the characteristics of peer assessment designs

Designed by Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education.* **Use that citation if you include the instrument in your publication**.

TITLE OF THE ARTICLE

Fostering higher education students' perceptions of trust and participation in peer feedback: Investigating the efficacy of an intervention combining training and backward evaluation

Authors

Morgane Senden, Mariane Frenay, Bernadette Charlier, & Liesje Coertjens

Our study investigates:

- ☒ The impact of peer assessment on interpersonal/intrapersonal factors
- ☐ The impact of interpersonal/intrapersonal factors on peer assessment
- ☐ The impact of peer assessment on performance/behaviour
- ☐ The impact of moderator/mediator variables on peer assessment outcomes

Describe the characteristics of your peer assessment study in the table below.

	Category	Description [i]	Our study
Context			
1	**Subject domain**	Subject domain the study was done in (e.g., mathematics, instructional sciences, accounting, etc.)	Education (psychology of learning)
2	**Place/Time**	Where was the PA conducted? (In class or out of class?)	☐ In class/class time ☒ Out of class/free time
3	**Setting**	Formal or informal education setting?	☒ Formal ☐ Informal
4	**Requirement**	Was PA compulsory or voluntary for assessor/assessee?	☒ Compulsory ☐ Voluntary
5	**Alignment**	Was the PA activity aligned to curriculum, learning goals or teaching?	☒ Yes ☐ No
Instructional design			
6	**Purpose**	What was the assessment purpose of the PA activity? (Formative, summative or both?)	☒ Formative ☐ Summative ☐ Both
7	**Object**	What was assessed? (e.g., written assignment, oral presentation, contribution to group work)	Written assignement (case analysis)
8	**Product/Output**	What was the output of the PA? (e.g., score, written feedback, oral feedback, or a combination)	Written feedback
9	**Relation to staff assessment**	Was PA done without staff assessment (substitutional) or in addition to staff assessment (supplementary)?	☒ Substitutional ☐ Supplementary
10	**Official weight**	Did participation in the PA activity or the grade given by peer(s) contribute to learners' final grades?	☒ No ☐ Yes-for participation in PA ☐ Yes-for PA grade ☐ Yes-both (PA & participation) ☐ Other: Click here to add text
11	**Reward**	Was there a reward for participation in PA?	☒ No ☐ Yes-course credit ☐ Yes-incentives (e.g., free time, money, etc.) ☐ Other: Click here to add text
12	**Directionality**	Was the learner assessing another without being assessed (unidirectional) or acting as both assessor and assessee (bidirectional)?	☐ Unidirectional ☒ Bidirectional
13	**Degree of interactivity**	How did the assessee demonstrate engagement and response to PA?	☒ Reactive: assessee responds to assessor ☐ Reciprocal: same learners assess each other on same task ☐ Negotiated: PA was done more than once on the same task and both parties negotiated it
14	**Frequency**	How often was the PA of the same task done? (Once, twice, etc.)	Four times on the same case analysis, but once on each question

15	**Group constellation**	Did members of the same group assess each other (intragroup) or peers from another group (intergroup) or both?	☐ Intragroup ☐ Intergroup ☐ Both
16	**Constellation assessor**	The number of assessors assigned to each assessee	3
17	**Constellation assessee**	The number of assessees per assessor	3
18	**Unit of assessment (assessor)**	At what level did the assessor(s) perform PA? Individual, group, or both?	☒ Individual ☐ Group ☐ Both
19	**Unit of assessment (assessee)**	At what level did the assessee(s) experience the PA? Individual, group, or both?	☒ Individual ☐ Group ☐ Both
20	**Privacy**	Did assessor and assessee know the identity of each other? (Was PA public, single-blind, or double-blind?)	☐ Public ☐ Single-blind (for assessor) ☐ Single-blind (for assessee) ☒ Double blind (anonymous) ☐ Other: Click here to add text
21	**Contact**	Was PA done face-to-face or online? How was the contact between assessor and assessee?	☐ **Face-to-face synchronous:** same time same place (ST SP) ☐ **Online synchronous:** same time different place (ST DP), e.g., skype, chat, etc. ☒ **Online asynchronous:** different time different place (DT DP), e.g., e-mail, SWoRD, etc. ☐ **Other:** Click here to add text
22	**Matching**	How were assessor and assessee matched for the PA activity? (e.g., random, skill, self-select, gender, friendship, performance, etc.)	Random
23	**Format**	How was the PA guided?	☐ Freestyle (no instruction provided to learers) ☐ Guided instruction ☐ Guided prompts ☒ Guided criteria/rubrics ☐ Guided criteria/rubrics and prompts ☐ Other: Click here to add text
24	**Training**	Did the learners receive PA training at any time? If yes, describe the moment in which they received it and the type of training.	They received a 1-hour session to prepare them before the peer assessment (they learned how to give effective feedback according Hattie & Timperley (2007) framework and they practiced peer feedback)
25	**Revision**	Did learners revise their work after receiving or providing PA?	☒ No ☐ Yes
26	**Scope of involvement**	How were learners involved in the PA activity?	☐ Development of assessment criteria ☐ Gave PA only ☐ Received PA only ☒ Gave and received PA ☐ Additional self-assessment ☐ Other: Click here to add text
Outcomes			
27	**PA Outcomes**	These variables are directly measured as outcomes of the PA activity (i.e., why was the PA activity conducted?) Select all the options that apply to your study from the right column.	☒ **Beliefs & perceptions:** including perceptions of learning capacity to perform PA or any perceptions related to the PA processes (e.g., fairness, usefulness), metacognition and self-regulation, attitudes and beliefs (e.g., self-efficacy), teachers' perceptions/conceptions. ☐ **Emotions and motivation:** emotions experienced by learners (e.g., achievement emotions, social emotions, etc.) & all motivational beliefs (e.g., learning motivation). ☐ **Performance:** academic/domain specific performance, achievement, improved draft/work (i.e., revision). ☐ **Skills:** quality of contribution to the group, professional behaviour, problem solving skills, work habits, interpersonal skills, metacognitive & self-regulatory skills. ☐ **Reliability of PA:** consistency of PA scores compared to other peer assessors, or PA over several rounds. ☐ **Validity of PA:** accuracy of PA compared to teachers/tutor/expert's assessment. ☐ **PF content:** characteristics of the feedback messages including type, focus, quality, frequency of comments/posts (i.e., participation), etc. ☐ **PF processing:** Includes implementation, reactions to PF, seeking PF, coping with PF, etc. ☒ **Other:** Participation in PF
Moderators/mediators			
28	**PA Moderators/mediators**	Variables that are not usually manipulated but are taken into account when investigating PA. Select the variables that have been explored in your study from the right column.	☐ **Gender:** of assessor/assessee. ☐ **Ability & Skills:** includes prior knowledge, prior performance, achievement level, GPA, finished high school, previous level of education, year of enrolment, etc. ☐ **Skills:** reviewing ability, computer skills, etc. ☐ **Age:** of assessor/assessee. ☐ **Culture, ethnicity, nationality or race** ☐ **Other:** Click here to add text
This tool is based on: Alqassab, M., Strijbos, J., Panadero, E., Fernández Ruiz, J., Warren, M., & To, J. (2023). A systematic review of peer assessment design elements. *Educational Psychology Review*.			

Note. PA = peer assessment; PF = peer feedback

5.2 Appendix B. Items used to measure students' reactions and attitudes

	Items in French	Translation in English
Impression of the training	La séance de formation était réaliste et axée sur la pratique	The training was realistic and practical
	La méthode était bien adaptée aux objectifs et au contenu de la séance de formation	The method was well-suited to the objectives and content
	La méthode utilisée nous a permis de participer activement à la séance de formation	The method used enabled us to take an active part in the training
	La séance de formation mérite une bonne note globale	The training merits a good overall rating
	Je garderai un bon souvenir de la séance de formation	I will keep the training in good memory
	J'ai beaucoup apprécié la séance de formation.	I enjoyed the training very much
	J'ai appris beaucoup de nouvelles choses au cours de la séance de formation	I learned a lot of new things in the training
	Après la séance de formation, je sais plus de choses sur les feedbacks entre les pairs qu'avant	After the training, I know substantially more about the training contents than before
Perceived impact of the training	La séance de formation m'a été très utile pour effectuer l'évaluation par les pairs	The training was very beneficial in conducting the peer assessment
	Participer à ce type de formation avant une évaluation par les pairs est très utile pour moi	Participation in this kind of training before a peer assessment is very useful for me
	Lors de l'évaluation par les pairs, j'ai utilisé les connaissances acquises lors de cette séance de formation	During the peer assessment, I used the knowledge I gained in the training
	J'ai réussi à appliquer les contenus de la formation lors de l'évaluation par les pairs	I successfully managed to apply the training contents during the peer feedback activity
Trust in the self as an assessor	Je me sens suffisamment compétent·e pour donner des feedbacks à mes pairs	I feel competent enough to give feedback to my peers
	Je pense être capable de donner à mes pairs des feedbacks sur leurs productions	I think I can give my peers feedback on their work
	J'ai confiance en ma capacité à évaluer mes pairs en leur donnant des feedbacks	I am confident in my ability to assess my peers by giving them feedback
	Je pense être capable d'évaluer mes pairs	I think I am able to assess my peers
Trust in the other as an assessor	Je pense que mes pairs sont capables de me donner des feedbacks objectifs	I think my peers are able to provide me objective feedback
	Je **ne** fais **pas** confiance aux feedbacks donnés par mes pairs sur mes productions (reverse item)	I don't trust the feedback that my peers provide on my work (reverse item)

5.3 Appendix C. Interview guide adapted to Eliot and material traces presented to him

A general interview guide was created and adapted to each student. The main questions are in regular font, while italics is used for follow-up questions.

Questions about the student and his previous experience with peer feedback

To begin with, can you introduce yourself?

Which studies did you do before? What is your professional background? What job do you do at the moment?

Why did you decide to start a master in education?

Is the master up to your expectations? Why or why not?

The psychology of learning course may not be your first experience of peer feedback. Can you tell me about any peer feedback experiences you may have had before?

What were your experiences like? What were the positive aspects of these experiences? What were the negative aspects of these experiences?

When you learned that there would be peer feedback for the psychology of learning course, what did you think of it?

How did you react to the announcement that there would be peer feedback?

If I'm not mistaken, you took part in the first three peer feedback activities, but not the last one. What were your reasons for participating or not?

What about the first one? What about the second? What about the third? What about the fourth? Were there any other reasons?

Questions about the first peer feedback activity

We're now going to go into detail about the different moments of the peer feedback, and I'm going to show you my computer. As a reminder, here is the feedback you gave during this first activity (**Trace 1**). Can you tell me how it went?

How did you go about giving feedback? How did you feel in the role of feedback provider? You left your name for the third feedback...

Trace 1

Le cas Anita – socio-constructivisme

4) Le dispositif de la formation pratique en atelier répond-il aux critères d'un dispositif coopératif ? Justifier votre réponse

Malgré le fait que Anita soit seule lorsqu'elle exécute les différents choses demandées (ligne 101). Elle est entourée par son formateur et par les autres membres du groupe.

Ligne 40 et + → elle dit qu'elle retrouve le même plaisir que chez elle en autodidacte et même plus car ici il y a le formateur présent et ça l'aide à bien progresser.

Ligne 47 → elle dit que dans le groupe il y a beaucoup d'entraide.

Ligne 101 → il travaille seul mais le formateur passe pour donner des conseils et pour les aiguiller.

Ligne 105 → elle dit que l'ambiance du groupe est bienveillante et qu'il arrive qu'il se donne l'un ou l'autre conseil.

Il y a de l'entraide mais je ne dirais pas que le dispositif mis en place par le formateur (lors des ateliers) est un dispositif coopératif. Les élèves travaillent majoritairement seuls.

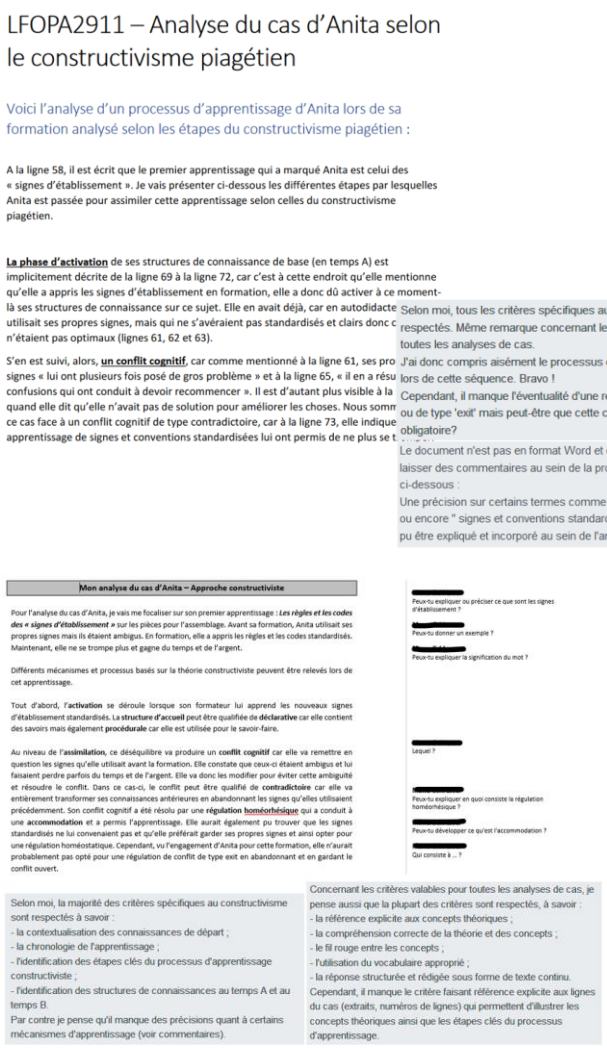
5) Trouvez-vous des traces de relation asymétrique ? pose-t-elle problème pour l'apprentissage ?

Oui il y a des traces de relation asymétrique entre le formateur et Anita.

Ligne 107 → elle dit que le formateur intervient souvent et de manière très directive pour corriger les erreurs. Mais ce n'est pas un frein à l'apprentissage car Anita comprend et a une grande confiance dans son formateur, elle sent de la bienveillance en lui.

Feedback général

Malheureusement l'analyse du cas lors de cette séquence porte sur le constructivisme et pas sur le socio-constructivisme.



LFOPA2911 – Analyse du cas d'Anita selon le constructivisme piagétien

Voici l'analyse d'un processus d'apprentissage d'Anita lors de sa formation analysé selon les étapes du constructivisme piagétien :

A la ligne 58, il est écrit que le premier apprentissage qui a marqué Anita est celui des « signes d'établissement ». Je vais présenter ci-dessous les différentes étapes par lesquelles Anita est passée pour assimiler cette apprentissage selon celles du constructivisme piagétien.

La phase d'activation de ses structures de connaissance de base (en temps A) est implicitement décrite de la ligne 69 à la ligne 72, car c'est à cette endroit qu'elle mentionne qu'elle a appris les signes d'établissement en formation, elle a donc dû activer à ce moment-là ses structures de connaissance sur ce sujet. Elle en avait déjà, car en autodidacte utilisait ses propres signes, mais qui ne s'avéraient pas standardisés et clairs donc c n'étaient pas optimaux (lignes 61, 62 et 63).

S'en est suivi, alors, **un conflit cognitif**, car comme mentionné à la ligne 61, ses pro signes « lui ont plusieurs fois posé de gros problème » et à la ligne 65, « il en a résu confusions qui ont conduit à devoir recommencer ». Il est d'autant plus visible à la quand elle dit qu'elle n'avait pas de solution pour améliorer les choses. Nous somm ce cas face à un conflit cognitif de type contradictoire, car à la ligne 73, elle indique apprentissage de signes et conventions standardisées lui ont permis de ne plus se t

Selon moi, tous les critères spécifiques au constructivisme sont respectés. Même remarque concernant les critères valables pour toutes les analyses de cas.
J'ai donc compris aisément le processus d'apprentissage analysé lors de cette séquence. Bravo !
Cependant, il manque l'éventualité d'une régulation homéostatique ou de type 'exit' mais peut-être que cette consigne n'était pas obligatoire?

Le document n'est pas en format Word et donc impossible de laisser des commentaires au sein de la production. Je les écris donc ci-dessous :
Une précision sur certains termes comme "signes d'établissement" ou encore " signes et conventions standardisées" auraient peut-être pu être expliqué et incorporé au sein de l'analyse.

Mon analyse du cas d'Anita – Approche constructiviste

Pour l'analyse du cas d'Anita, je vais me focaliser sur son premier apprentissage : ***Les règles et les codes des « signes d'établissement »*** sur les pièces pour l'assemblage. Avant sa formation, Anita utilisait ses propres signes mais ils étaient ambigus. En formation, elle a appris les règles et les codes standardisés. Maintenant, elle ne se trompe plus et gagne du temps et de l'argent.

Différents mécanismes et processus basés sur la théorie constructiviste peuvent être relevés lors de cet apprentissage.

Tout d'abord, l'**activation** se déroule lorsque son formateur lui apprend les nouveaux signes d'établissement standardisés. La **structure d'accueil** peut être qualifiée de **déclarative** car elle contient des savoirs mais également **procédurale** car elle est utilisée pour le savoir-faire.

Au niveau de l'**assimilation**, ce déséquilibre va produire un **conflit cognitif** car elle va remettre en question les signes qu'elle utilisait avant la formation. Elle constate que ceux-ci étaient ambigus et lui faisaient perdre parfois du temps et de l'argent. Elle va donc les modifier pour éviter cette ambiguité et résoudre le conflit. Dans ce cas-ci, le conflit peut être qualifié de **contradictoire** car elle va entièrement transformer ses connaissances antérieures en abandonnant les signes qu'elles utilisaient précédemment. Son conflit cognitif a été résolu par une **régulation homéorhésique** qui a conduit à une **accommodation** et a permis l'apprentissage. Elle aurait également pu trouver que les signes standardisés ne lui convenaient pas et qu'elle préférait garder ses propres signes et ainsi opter pour une régulation homéostatique. Cependant, vu l'engagement d'Anita pour cette formation, elle n'aurait probablement pas opté pour une régulation de conflit de type exit en abandonnant et en gardant le conflit ouvert.

[redacted] Peux-tu expliquer ou préciser ce que sont les signes d'établissement ?
[redacted] Peux-tu donner un exemple ?
[redacted] Peux-tu expliquer la signification du mot ?
[redacted] Lequel ?
[redacted] Peux-tu expliquer en quoi consiste la régulation homéorhésique ?
[redacted] Peux-tu développer ce qu'est l'accommodation ?
[redacted] Qui consiste à ... ?

Selon moi, la majorité des critères spécifiques au constructivisme sont respectés à savoir :
- la contextualisation des connaissances de départ ;
- la chronologie de l'apprentissage ;
- l'identification des étapes clés du processus d'apprentissage constructiviste ;
- l'identification des structures de connaissances au temps A et au temps B.

Par contre je pense qu'il manque des précisions quant à certains mécanismes d'apprentissage (voir commentaires).

Concernant les critères valables pour toutes les analyses de cas, je pense aussi que la plupart des critères sont respectés, à savoir :
- la référence explicite aux concepts théoriques ;
- la compréhension correcte de la théorie et des concepts ;
- le fil rouge entre les concepts ;
- l'utilisation du vocabulaire approprié ;
- la réponse structurée et rédigée sous forme de texte continu.

Cependant, il manque le critère faisant référence explicite aux lignes du cas (extraits, numéros de lignes) qui permettent d'illustrer les concepts théoriques ainsi que les étapes clés du processus d'apprentissage.

You also received feedback from two of your peers. You can see the feedback you received here (**Trace 2**). Can you tell me what it was like to receive this feedback?

What did you do with the feedback you received? How did you feel in this role? Both people left their names[1]...

[1] Students received instruction to maintain their identity anonymous, but some of them did not respect the instruction.

Trace 2

Elle concerne l'apprentissage d'une technique nécessaire à la réalisation d'un assemblage à mi-bois (« *Un autre apprentissage… concernait la technique pour réaliser un* d'utiliser le ciseau à bois afin d'obtenir un résultat optimal. Il y a donc eu une régulation homéorhésique. Autrement dit, Anita a modifié sa structure d'accueil concernant la technique d'assemblage à mi-bois en modifiant ses connaissances antérieures (utilisation unique du ciseau à bois et manière de l'utiliser) grâce à l'intégration de nouvelles informations par le formateur (utilisation du ciseau à onglet dans un temps et, changement dans la manière d'utiliser le ciseau à bois dans un temps).

À propos de cette séquence, je pense qu'une régulation du conflit de type e à-dire l'abandon du conflit cognitif par le sujet car il n'est pas suffisamment engagé pour trouver une solution, aurait pu se produire si Anita n'avait pas expérimenté cette nouvelle technique à plusieurs reprises et constater la supériorité du résultat (« *Elle a*

cette technique auparavant (« *Elle s'y était déjà essayée à plusieurs reprises par le passé, en autodidacte » ligne 82*) et, en utilisant seulement un outil (« *utilisant uniquement un ciseau à bois » ligne 83*) n'obtenant dès lors pas le résultat escompté (« *un résultat assez approximatif et peu satisfaisant » ligne 83*). La nature du conflit cognitif est de type lacunaire, c'est-à-dire que

progressive de la structure d'accueil en une structure nouvelle, plus adaptée aux caractéristiques propres des éléments à assimiler. Cette accommodation s'est faite par l'incorporation d'éléments nouveaux au sein de la structure de connaissances sans

nouvelle technique à plusieurs reprises et constater la supériorité du résultat (« *Elle a donc pu expérimenter à plusieurs reprises cette nouvelle technique et réaliser la supériorité impressionnante du résultat final. » lignes 88 et 89).*

Questions about the training session

During the first practical session, you were divided into small groups to follow a training on peer feedback. What do you remember about this training?

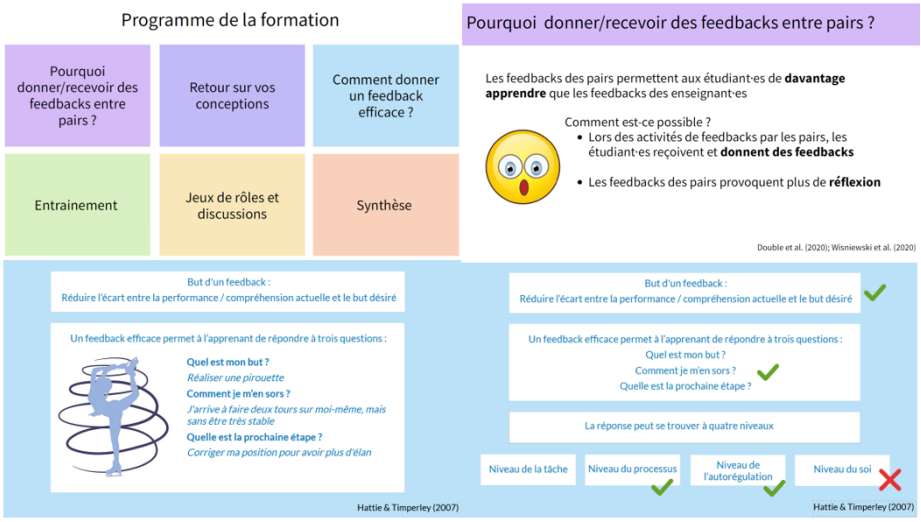
Generally speaking, what did you think of this training (**Trace 3**)?

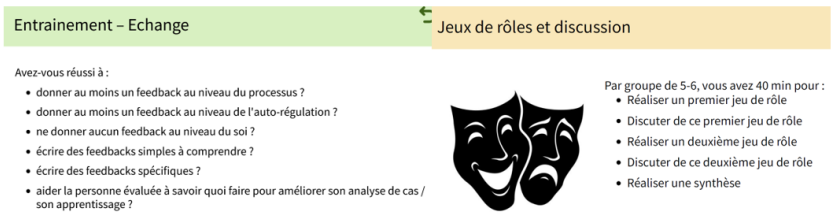
What parts of the training did you find most important? What did you learn from the course?

If you were to create a peer feedback training course yourself, what would you put in it?

What would you have liked to be included in this training?

Trace 3





Etape 2 – Premier jeu de rôle (5 à 10 min)

Mise en situation

Marie, Inès et Liam sont des étudiant·es en master en économie. Pour leur cours d'économie politique, ils doivent réaliser un travail individuel de 8 pages à rendre en janvier. Ils ont la possibilité de participer à une évaluation par les pairs, facultative et formative, en décembre. S'ils décident de participer, ils devront lire les travaux de quatre de leurs pairs et rédiger des feedbacks pour chaque travail. En échange, ils recevront des feedbacks écrits de quatre de leurs pairs sur leur travail. Marie est fort emballée par l'idée, Liam est hésitant et Inès plutôt réticente. Ils en discutent ensemble.

Mettez fin au jeu de rôle lorsque, après avoir écouté les arguments des autres, vous avez chacun·e décidé si vous alliez ou non participer à l'évaluation par les pairs facultative.

Etape 3 – Première discussion (5 min max.)

Utilisez les questions ci-dessous pour lancer la discussion. Les questions servent de point de départ, mais vous pouvez vous en éloigner. Le scribe prend note.

Questions pour Marie, Inès et Liam :

- Votre position a-t-elle évolué au cours de la situation ? Pourquoi ? Quels arguments vous ont le plus marqué ?

Questions pour les observateur·trices :

- La discussion qui a eu lieu vous paraissait-elle réaliste ? Quels arguments vous ont le plus marqué ?

Questions about the second peer feedback activity

As a reminder, here is the feedback you gave during this second activity (**Trace 4**). Can you tell me how it went?

How did you go about giving feedback? How did you feel in your role of feedback provider?

Trace 4

Mon analyse du cas d'Anita – Approche socio-constructiviste

Le dispositif pratique en atelier ne répond pas vraiment aux critères d'un **dispositif coopératif** car il n'y a pas de coopération entre les différents élèves (*lignes 99 à 111*). Ils ne travaillent pas pour un projet de groupe mais de manière individuelle donc il n'y a pas de **responsabilisation individuelle** (pas de projet collectif) ni **d'interdépendance positive de buts** (la pièce demandée par le formateur est individuelle, ils ne dépendent pas des autres élèves, l'avancée des autres élèves ne leur apporte rien) (*lignes 99 à 102*). Les seules conditions remplies sont la **taille du groupe** qui peut être qualifié de restreint (8 à 10 élèves) et permet donc des interactions riches et individualisées (*ligne 99*).

Il y a des traces de **relation asymétrique** car elle reconnait la supériorité du formateur et a confiance en ses compétences (*lignes 47 à 50*). Le formateur ne se met pas juste en position d'expert mais plutôt à l'écoute des progrès ce qui favorise la relation qui se base sur la bienveillance (*lignes 108 à 110*).

En ce qui concerne les **axes**, la configuration observée est l'**aménité**. Elle pense qu'il est directif pour de bonnes raisons et elle le trouve bienveillant (*lignes 108 à 111*).
On observe également de la **contradiction** car même si elle trouve parfois le comportement de son formateur pénible, elle le comprend (*lignes 173 à 174*). Il la confronte à ses propres erreurs en remettant en cause une erreur dans ses calculs malgré sa formation de comptable. Dans ce cas-ci, la confrontation n'est pas par la parole mais par la surprise du formateur face à l'erreur d'Anita (*lignes 120 à 125*).

On constate des signes de **conflit socio-cognitif** au moment où le stage lui casse l'image qu'elle se faisait du métier. Elle est alors confrontée à une réalité qui ne colle pas du tout à ce qu'elle pensait (*lignes 176 à 188*).
Au niveau de la **régulation**, elle peut être définie comme **épistémique**. Un premier exemple est lorsqu'elle accepte les remarques du formateur dans le but de progresser (*lignes 107 à 111*). Un deuxième exemple est lorsqu'elle se rend compte qu'elle a négligé un aspect de la question, avec la moue dubitative de son formateur (*lignes 121 à 130*).
On peut également être définir la régulation comme **relationnelle protective**. Un premier exemple est lorsqu'elle rend sa copie sans contester les autres étudiants ni le formateur (*lignes 126 à 128*). Un deuxième exemple est lorsqu'elle abandonne le lieu de stage en ne maintenant pas son point de vue face aux autres. Elle préfère partir (mode « exit ») (*lignes 121 à 130*).

Anonyme
Quels sont ces critères ?

Anonyme
A définir

Anonyme
A définir

Anonyme
Je ne sais pas s'il y a réellement conflit socio-cognitif ... conflit certainement mais "socio", il me semble que c'est l'ensemble des discussions et contributions de chacun qui permet de résoudre le conflit et donc ici Anita est seule à subir cette situation, il n'a pas vraiment de conflit ....à mon sens ! Je te fais juste part de mon point de vue! Consulte la slide 15.

Anonyme
A définir

A part, le conflit socio-cognitif sur lequel j'ai un point de vue différent du tien, l'ensemble de l'analyse me semble correcte.
J'ai cependant remarqué que tu n'avais pas expliquer théoriquement les concepts mobilisés dans les situations analysées.
Je propose donc dans le feedback général, de te montrer ma manière de définir, par exemple, le concept de la configuration des "axes". Pour ce qui est de l'analogie avec la situation concrète et les extraits choisis c'est top, donc je te montre uniquement au niveau théorie.

N'hésite pas à t'entrainer avec les autres concepts que je t'ai mentionnés lors des commentaires (à définir sur le doc Word).
Bonne chance !

Feedback général

Une des conditions d'efficacité d'un conflit sociocognitif est la dimension affective que prennent les interactions sociocognitives. Il faut distinguer deux dimensions : celle liée au relationnel : aménité (bienveillance) >< contrariété (malveillance) et celle liée aux idées (savoirs) : contradiction (ne pas être d'accord) >< approbation (être d'accord).

La solution la plus favorable à l'apprentissage consiste en la combinaison de l'aménité et de la contradiction. En effet, un climat de bienveillance favorise les divergences de point de vue.

Ce texte permettrait donc d'introduire la notion des axes. Ensuite, viendrait s'y greffer ton illustration avec tes références (lignes, ...).

Pour l'explication de la théorie, tu peux trouver les informations dans les slides du cours se rapportant au socio-constructivisme. Notamment, dans cette exemple, la slide 19.

Anita – Socio-constructivisme

Q4 : Le dispositif de la formation pratique en atelier répond-il aux critères d'un dispositif coopératif ? Justifier votre réponse (pour chaque critère pris en considération).

La séquence analysée se trouve entre la ligne 99 et la ligne 111.

Les critères d'un dispositif coopératif sont :
Une situation d'interdépendance positive des buts entre les membres du groupe.
Un groupe restreint.
Un temps suffisamment long et significatif sans supervision.

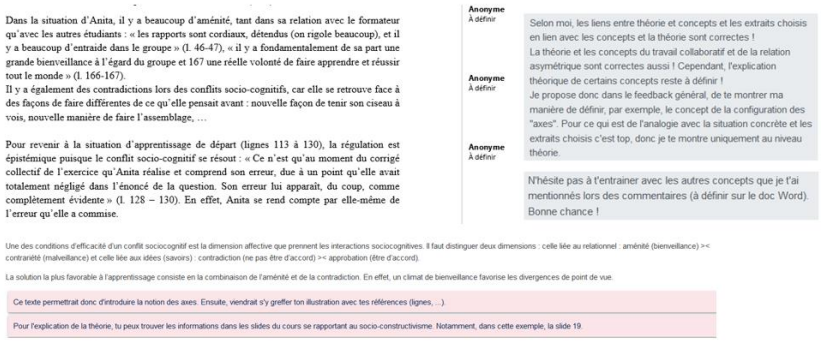
Y-a-t-il une situation d'interdépendance positive entre les membres du groupe ? Non, chaque élève travaille sur son établi (ligne 100) malgré le fait que les élèves se donnent l'un ou l'autre conseil (ligne 105) car il est précisé que cela reste rare et le but n'est pas d'y arriver ensemble mais bien individuellement (individualisation).

Le groupe est-il restreint ? Oui, puisque les élèves sont en petits groupes de 8 à 10 participants (ligne 99).

Le temps sans supervision est-il suffisamment long ? Oui, il laisse les participants travailler à leur rythme (ligne 101), les élèves peuvent être seul sans supervision face à une tache d'usinage sur une machine (ligne 104).

Ce dispositif ne répond pas totalement aux critères d'un dispositif coopératif : 2 critères sur 3.

Selon moi, tout est maitrisé : les concepts, la théorie, les liens entre théorie et concepts et les extraits choisis en lien avec les concepts et la théorie ! Je n'ai donc aucun commentaire écrtit sur le doc Word !
Bravo !



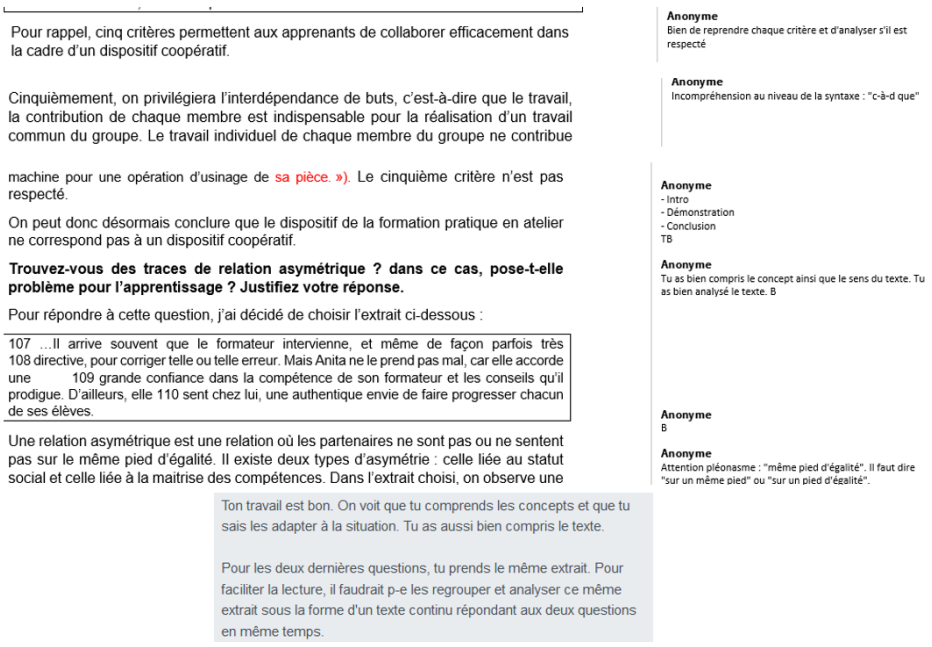
You also received feedback from one of your peers. You can see the feedback you received here (**Trace 5**). Can you tell me what it was like to receive this feedback?

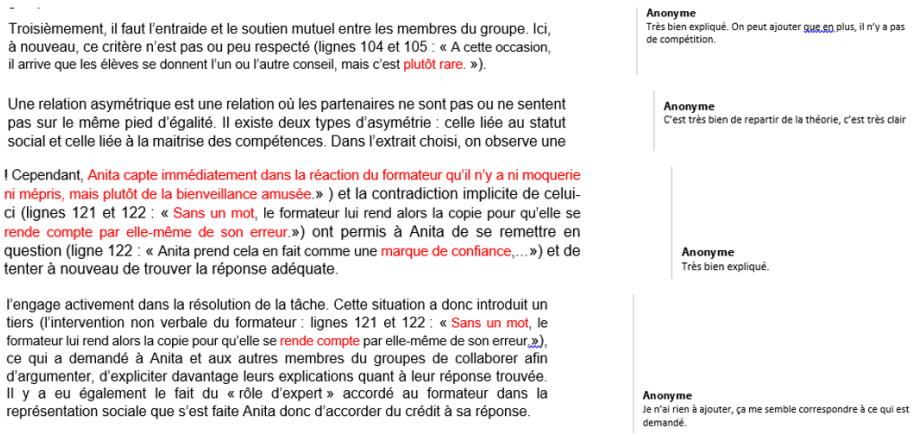
What did you do with the feedback you received? How did you feel in this role?

Do you feel you experienced this second peer feedback activity differently from the first?

Did you use what you learned during the training? Did you notice any difference in the quality of the feedback you received?

Trace 5





Questions about the backward evaluation

After this second peer feedback activity, you were asked to give your opinion on the three feedback you had received via an online survey. Here's what you said about the only feedback you had at the time (**Trace 6**). How did you reach this opinion?

In your opinion, what were the advantages/disadvantages of answering this form?

Trace 6

A quel point les feedbacks donnés par l'évaluateur 1 vous ont-ils été utiles ?

	Pas du tout utile	Pas tellement utile	Assez utile	Fort utile	Extrememement utile
Evaluateur1				x	

Si vous le voulez, expliquez pourquoi vous trouvez utile ou non les feedbacks donnés par l'évaluateur 1

Ils me permettent de peaufiner mon analyse et les détails qu'il m'a fournis sont importants.

Comment les feedbacks de l'évaluateur 1 pourraient être améliorés ?

Autre suggestion d'amélioration :

L'évaluateur aurait pu me montrer une phrase ou deux en exemple afin d'illustrer ses propos.

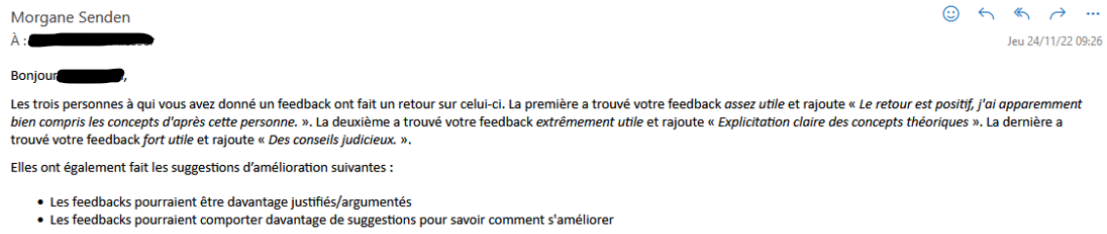
Your peers have formulated an opinion on the feedback you gave them. I e-mailed it to you (**Trace 7**). What did you think of it?

What kind of feedback would you have expected? What kind of feedback would have been useful to you?

What impact did the feedback you received have on your motivation to do the next peer feedback activity?

What kind of feedback would have motivated you to do the next peer feedback activity?

Trace 7



Morgane Senden
À : [redacted]
Jeu 24/11/22 09:26

Bonjour [redacted],

Les trois personnes à qui vous avez donné un feedback ont fait un retour sur celui-ci. La première a trouvé votre feedback *assez utile* et rajoute « *Le retour est positif, j'ai apparemment bien compris les concepts d'après cette personne.* ». La deuxième a trouvé votre feedback *extrêmement utile* et rajoute « *Explicitation claire des concepts théoriques* ». La dernière a trouvé votre feedback *fort utile* et rajoute « *Des conseils judicieux.* ».

Elles ont également fait les suggestions d'amélioration suivantes :

- Les feedbacks pourraient être davantage justifiés/argumentés
- Les feedbacks pourraient comporter davantage de suggestions pour savoir comment s'améliorer

Questions about the third peer feedback activity

As a reminder, here is the feedback you gave during this third activity (**Trace 8**). Can you tell me how it went?

How did you go about giving feedback? How did you feel in your role as feedback provider?

Trace 8

- *Q8 : Les apprentissages auxquels Anita a été confrontée dans sa formation se situaient-ils dans sa ZPD ? Justifier la réponse*

Dans la formation d'Anita, elle n'est pas confronté à une situation de ZPD car il n'y a pas de début à son apprentissage. Nous avons seulement le cas d'un apprentissage qui évolue. Nous n'avons pas tous les éléments qui nous aideront à situer la zone proximale de développement.

- *Q9 : Trouver une illustration de la relation « participation-identité-apprentissage »*

Dans le cas, à analyser, Anita n'est pas à l'écart du groupe. Elle n'est pas incluse dans les activités non plus. Selon moi, il n'y a pas d'illustration qui pourrait justifier la relation « participation-identité-apprentissage » car elle est uniquement en interaction avec le professeur.

Difficile de commenter cette analyse car le document est au format PDF. De plus, il manque énormément d'informations si pas la quasi-totalité En effet, tu t'es contenté(e) de donner un bref avis sur la question. Il manque l'ensemble d'es critères généraux à l'étude des cas ainsi que tous les critères concernant l'approche historico-culturelle. Voici un exemple d'étude de cas pour la question 1 que je te propose dans la rubrique "feedback général". C'est mon étude de cas, elle n'est peut-être pas optimale mais elle pourra te donner des idées.

Pour contextualiser ces dires dans un de ses apprentissages, je vais reprendre celui des signes d'établissement. Il est décrit de la ligne 58 à la ligne 79. Aux lignes 60 et 61, elle explique qu'elle utilisait, précédemment, ses propres signes. Néanmoins, aux séances d'apprentissage de la formation sur cette matière, grâce aux cours donnés p et l'entraide de ses pairs, elle y a amélioré ses connaissances. Cela se confirm à la ligne 79. Après ces cours, elle ne se trompait plus, elle gagnait du temps sentiment de s'inscrire comme membre à part entière dans cette communau professionnelle.

Tous les critères tant les généraux valables pour toutes les analyses que les spécifiques concernant la cadre historico-culturel sont présents. Cette étude de cas est vraiment riche et complète, bravo ! Si je peux me permettre un petit conseil : je te propose de recopier des extraits un peu plus longs afin que nous n'ayons pas besoin de retourner dans le texte de base. En effet, même si les lignes sont mentionnées, le fait de reprendre des extraits est plus visuel à mon sens. Encore bravo !

L'écart entre ces deux niveaux définit la zone proximale de développement (ZPD). Celle-ci définit l'espace dans lequel peuvent et doivent prendre place les processu activités d'enseignement, lesquels transforment et déterminent égalemer lui donnant forme et contenu.

Je n'ai fait aucun commentaire concernant le cas d'Anita pour la simple et bonne raison que je compte me servir de cette analyse afin de peaufiner la mienne qui me parait moins complète !
Donc merci à toi ! Cette analyse est parfaite à mon sens ! Tout y est : que ce soit les critères généraux pour les analyses ainsi que ceux spécifiques à l'approche historico-culturelle !
Encore bravo!

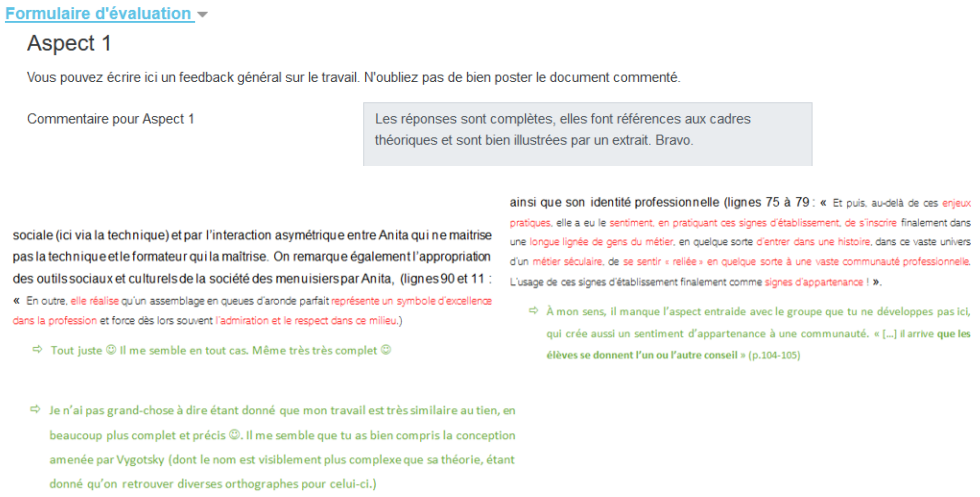
You also received feedback from two of your peers. You can review the feedback you received here (**Trace 9**). Can you tell me what it was like to receive this feedback?

What did you do with the feedback you received? How did you feel in this role?

Do you feel you experienced this third peer feedback activity differently from the first two?

Did you use what you learned during the training? Did you notice any difference in the quality of the feedback you received?

Trace 9



Questions about the student's current perceptions of peer feedback

At the beginning of the term, I came to a class and asked you to take part in a Wooclap. I asked you what you thought the advantages and disadvantages of peer feedback were, and what your fears were about these activities. This is what came out of the Wooclap (**trace 10**). Do you currently agree with the answers?

Why or why not? How long have you felt this way? And for the advantages? What about fears? What experiences have led you to think this? What advantages would you add now? What disadvantages/concerns would you add now?

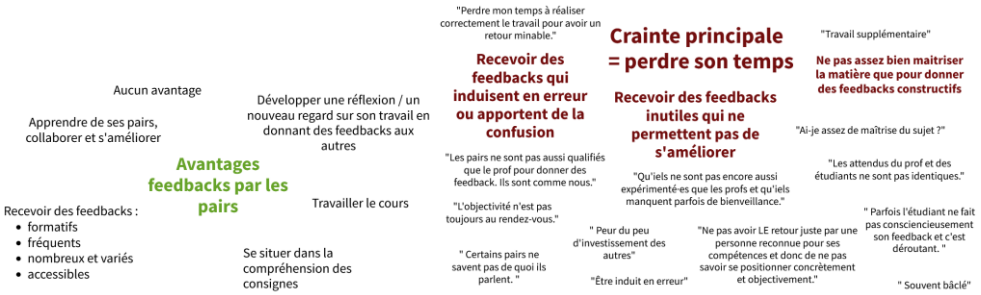
If, next term or next year, you were told that there would be peer feedback in a course, what would you think of it?

What type of courses are best suited to peer assessment?

What would you recommend be implemented for students taking the psychology of learning course next year?

Why would you/ wouldn't you keep peer feedback activities? How would you organise them?

Trace 10



Supplementary questions to ask if the subject has not been addressed by the students

What did you think of the fact that peer feedback was anonymous?

What are the advantages and disadvantages of anonymity in peer review?

Anonymity of the assessor? Anonymity of the evaluated? In this case, feedback and work were theoretically anonymous, but was this always the case?

What impact does the fact that most students are also teachers have on peer feedback activities?

What aspect of the peer feedback activities did you learn the most from?

Do you feel you have learned more by evaluating or being evaluated? Why?

If you had to choose in which course to put a peer feedback activity (psychology of learning or any other course), which course would you choose?

Why?

I asked all the questions I had. Do you have anything you'd like to add?

Chapter 6

General discussion

General discussion

1. Aim and structure of the dissertation

Feedback is recognized as an indispensable tool to foster students' learning: effective feedback helps students identify their strengths and areas for improvement as well as actions to take to improve and achieve their desired goal (e.g., Hattie & Timperley, 2007; Sadler, 1989). Therefore, it is an essential component of higher education. Yet, with the massification of higher education, teaching staff rarely have the opportunity to provide feedback to students, which means that alternatives to teachers' feedback are required, such as the peer feedback-process (Huisman et al., 2019).

In this dissertation, "the peer feedback-process" refers to formative uses of peer assessment, which is itself defined by Alqassab et al., (2023, p.2) as:

> A learning phenomenon where individuals or social constellations (e.g., pair, group, team, community)—within a physical and/or virtual environment— exchange, react to, interact and/or act upon information about individual performance and/or individual contributions to the process and/or product of a social constellation, with the purpose to accomplish implicit or explicit shared and individual learning goals (e.g., domain-specific knowledge or skills, social skills, etc.)

"Process" is added to distinguish between two conceptions of feedback: a process-centric conception that considers feedback as an interactive process between a learner and someone else, and an information-centric conception that considers feedback as information presented to a learner (Winstone et al., 2022). Therefore, in this dissertation, "peer feedback-information" refers to feedback provided by a peer when following the definition of feedback proposed by Lipnevich and Panadero (2021, p. 25):

> Feedback is information that includes all or several components: students' current state, information about where they are, where they are headed and how to get there, and can be presented by different agents (i.e., peer, teacher, self, task itself, computer). This information is expected to have a stronger effect on performance and learning if it encourages students to engage in active processing

The educational value of the peer feedback-process in higher education is well-established. It is an effective way to foster students' learning, even more effective than the teacher feedback-process (Double et al., 2020: Wisniewski et al., 2020). Moreover, it contributes to the development of students' feedback literacy, which is required for students to be able to act upon feedback-information and to become independent lifelong learners (Carless & Boud, 2018; Tai et al., 2018).

However, the peer feedback-process is not without challenges. It is a social experience that involves interactions between students and, as such, interpersonal processes of peer feedback-process are important to consider (Panadero et al., 2023; Topping, 2017; van

Gennip et al., 2009). Psychological safety, trust in the self as an assessor, and trust in the other as an assessor are three important interpersonal processes that seem to affect the benefits students can reap from the peer feedback-process (e.g., K. Cheng & Tsai, 2012; van Gennip et al., 2012). Therefore, peer feedback activities should be designed with these interpersonal processes in mind. This is why the current dissertation aims to better understand how peer feedback activities can be designed to ensure that higher education students feel psychologically safe and trust their own and their peers' abilities.

The first step in answering our research question was to draw a clear picture of the current state of knowledge on the subject. Therefore, the **first study** systematically reviewed the literature on the impact of instructional design elements of peer assessment on interpersonal processes. Based on 27 included studies, the results revealed that the most investigated design elements were privacy (i.e., who is anonymous for whom), the format of peer assessment (i.e., the kind of guidance provided to students), and training. Other design elements were also investigated, but no conclusion can be drawn given the small number of studies and their weak research design. Of the three most investigated design elements, training seemed to be the most promising to address the interpersonal factors of the peer feedback-process. Two studies (Ocampo et al., 2023; Parikh & Sheehan, 2016) found a positive impact of a general training session on students' feeling of comfort, while another study (L. Li, 2017) confirmed the effectiveness of a training session specifically designed to reduce students' feeling of peer pressure. The impact of anonymity was mixed: it seemed to positively affect some interpersonal processes, while negatively affecting others. Concerning the format of peer assessment, it seemed that providing more guidance (such as providing a guide with specific steps to take) could have a positive impact. However, there were some important weaknesses in the design of the study investigating it. Only two studies had (quasi)-experimental design: one found that providing students with a detailed regulation script for a dialogic peer feedback-process had a positive impact on trust (L. Cheng et al., 2023), while another did not find any positive impact of the use of a rubric on comfort or fairness (Panadero et al., 2013). Hence, the impact of the format of peer assessment on interpersonal processes remains unclear.

Given the potential of training to address interpersonal processes, I decided to design a training session specifically targeting trust in the self, trust in the other, and psychological safety. The **second study** detailed the design of an online version of the training and investigated its feasibility and efficacy. The training, which mostly contained active learning methods, was successfully delivered online. Results from questionnaire data and semi-directed interviews suggested that most students reacted positively to it. However, students' perceptions of the training varied strongly: some students saw little value in the training, while others had very positive perceptions of it and reported that it positively affected their behaviour during the peer feedback-process. This important variability could be explained by the diversity in students' initial needs and by a variability in students' engagement during the training session. I hypothesized that students' engagement in the training would vary less if delivered on campus, with the physical presence of educators and other students. Indeed,

the virtual presence of others does not have the same positive impact on engagement in learning as the physical presence of others (Sarasso et al., 2022).

Considering the encouraging results of the second study, I decided to investigate the efficacy of the training when delivered on campus. The **third study** is a quasi-experiment with a pre-and post-test design. Given that it is well-known that a general training session on how to provide peer feedback-information is a requirement for the success of the peer feedback-process (van Zundert et al., 2010), such training was provided to all students. Therefore, the impact of the comprehensive training session designed in Study 2 (including a part on how to provide feedback-information and a part targeting trust and psychological safety) was compared to that of a 1-hour general training session on how to provide feedback-information. Data were collected from two cohorts of students: students who followed the course in 2019-2020 (N = 47) served as a control group (only receiving the 1-hour general training session), while students who followed the course in 2021-2020 (N = 42) received the comprehensive 2-hour training session. The impact of the comprehensive training session was investigated on (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-information students provided to their peers, and (3) students' performance improvement following the peer feedback activity. The effectiveness of the comprehensive training was not supported by the results: it did not have a significant positive impact on any of these outcomes.

Two aspects may have limited the training effectiveness: (1) the student-to-instructor ratio and (2) the duration of training. Indeed, our objective was to create a training session that could easily be implemented in higher education. Therefore, the training was designed to be provided by a single instructor for a group of approximately 50 students and to be short enough to be implemented in busy schedules. However, it is recommended that instructors visit each group during the role-plays, to ensure that students are on task and to answer any questions they may have (Bolinger & Stanton, 2020). Moreover, active methods take time (Martin et al., 2014) and by keeping the training short it is not possible to combine various training methods or to include more than one role-play for each target (trust and psychological safety), both of which are recommended to foster training effectiveness (Blanchard & Thacker, 2012; Martin et al., 2014).

Several adaptations of the training were made to overcome these two limitations. The first major change was that the training was adapted to only focus on trust. This way, the training could be kept short while still dedicating enough time to foster trust (e.g., both role-plays were focused on trust). The decision to dismiss psychological safety was justified by the fact that psychological safety is difficult to foster (Hunt et al., 2021) and that previous evidence also suggested that a training session may not be enough to develop it (Dusenberry & Robinson, 2020). The second major change was that the training session was not the only element of our intervention anymore, a backward evaluation was added. As backward evaluation can be organised outside class time, it can reinforce the training without impacting its feasibility. The last major change was to adapt the training to be provided with smaller groups (of approximately 25 students), and therefore to decrease the student-to-instructor ratio.

The absence of results supporting training effectiveness in the third study may also be explained by limitations in the study design. First, for ethical and practical reasons, the third study adopted a pre-and post-test design with two cohorts of students. This design is relatively strong, provided that no major changes occurred between the two data collections (Cohen et al., 2007). Unfortunately, the COVID-19 pandemic took place and strongly affected students (e.g., Alger & Eyckmans, 2022; Di Pietro, 2023), which hindered the comparability between the control and the experimental group. Moreover, another limitation of the third study was that students participating in it already had quite high levels of trust and psychological safety at the pre-test, which could have made the training less needed and its possible impact more difficult to observe. These positive perceptions could be explained by the fact that the students in this group knew each other well and were used to helping one another. Another limitation of the study was that students' reactions to the training were not measured. When investigating the impact of an intervention, it is important to investigate three types of outcomes which are affected by one another: (1) participants' reactions to the intervention, (2) the impact of the intervention on the knowledge, skills or attitudes it was designed to address, and (3) the impact of the intervention on participants' behaviours (Blanchard & Thacker, 2012). If not all outcomes are investigated, it is difficult to determine what causes an absence of effects. For example, it may be that students did not like the training and, in consequence, were not attentive during it and did not learn from it. Finally, a last limitation was that attitudinal changes were only measured a month after the training. Therefore, it is possible that the training affected students' attitudes, but that this effect faded away and was not detectable anymore a month later.

The fourth study was designed to overcome these limitations. First, to ensure comparability between the control and experimental groups, the study was conducted with a large cohort of students that was divided between an experimental and a control group. Moreover, conducting the study with a large cohort of students, and therefore with students who did not know their peers as well as the students in the previous study, increased the likelihood of low levels of trust at the pre-test. Finally, the intervention was implemented in a course with multiple peer feedback activities and, to better understand the effects of the intervention, students' reactions, attitudes, and behaviours were all measured on different occasions. A downside of the large number of students included and the multiple peer feedback activities was that it was not realistic anymore to code all the feedback-information provided by students and to assess all students' productions. Therefore, the impact of the intervention on students' behaviours was only measured through participation rates.

The **fourth study** aimed to investigate the effectiveness of an intervention composed of two elements: (1) the provision of a training session that aimed to increase trust in the self and in the other as an assessor and (2) the presence of a backward evaluation. The intervention was implemented in a course followed by master students in educational sciences. Students assigned to the experimental condition received the comprehensive 2-hour training session (one hour on how to provide feedback-information and one hour on targeting trust) at the start of the course and participated in a backward evaluation after the second peer feedback activity. Students assigned to the control condition received the 1-hour training on how to

provide effective feedback-information at the same time as students in the experimental condition, but only received the second hour of the training (targeting trust) at the end of the semester. Moreover, they did not participate in the backward evaluation. Quantitative analyses were conducted to investigate differences between students in the two conditions regarding students' reactions to the training sessions, students' perception of trust in the self and in the other, and students' rate of participation in the last peer feedback activities.

The results revealed an unexpected pattern. Even though students who received the comprehensive training session perceived it more positively than students who only received the 1-hour training session, there was no significant impact of the intervention on students' perceptions of trust. This suggests that the intervention, even though well-received, was ineffective. However, the intervention did have a positive impact on students' behaviours: the participation rate in the last peer feedback activity was larger in the experimental than in the control condition, which supports the effectiveness of the intervention. To better understand these results, sixteen follow-up interviews were conducted in a second phase. Qualitative analyses confirmed that students appreciated the intervention, which seemed to have impacted them in several ways. First, some students explained that the training and, to a lesser extent, the backward evaluation allowed them to feel more legitimate in providing feedback-information to their peers, something that seemed crucial for them. Second, some students reported that it was thanks to the training session that they realized that the peer feedback-process had value for their learning.

In this general discussion, the main findings of the chapters will be integrated to provide answers to our research question. Next, the limitations of the dissertation will be discussed and perspectives for future research will be proposed to overcome these limitations. Moreover, I will also propose avenues of research complementary to those explored in this dissertation and provide implications for educational practices.

2. Integrative findings

In this section, findings from the three empirical studies (studies 2, 3, and 4) will be integrated to answer the main research question of this dissertation: *How can peer feedback activities be designed to ensure that higher education students feel psychologically safe and trust their own and their peers' abilities?* More precisely, I will try to determine if incorporating a training session targeting trust and psychological safety (Study 2 and Study 3) or an intervention combining a training session targeting trust and a backward evaluation (Study 4) is beneficial. This section will be divided into four sub-sections: (1) students' concerns related to interpersonal processes, (2) how these interventions were received by students, (3) their impact on students' attitudes, and (4) their impact on students behaviours.

2.1 Students' concerns related to interpersonal processes

To tailor the interventions to their specific needs, students were asked to provide their thoughts and concerns about the peer feedback-process at the start of each of the three experimental studies. Students tended to see both advantages (e.g., the possibility of learning from seeing other students' work) and disadvantages (e.g., peers are more subjective than teachers). They also expressed many concerns linked to the peer feedback-process. Interestingly, many of these concerns were related to a lack of trust in the self (e.g., fear of misleading their peers), a lack of trust in the other (e.g., fear of a lack of involvement from their peers which would result in low-quality feedback-information), and, to a lesser extent, a lack of psychological safety (e.g., fear of being perceived as stupid). In line with previous research (e.g., Hanrahan & Isaacs, 2001; L. Li & Steckelberg, 2013), students' answers confirmed that they had concerns related to the three interpersonal processes investigated in this dissertation. This means that it is relevant to design interventions that aim to increase students' levels of trust in the self, trust in the other, and their perceptions of psychological safety.

Even though all students expressed concerns related to interpersonal processes, there was a striking difference between students enrolled in a master in motricity sciences (Study 2 and Study 3) and students enrolled in a master in educational sciences (Study 4). When asked about the advantages of the peer feedback-process, motricity science students seemed to see as many advantages as disadvantages, while the most common answer for educational science students was "there are no advantages". Moreover, the concerns expressed by students in motricity science seemed to reflect good relationships between students: they were afraid to receive undeserved positive feedback-information because their peers would be "too nice" or because feedback-information would be distorted by friendship. It was not the case for the educational sciences students, who principally feared receiving useless feedback-information due to a lack of involvement from their peers. Students from the two groups also seemed to have different levels of trust in the self and trust in the other, as measured with our questionnaire (the means were of approximately one more point on a 7-point scale for students in motricity sciences compared to students in educational sciences). These findings confirmed that the general classroom climate seems to be of major influence on students' perceptions related to the peer feedback-process (Rotsaert, 2017) and that students in different majors have different perceptions (Praver et al., 2011; Wen & Tsai, 2006).

The difference between the two groups could be linked to the difference in class size, with larger classes in educational sciences (with more than 100 students) than in motricity sciences (with no more than 50 students). Moreover, students in motricity sciences often interact with each other by playing sports during their lessons and are used to situations that require psychological safety (e.g., dancing in front of the whole class) or trust (e.g., doing an acrobatic exercise with the support of a peer). Another difference between the two groups is that most educational sciences students combine their studies with a part-time or full-time job and, as they tend to be older than students in motricity sciences, they are more likely to

have children. Therefore, even though students in motricity sciences also have busy lives (e.g., with sports competitions or student jobs), the question of time is not as central for them as it is for students in educational sciences. It could explain why educational sciences students seemed less inclined to participate in peer feedback activities, why their main concern was to waste their time by not reaping enough benefits from the activity, and why they feared that their peers would not devote enough time to provide relevant feedback-information. Finally, a last important difference between the two groups was that almost all students in educational sciences participated together in peer feedback activities the year before, while students in motricity sciences had more diverse previous experiences with peer feedback activities. The interviews revealed that most students in educational sciences perceived these previous experiences as being of little use, which may have shaped their perceptions of the peer feedback-process.

2.2 Students' reactions to the training

Students' reactions to the training were investigated in Study 2 and Study 4. In both cases, students had to indicate their percentage of agreement with items adapted from Grohmann and Kauffeld's (2013) Questionnaire for professional training evaluation and from Holgado-Tello et al.'s (2006) Training satisfaction rating scale. In Study 2, due to the small number of students (N=27), the five sub-scales used were analysed separately (general reaction, satisfaction, perceived learning, perceived utility, and application to practice), while in Study 4 a factor analysis suggested a two-factor structure: impression of the training (general reaction, satisfaction, and perceived learning) and perceived impact of the training (perceived utility and application to practice).

Study 2 was a study without a control group, so students' reactions were only descriptively analysed. A striking finding was that students' reactions to the training varied strongly from one student to another. The general impression of the training was mostly positive for every student but, for the other subscales, the score ranged from 10 to 90%, with standard deviations ranging from 18 to 24. This means that, while some students were very satisfied with the training and reported to have learned from it, found it useful, and managed to apply it to the peer feedback activity, others students were much more critical. This variability between students was also present in Study 4 but to a lesser extent. For the impressions of the training, the scores ranged from 45 to 100% and the standard deviation was 14.3, while for the perceived impact of the training, the scores ranged from 35 to 100% and the standard deviation was 16.7. This smaller variability in students' reactions to the training supports the hypothesis that students' reactions are less varied when the training is delivered on campus than when it is delivered online.

In Study 4, the reactions of students from the experimental group were compared to those of students from the control group. The reactions measured were the reactions to the comprehensive 2-hour training session (one hour on how to provide peer feedback-information and one hour targeting trust) for students in the experimental condition and the reactions to the 1-hour training session on how to provide peer feedback-information for students in the control group. Students' impressions of the training (general reaction,

satisfaction, and perceived learning) and perceived impact of the training (perceived utility and application to practice) were significantly more positive in the experimental than in the control condition. This means that students preferred the comprehensive training session that also targeted trust over a general training session on how to provide effective feedback-information and that they found it more impactful.

2.3 Impact of the intervention on students' attitudes

After investigating how an intervention was received, it is important to assess if participants' knowledge, attitudes, or skills evolved positively (Blanchard & Thacker, 2012). The interventions developed in this dissertation aimed to affect students' attitudes. More precisely, psychological safety, trust in the self, and trust in the other as an assessor were targeted in Study 3, while Study 4 only aimed to increase trust in the self and trust in the other. The results were identical for both interventions: no significant impact of the intervention was found on students' attitudes.

I see two main explanations for these unexpected results. First, as will be further detailed in the next section (see 3.1), the measurement tools used had some limitations: the reliability of the scales was too low on some occasions and the translated version of the scales was not validated through a rigorous process. Because of these limitations, it is possible that an effect was present but could not be detected. Second, for trust in the self specifically, the intervention may have had a positive impact on students' perceptions of trust, as expected, but also a negative one and these two effects may have cancelled each other out. According to the Dunning-Kruger effect, people with few skills in a particular area tend to overestimate their skills because their lack of skills also prevents them from assessing their skill level accurately. When they develop skills, they are more able to assess themselves accurately, and, paradoxically, judge their skill level lower than before, because they are less likely to overestimate themselves (Kruger & Dunning, 1999). Applied to our results, it could mean that the intervention allowed students to develop their skills, but also to realize that they may have overestimated their skills before the intervention. This would be coherent with the results from a previous study showing that students felt less confident regarding feedback-information provision after receiving training on feedback provision skills, even though the training allowed them to provide better feedback-information (Alqassab et al., 2018a). Obviously, another explanation for the unexpected results could be that both interventions were ineffective in affecting students' attitudes. However, this explanation is unlikely given that the intervention did have a positive impact on students' behaviours outcomes in Study 4.

2.4 Impact of the intervention on students' behaviours

Different behaviours were measured in Study 3 and Study 4. In study 3, the impact of the intervention was investigated on the type of feedback-information students provided to their peers (according to the distinction between verification and elaboration feedback-information proposed by Shute, 2008) and on their performance improvement following the peer feedback-process. Neither of them was significantly different between the control and the

experimental group, which suggested that they were not affected by the intervention. In Study 4, students' feedback-information and productions were also collected but, due to their large number, are yet to be analysed. Therefore, participation rates were the only behaviours investigated. This outcome is essential because if students do not participate in peer feedback activities they will not reap the benefits of the peer feedback-process, no matter how effective and well-designed the activities are. A significant impact of the intervention was present: students from the experimental group participated more in the last peer feedback activity than students in the control group.

In sum, no behavioural effects were found in Study 3, while the intervention affected students' participation rate in peer feedback activities in Study 4. Study 4 was created to overcome limitations in the study design and the intervention of Study 3. Therefore, the difference in findings between the two studies could either be explained by the fact that the intervention in Study 4 was more effective than that in Study 3 or by the fact that the potential efficacy of the training remained veiled due to methodological limitations in Study 3. Moreover, as the behaviours measured were not the same, it is possible that those of Study 3 were more difficult to influence with an intervention than those of Study 4 because their relationships with students' perceptions of trust and psychological safety were more indirect. Indeed, as illustrated in Figure 1, I hypothesised that students would be more engaged in the peer feedback-process if they had higher levels of trust in the self and in the other and that this higher engagement would be reflected by higher quality feedback-information and by a greater performance improvement following the peer feedback-process. The first part of the hypothesis was supported in Study 4: the intervention targeting trust increased students' engagement in peer feedback activities, as measured by the participation rates. This finding adds to a previous study that already showed a link between trust and students' engagement in the peer feedback-process (Zou et al., 2018). The quality of the provided feedback-information and performance improvement are more distal outcomes of the intervention given that they required that students' engagement increased and affected them. Even though it seems logical that an increase in students' engagement will positively affect peer feedback-process outcomes, there is a lack of research on this. Investigating the efficacy of the intervention from Study 4 on the quality of the provided feedback-information and the performance improvement are important next steps. Even though it is already positive to have an intervention that increases students' participation in peer feedback activities, the final goal is to maximise the benefits of the peer feedback-process and, consequently, have a positive impact on students' performance.

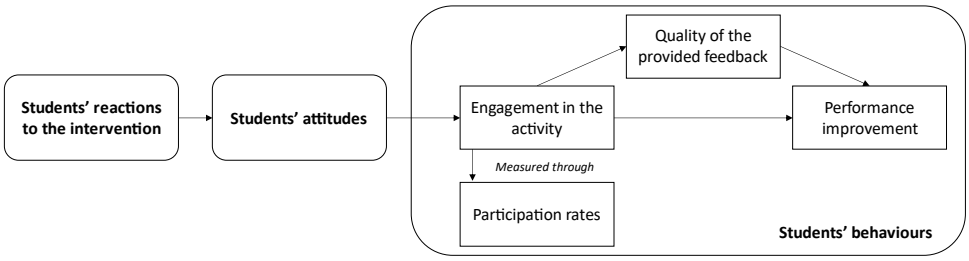


Figure 1. Hypothesized relationships between students' reactions to the interventions, students' attitudes and students' behaviours

The general research question guiding this dissertation was: How can peer feedback activities be designed to ensure that higher education students feel psychologically safe and trust their own and their peers' abilities? To provide a partial answer to this question, it seems that it is beneficial to design peer feedback activities with the presence of a training session targeting trust and a backward evaluation. This result is consistent with previous studies showing that implementing a training session targeting specific interpersonal processes is beneficial (L. Li, 2017), as well as implementing a backward evaluation (M. Kim, 2009; Ng & Yu, 2023).

However, an important question remains unanswered: Is the impact of the intervention on participation rate explained by an increase in students' perceptions of trust in the self and in the other as an assessor? Indeed, even though the intervention positively affected the participation rate, it did not significantly affect students' perceptions of trust. This result is surprising given that attitudinal changes influence how much behaviour can change (Blanchard & Thacker, 2012). It may be that the scales used were not able to detect any effect on trust or that the intervention did not increase students' perceptions of trust but positively affected other important variables. The qualitative findings suggested that two variables may have been affected by the intervention: students' feeling of legitimacy and the meaning given to the peer feedback-process. As such, it is possible that the intervention targeting trust did not increase students' perceptions of trust but raised their awareness about the question of trust in the peer feedback-process. This could have allowed them to feel legitimate in providing peer feedback-information and to see meaning in the peer feedback-process even without fully trusting their peers and themselves. Indeed, according to Rotsaert (2017), an effective way to deal with the interpersonal processes of peer feedback-process is to organise activities that raise students' awareness of these processes, as higher students' awareness should lead to lower negative consequences. This means that our research question should be formulated in the following way: How can peer feedback activities be designed to raise students' awareness concerning interpersonal processes of peer feedback-process and more particularly concerning trust? To this question, I feel confident to answer that it seems effective to design peer feedback activities with (1) a comprehensive training session explaining how to provide effective feedback-information and targeting trust and (2) a backward evaluation.

3. Limitations

Limitations specific to each study have already been discussed in the previous chapters. In this section, I will develop the main limitations of this dissertation as a set of studies, alongside ideas on how future research could overcome these limitations. But before that, I would like to highlight two of its strengths: the use of robust study designs and the emphasis put on ecological validity. Multiple authors have pointed out that there is a predominance of weak research design in the field of peer assessment (e.g., cross-sectional studies without pre-test or control group), which limits the development of the field (e.g., Alqassab et al., 2023; Strijbos & Sluijsmans, 2010; van Zundert et al., 2010). Therefore, in the current dissertation, quasi-experimental designs were used in Study 3 and Study 4, which increased the internal validity of the findings (Creswell, 2012). Moreover, an emphasis was put on the ecological validity of the studies: they were conducted within real educational contexts and in collaboration with higher education professors. This important ecological validity means that the positive effects of the intervention (in Study 4) can be expected if the intervention is re-used by teaching staff implementing peer feedback activities in similar contexts.

3.1 Limitations regarding the variables' conceptualization and measurement

The first limitations concerned our variables of interest: psychological safety, trust in the self, and trust in the other as an assessor. As highlighted in Study 1, the interpersonal processes of peer feedback-process are ill-defined and there is no consensus on which scale to use to measure which interpersonal process. Recently, important efforts have been made to clarify the key concepts in the field of interpersonal processes of peer assessment. For example, Panadero et al. (2023) proposed clear definitions for a large number of interpersonal processes. However, some problems persist such as a lack of concept delimitation and a lack of validated scales to measure them.

Trust in the self, trust in the other, and psychological safety are not immune to these problems. First, the limits of these concepts are blurry, especially for trust. While trust in the self as an assessor and trust in the other as an assessor are sometimes operationalised as a single variable (e.g., L. Cheng et al., 2023; Ching & Hsu, 2016), other researchers consider them as two distinct concepts (e.g., L. Li & Steckelberg, 2013; Ocampo et al., 2023; Panadero, 2016; Rotsaert et al., 2018). In this dissertation, I follow the latter approach as a previous study found they had different predictors (van Gennip, Gijbels et al., 2010). In Study 3, there was no significant correlation between trust in the self and trust in the other (see Appendix A), confirming that these are two different concepts. However, in Study 4, there were moderate to strong correlations between trust in the self and trust in the other measured at the same measurement time (see Appendix B). It is clear that more research is needed to better understand the links between trust in the self and trust in the other.

Second, it would be interesting to determine whether it is useful to distinguish between sub-concepts of trust, and if so, which sub-concepts. In the introduction, I proposed working definitions of trust in the self and trust in the other as an assessor that distinguished between

two aspects: the ability to provide (1) accurate feedback-information, and (2) fair feedback-information. However, these two aspects were not distinguished in the measure of trust in the self and trust in the other as an assessor and in the analyses. Future studies could investigate whether it is more relevant to measure them as a whole or as differentiated sub-concepts. Furthermore, the results of Study 4 suggest that another aspect of trust may be missing: benevolence. Indeed, benevolence appears to be of great importance to some students, which is consistent with the literature on trust as a general concept. In organisational research, three characteristics of the trustee are often identified as responsible for trust: the trustee's ability, the trustee's benevolence, and the trustee's integrity (Mayer et al., 1995; Schoorman et al., 2007). The trustee's ability can be equated with the issue of the accuracy of the feedback-information, while the trustee's integrity can be equated with the issue of the fairness of the feedback-information. Benevolence, the extent to which the trustee has the truster's interest at heart (Mayer et al., 1995), is therefore not included in our working definitions, but may be important to add.

Third, the links between trust and other interpersonal processes should also be investigated. For example, Study 4 revealed that the question of legitimacy was important for students but it is not clear how it is linked to trust in the self (are they two associated but distinct variables? Is one a component of the other?). For trust in the other, an example is the relationship between trust and the perceptions toward the assessor measured in Yu's studies (F. Yu, 2012; F. Yu & Wu, 2011). Given the items used to measure perceptions toward the assessor (e.g., "I felt that my assessors did not know enough on the learned subject", F. Yu, 2012, p. 75), it seems that trust in the other is a component of it, yet it is not explicitly said by the authors.

Finally, this conceptual ambiguity is reinforced by an absence of consensus on how to measure psychological safety, trust in the self, and trust in the other. Seminal works in the field of interpersonal processes of peer feedback-process are the papers published by van Gennip and colleagues (van Gennip et al., 2009; van Gennip, Segers & Tillema, 2010). Based on the literature on team learning in work psychology, these researchers identified four variables (psychological safety, trust, interdependence and value congruency) that could play an important role in peer feedback-process (van Gennip et al., 2009) and developed items to measure these concepts (van Gennip, Segers & Tillema, 2010). These items were created based on previous research (e.g., Edmondson, 1999), adapted to the context of peer assessment, and pretested, but to the best of my knowledge, no information on this pretest or the validity of the scales has been reported. Only the reliability of the scales was provided, which was low for psychological safety and trust in the self as an assessor and good for trust in the peer as an assessor. The development of these scales was the first important step in exploring the interpersonal processes of peer assessment. But, ideally, they should have been followed by studies targeting the validation of the scales, by following a rigorous process of validation as described in the literature (e.g., Boateng et al., 2018). To the best of my knowledge, this has never been done for the three central concepts of this dissertation (psychological safety, trust in the self, and trust in the other). Moreover, other researchers have created their own scales to measure trust and psychological safety, with items

sometimes remarkably distinct from those initially used. For example, in an article by Ching and Hsu (2016), the authors cited van Gennip and colleagues (2009) to introduce trust and psychological safety, yet created different items to measure them (e.g., for psychological safety: “I believe it is important for me to learn how to give constructive feedback to peers“ and, for trust: “I have benefited from providing feedback to peers' work”, Ching & Hsu, 2016, pp. 113-114).

These general limitations of the field had direct impacts on this dissertation. For the literature review (Study 1), as the included studies often did not provide clear definitions of the interpersonal processes studied nor the complete scales used to measure them, we were constrained to use the terms used by the authors of the studies, sometimes without being entirely sure of what they meant by it. This means that two studies whose results were linked to the same variables may not have measured the same concepts. In Study 3 and Study 4, trust in the self, trust in the other, and psychological safety (for Study 3) were measured with a French translation of scales used by Rotsaert et al. (2018). The scales were chosen because (1) the items operationalise the concepts as defined in the introduction of this dissertation, (2) they have been developed to be used with higher education Belgian students, which are the same participants as in this dissertation, and (3) they have shown good reliability (Rotsaert et al., 2018). However, as with other scales measuring these variables, no information has been published on their dimensionality and validity. Moreover, to use them, the scales had to be translated from Dutch to French. Even though a rigorous process to translate the items with forward and backward translation was adopted, as recommended (e.g., Tsang et al., 2017), I did not conduct a validity study before using the translated scales. A validity study is a study that aims to verify if a newly developed scale has good construct, convergent, and discriminant validity (e.g., Roy et al., 2013; Willems et al., 2023). In addition, the reliability of the scale measuring trust in the other as an assessor was low on some occasions (for students in the experimental group in Study 4). Hence, it is possible that students’ perceptions were not adequately measured with the scales used.

To advance the research on psychological safety and trust in the peer feedback-process, it is important to develop validated tools to measure these concepts. This will help avoid situations where it is not possible to disentangle if an absence of significant results means that the hypotheses were not confirmed or that the variables were not adequately captured by the scales. Therefore, I recommend creating scales measuring these concepts by following a rigorous complete process, such as Boateng et al.'s (2018) primer for scale development. The first phase proposed by this primer is the item development phase, which is composed of three steps (identification of domain, item generation, and content validity). For item generation, the best practice is to combine deductive and inductive methods. The literature review of existing scales provides the required theoretical basis for defining the domain while it is possible to identify its manifest forms through qualitative techniques such as interviews and focus groups with the population of interest (Boateng et al., 2018). Conducting this phase with a mixed method for trust and psychological safety would allow researchers to profit from existing scales while ensuring that items reflect students' experiences (e.g., in Study 4 a central concept for students was that of legitimacy, which is not to be found in the research

on peer feedback-process). The next two phases of Boateng et al.'s (2018) primer are scale development (pre-testing of questions, sampling and survey administration, items reduction, extraction of factors) and scale evaluation (tests of dimensionality, tests of reliability, and tests of validity), two steps that are also recommended when translating existing scales (Tsang et al., 2017).

3.2 Limitations regarding the samples and methodologies

The fact that the experimental studies were conducted in real education contexts implied two limitations: it was not possible to use a true experimental design and the sizes of the samples were determined by the number of students participating in the courses.

First, as it is unethical and impractical to use random sampling in natural settings (Cohen et al., 2007), Study 3 and Study 4 relied on quasi-experimental designs. In study 3, a quasi-experiment with pre-and post-test design with two cohorts was adopted (Creswell, 2012). The design is strong if the two cohorts are comparable. In this case, the data for the first cohort were collected just before the COVID-19 outbreak. The COVID-19 pandemic and the restrictions implemented to prevent its spread had important psychological, social, educational, and economic effects, which caused a notable increase in psychological distress and a decrease in the mental well-being of young adults (Wiedemann et al., 2022). The data collection for the second cohort, which was supposed to take place in 2021, was reported to 2022 in the hope of having more comparable groups of students. However, even if most restrictions implemented to prevent the spread of COVID-19 had been lifted by February 2022 in Belgium, some of their negative impacts were still present, such as a decrease in life satisfaction (Neugebauer et al., 2023) or an increase in learning difficulties (Hu et al., 2022). Therefore, it was not surprising to find significant differences between students from both cohorts at the pre-test, which affected the validity of Study 3. To overcome this limitation and maximize the comparability between students in the control and the experimental groups, both groups in Study 4 were composed of students following the course in which the study was implemented during the same year.

A second limitation when conducting studies in natural educational settings is that the sizes of the samples are limited to the number of students enrolled in the course. To have large samples, Study 3 was conducted over two years (which doubled the number of participants) and Study 4 was conducted in a course followed by an large number of students (more than 200). However, in both studies, not all students participated in every activity (e.g., training session, peer feedback activity), even if they were mandatory, which resulted in smaller samples. The quantitative analyses were conducted on 89 students in Study 3 and 94 students in Study 4. Even though samples of this size are the standard in studies on the subject (e.g., the majority of studies included in our literature review included less than 100 participants, see Chapter 2), they may have been too small to detect small effects. As the sample size is a critical factor to take into consideration when designing studies, future studies should determine the adequate number of participants through power analyses (Howell, 2007). However, deciding on which course to conduct a study is not an easy task

and many other factors are important to take into consideration to find an appropriate context (e.g., professors open to collaboration, schedules flexible enough to implement the study, etc.).

Another consequence of having the number of participants limited by the number of students enrolled in the course is that only two conditions were compared in Study 4 to have enough students in each condition. As the intervention tested in Study 4 consisted of two components, it would have been ideal to create four conditions: a control condition, an experimental condition receiving the full intervention, an experimental condition only receiving the training session, and an experimental condition only receiving the backward evaluation. Comparing these four groups would allow us to investigate if both components are required for the intervention to be effective and if the combined effect of both components is more important than the sum of the effects of each component taken separately. I hypothesize that the training session would be effective even without the backward evaluation, but that its effectiveness would be enhanced by the backward evaluation. Concerning the backward evaluation, I hypothesize that it would not have a positive effect without the training, as students may not understand the value of it if their awareness is not first raised through the training session. It is clear that further research on this is warranted.

4. Perspective for further fundamental research

Besides overcoming the limitations of the current dissertation, future research could also extend the findings of the dissertation by exploring complementary areas of research. In this section, I present six main perspectives for future research: implementing the intervention in other contexts to investigate its generalizability, emphasizing the processing phase of the peer feedback-process in the intervention, implementing the intervention in the context of dialogic peer feedback-process, investigating if it is more effective to target several interpersonal processes, investigating how trust in the self and trust in the peer evolve, and considering the role of students characteristics.

4.1 Implementing the intervention in other contexts

The studies conducted in this dissertation have all been implemented in the same Belgian university, with students in motricity sciences for the first version of the intervention (Study 2 and Study 3) and students in educational sciences for the improved version of the intervention (Study 4). This raises the question of the generalizability of the results to (1) students in different majors, (2) students from different countries and cultures, and (3) students from different levels of education. In any case, to strengthen the effectiveness of the intervention, I recommend adapting it to each group of students, for example by asking students about their thoughts on the peer feedback-process (Dusenberry & Robinson, 2020).

First, future studies should investigate if the intervention is effective with students from different majors. The improved version of the intervention was only implemented in a course on the psychology of learning attached to the master in educational sciences. Previous research suggests that students in different majors have different perceptions regarding the

peer feedback-process (Praver et al., 2011; Wen & Tsai, 2006, Zou et al., 2018), which I also noticed when asking students about their thoughts on the peer feedback-process. Indeed, there were important differences between students in motricity sciences and students in educational sciences (see section 2.1). It is therefore possible that the intervention may not be effective when using it with students from different majors, especially given that the intervention took place in a course on the psychology of learning. The content of the course may have supported the impact of the intervention on students' participation in peer feedback activities as students studied theories and concepts that can explain why the peer feedback-process is useful (e.g., socio-constructivism).

Second, future studies should investigate if the intervention is effective in other cultures and countries. It has been argued that culture plays a role in students' perceptions of the peer feedback-process and that students from different cultures have different concerns linked to the peer feedback-process (e.g., Fithriani, 2018; F. Zhang et al., 2023). Moreover, besides students' culture, it is possible that organisational aspects of higher education that vary from one country to another also affect students' experience of the peer feedback-process. For example, in a study conducted in Australia, some students explained that it was not their role to provide feedback-information as teachers were paid to do so (Wilson et al., 2015). I never found this idea in the interviews with students, which may be explained by the fact that university tuition fees are considerably less expensive in Belgium than in Australia (OECD, 2021).

Third, it would be interesting to investigate if the intervention could be effective at other educational levels than higher education, the specific focus of this dissertation. This would require adapting the intervention to the age of students (e.g., changing the formulation of the backward evaluations' questions, adapting the roleplays) and investigating how the students react to the training, but also to the backward evaluation. Anonymous backward evaluation may not be adapted to primary students as a study revealed that students can write harsh unjustified comments to their peers when anonymous (F. Yu & Wu, 2011). Hence, a face-to-face non-anonymous exchange on the usefulness of the received feedback may be preferable with younger students.

4.2 Emphasizing the feedback processing phase

The peer feedback-process includes two phases: feedback-providing, when students assess their peers' work and provide feedback-information on it, and feedback-processing, when students make sense of the received feedback-information (Aben et al., 2019; Kollar & Fischer, 2010). Like feedback providing, feedback processing is an important activity that needs to be supported by an adequate learning context and appropriate scaffolding as students tend to struggle to make sense of their peers' feedback-information and to efficiently revise their work based on it (Wichmann et al., 2018). Moreover, interpersonal processes play a role in both phases (Aben et al., 2019). For trust, for example, trust in the self should play an important role in the provision phase, while trust in the other should be more important in the processing phase.

Even though both the feedback-providing and feedback-processing phases were taken into account in this dissertation, the emphasis was put on the first phase. The interventions were provided in a context where all students received a general training session on how to provide effective peer feedback-information, thus supporting the feedback-providing phase. Moreover, the backward evaluation that was included in the improved version of the intervention principally supports the feedback-providing phase as it provides information to students on the quality of the feedback-information they provided and suggestions to improve them.

Future studies could investigate if the impact of the intervention is strengthened when implemented in a context where there is also adequate scaffolding for the feedback processing phase. Interventions to support the feedback-processing already exist. For example, Wichmann and colleagues (2018) developed a document to encourage students to reflect on and make sense of the received feedback-information. Students wrote all the feedback-information they received and indicated whether they understood the comment, agreed with it, planned to use it, and how they could improve their work based on it. Students who used this support made fewer errors in their final work and rejected more incorrect feedback-information than students from a control condition. Moreover, peer feedback processing could also be supported through general scaffolds that are not specific to the peer feedback-process, such as the Developing Engagement with Feedback Toolkit (Winstone & Nash, 2016). This toolkit – composed of a feedback guide for students, feedback workshop activities that can be implemented by teachers and a feedback portfolio – was developed so that students could overcome barriers preventing them from using feedback-information provided by teachers (Winstone, Nash, Rowntree & Parker, 2017). Besides a few aspects of the toolkit (e.g., a guide provided to students so that they can understand the vocabulary often used by their teachers in feedback-information), it should also be beneficial for processing peer feedback-information.

4.3 Implementing the intervention in the context of more dialogic peer feedback

As explained in Chapter 1, the two main benefits of the peer feedback-process are that it helps foster students' learning (e.g., Double et al., 2020) and contributes to the development of feedback literacy skills (e.g., Carless et al., 2011). According to Carless & Boud (2018), if the peer feedback-process can strengthen students' feedback literacy, it is because it provides opportunities for students to take responsibility, dialogue with peers around feedback-information, make judgements, and act upon feedback-information. The second point assumes that students are offered the possibility to dialogue with each other, for example by having a verbal exchange about the written comments. The presence of dialogue around feedback-information is recognized as a way to increase the feedback-process effectiveness, whether the feedback-information is provided by teachers (Nicol & Macfarlane-Dick, 2006) or by peers (Zhu & Carless, 2018).

In the empirical studies of this dissertation, the opportunities for students to dialogue around peer feedback-information were absent (in Studies 2 and 3) or limited (in Study 4). Students' interactions could be developed by allowing students to not only react on the feedback they received through backward evaluation, as was the case in Study 4 but also through what Zhao et al. (2024) have called re-feedback. Re-feedback comes after backward evaluation and allows feedback providers to respond to the backward evaluation to clarify misunderstandings, confirm or refine previous comments or propose new ones (Zhao et al., 2024). However, this type of dialogue may not be as beneficial as a verbal exchange between students, which is not possible for anonymous online feedback. The choice for such anonymous online peer feedback activities was made based on unstructured interviews I conducted with university professors at UCLouvain at the beginning of my dissertation to have an idea of the obstacles they encountered with peer feedback and to identify potential contexts in which to conduct my studies. Most of the university professors were very reluctant to the idea of having students' identities revealed during the peer feedback-process and told me they would not want to have a study on the peer feedback-process organised in their course if the study implied that peer feedback activities had to be organised non-anonymously. Therefore, to be able to find adequate contexts to conduct the studies and to develop an intervention that would be relevant for professors at UCLouvain, I decided to design anonymous peer feedback activities in each study, which precluded the possibility of verbal exchanges around feedback-information.

Therefore, future studies could adapt the intervention to a context where students are non-anonymous and can dialogue around feedback-information, then investigate the impact of this adapted intervention on students' feedback literacy. For example, the training could be adapted to be focused on oral feedback-information or the questions used in the backward evaluation could be transformed into a guide that could support a face-to-face discussion between students around the usefulness of the feedback-information received/provided. A study conducted by L. Cheng et al. (2023) found that students' attitudes toward the climate of trust (a concept similar to trust in the self and trust in the other) were more positive when students received a detailed regulation script with prompts to support peer feedback dialogue. It would be interesting to investigate if it is beneficial to combine the intervention from Study 4 with the regulation script developed by L. Cheng et al. (2023).

4.4 Investigating if it is more effective to target several interpersonal processes

A decision taken to improve the intervention between Study 3 and Study 4 was to target the intervention on trust only to have more time dedicated to it. Psychological safety was dismissed because Study 2 and Study 3 revealed that students were more concerned about trust and because training may not be an effective way to foster psychological safety. Indeed, it has been argued that psychological safety is difficult to foster (Hunt et al., 2021) and our study is not the first to fail to significantly affect students' level of psychological safety through training (Dusenberry & Robinson, 2020). This decision raises an important question: is trust a process that can be isolated from other interpersonal processes? Even if this study

is not the only one that focuses specifically on trust (e.g., L. Cheng et al., 2023), interpersonal processes of peer feedback-process are often studied as a multi-faceted experience including multiple interpersonal processes (e.g., Rotsaert et al., 2018; van Heerden & Bharuthram, 2021; Vanderhoven et al., 2015). However, the correlations between the different interpersonal processes are rarely reported in the literature. In Study 3, psychological safety was correlated with trust in the self (moderately on the pre-test and strongly on the post-test) but not with trust in the other (see Appendix A). The opposite pattern has been found in two other studies: psychological safety correlated moderately with trust in the other but not with trust in the self (van Gennip, Gijbels et al., 2010; van Gennip, Segers & Tillema, 2010). This illustrates that the link between the different interpersonal processes is not clear yet and should be studied more in-depth. If future studies reveal important associations between trust and other interpersonal processes of peer feedback-process, it should be interesting to compare the efficacy of our intervention specifically targeting trust with that of an intervention targeting trust as well as other associated interpersonal processes. It may be possible that if another interpersonal process associated with trust is not also targeted, it will hamper the effectiveness of the intervention (e.g., if a minimum level of psychological safety is required to achieve a desired level of trust).

4.5 Investigating how trust evolves

One of the explanations proposed for the absence of a significant effect of the intervention in Study 3 was that students' levels of trust were already quite high at the pre-test, and, therefore, may have been more difficult to influence (Ritter & Schooler, 2001; Rosenbaum et al., 2001). This raises two important questions: (1) what kind of learning curve does trust in the self and in the peer follow, and (2) is it worth it to devote time to increase trust when students already have high levels of it? In Study 3, we hypothesized that trust in the self and in the peer may follow a diminishing-returns curve as most intellectual and perceptual motor skills do (Rosenbaum et al., 2021): slowly evolving at the beginning, increasing exponentially, and then increasing very slowly when approaching mastery (Ritter & Schooler, 2001). However, trust in the self and in the other could follow other curves, such as an all-or-nothing evolution (students would either trust or mistrust their and their peers' abilities). Future research could investigate how trust in the self and in the peer evolves. They could also investigate what levels of trust in the self and in the other can be considered optimal levels, which would allow teaching staff to identify when it is relevant to implement an intervention targeting trust.

4.6 Considering the role of students' characteristics

In the three empirical studies of this dissertation, the analyses were conducted on students as a group, without considering their individual characteristics. However, the results of Study 2 suggest that there is considerable variability in students' perceptions of the intervention. Two characteristics may be particularly important to consider in future studies: students' gender and their conceptions of and approaches to learning.

Previous studies investigating the role of gender on trust in the self and in the other as an assessor have found mixed results. While one study found no differences in levels of trust in the self and in the other between men and women (Rotsaert et al., 2017), another study found that, without training, men have less trust in the self than women, but only when assessing productions from women (Ocampo et al., 2023). Inconsistent results have also been found for other interpersonal processes, such as comfort, with some studies finding that women tend to have more interpersonal concerns than men (Wen & Tsai, 2006; Zou et al., 2018), while another study found the opposite pattern (Ocampo et al., 2023). It would therefore be interesting to investigate whether men and women respond similarly to our interventions.

Other variables worthy of further investigation are students' conceptions of and approaches to learning. Conceptions of learning refer to what students think they are learning and can either be fragmented, meaning that students have no understanding of the relationship between the learning environment and their learning, or they can be cohesive, meaning that students have a better understanding of this relationship (Yang & Tsai). Students can also adopt different approaches to learning: they can adopt surface approaches, which means that they engage with the learning environment in a way that promotes reproduction, or deep approaches, which means that they engage with the learning environment in a way that promotes real understanding (Yang & Tsai, 2010). Students with more cohesive conceptions and deeper approaches to learning make more progress through peer assessment than students with fragmented conceptions and superficial approaches to learning (Yang & Tsai, 2010). In addition, students with cohesive conceptions of learning tend to have high levels of trust in the self and in the peer as an assessor, whereas students with deep approaches to learning tend to have high levels of trust in the self. Trust in the other and approaches to learning are not associated (K. Cheng & Tsai, 2012).

These findings suggest that conceptions of and approaches to learning may be important to consider in order to better understand the impact of our interventions. This could be explored in two steps. First, it would be interesting to conduct a phenomenographic study to identify students' conceptions of and approaches to learning in specific groups. Indeed, within each of the main categories (cohesive or fragmented conceptions and deep or surface approach), there are different conceptions and approaches that are not identical across diverse cultural and educational contexts (Paakari et al., 2011). Phenomenography is often used to study learning conceptions as it is a qualitative research method that allows for a collective analysis of individual experiences (Paakari et al., 2011; Yang & Tsai, 2010). Once the conceptions and approaches of a particular group of students have been identified, a second step could be to investigate whether these conceptions and approaches influence the way students experience the peer feedback process and the intervention.

5. Implications for educational practice

The findings of the current dissertation have implications for teaching staff implementing peer feedback activities. As discussed above (see 4.1), the studies were conducted in higher education, and their results may not be generalizable to other levels of education. Therefore,

although the implications could also be of interest to educators from other levels, they should be taken with more caution. In this section, the implications will be presented in the form of two main recommendations for designing peer feedback activities.

5.1 The importance of carefully designing peer feedback activities

A first important implication for practice is that teaching staff should carefully design peer feedback activities based on the characteristics of the group of students that will participate in them. Indeed, the findings from the experimental studies supported previous results showing that students' perceptions of the peer feedback-process were influenced by their major (Praver et al., 2011; Wen & Tsai, 2006). Educational sciences students (Study 4) were more resistant to the peer feedback-process than students in motricity sciences (Study 2 and Study 3) and had different concerns. Therefore, the design of peer feedback activities could be adapted to students' perceptions and concerns. For privacy specifically, the literature review (Study 1) revealed that this design element can have a positive impact on some interpersonal processes (e.g., students' feeling of comfort, S. Kim, 2019; Lim et al., 2021; Raes et al., 2015; Su, 2023) and a negative impact on others (e.g., perceptions of fairness, Açkay et al., 2021; Lin, 2018). Taken together, these results suggest that peer feedback activities should be adapted to the perceptions of the group of students that will participate in them. For example, if students are uncomfortable with the peer feedback-process and fear their peers' disapproval if providing critical feedback-information, it should be better to design an anonymous online peer feedback activity (e.g., Su, 2023; Vanderhoven et al., 2015). On the contrary, a face-to-face non-anonymous peer feedback-process should be preferable if students' main concerns are linked to the perceived fairness of peer feedback-process or the importance of face-saving (e.g., Akcay et al., 2021; S. Kim & Lan, 2021).

Hence, teaching staff first need to understand students' previous experiences, their perceptions, and their concerns. For small groups of students, I recommend to discuss with the students to gauge their perceptions. For larger groups, I recommend using interactive online tools, such as Wooclap, as was done in studies 2, 3, and 4, to ask students about the perceived advantages and disadvantages of peer feedback and about their concerns. Then, peer feedback activities can be adapted to their needs. As students' major seems to be an important determinant of their perceptions (e.g., Praver et al., 2011; Zou et al., 2018), the perceptions of students enrolled in the same course unit will likely be similar from one year to another. Hence, it is likely that teachers will not have to redesign their peer feedback activities each year.

Even though privacy is the most-studied design element, and the one professors seem to be the most preoccupied with, it is not the only important one. Teaching staff need to make other important decisions when designing peer feedback activities, such as the type of feedback-information students are expected to provide, the frequency of peer feedback, the relationship between peer and teacher feedback or the matching of assessors and assessees (Alqassab et al., 2023). While some elements should only have a positive effect (e.g., the presence of training, van Zundert et al., 2010), other less-studied design elements could also

have a mixed impact. Hence, the choices made when designing peer feedback activities could positively affect one interpersonal process while negatively affecting another interpersonal process and students' concerns should also be considered.

Given that the research is not advanced enough to know what are the positive and negative impacts of each instructional design element on interpersonal processes, a solution could be to directly ask students about their preferences. For example, as it is not clear how the matching of assessors and assessees affects interpersonal processes, teachers could ask students if they have any preferences regarding that. However, the design of peer feedback activities should not only be based on students' preferences but first and foremost on the consequences it will have on their learning.

5.2 The importance of providing peer feedback training

It was found in the literature review (Study 1) that the presence of a training session can positively affect interpersonal processes, whether it was a general training session on how to provide feedback (Ocampo et al., 2023; Parikh & Sheehan, 2016) or a specific session targeting interpersonal processes (L. Li, 2017). Except for Study 3 which did not find any effect (either positive or negative) of a training session, all the studies investigating the impact of training on interpersonal processes have found positive results. These findings add to the already extensive evidence that it is preferable to train students before they participate in peer feedback activities. Indeed, previous reviews already showed the positive impacts of training on other outcomes such as the quality of the provided feedback or students' performance improvement (H. Li et al., 2020; van Zundert et al., 2010). Therefore, I feel confident to recommend every teaching staff who implements peer feedback activities to provide training to students.

Moreover, findings from Study 4 suggest that a comprehensive training session composed of a first part on how to provide peer feedback-information and another part targeting trust is more beneficial than a shorter training session with only the former. Students who received the comprehensive training session reported being more satisfied with it, finding it more useful, and having applied it more to the peer feedback-process than students who only received a training session teaching them how to provide effective feedback to their peers. Moreover, the comprehensive training session, in combination with a backward evaluation, had a positive impact on students' rate of participation in the following peer feedback activity. Therefore, I recommend implementing the intervention developed in Study 4 (comprehensive training session and backward evaluation), especially for groups of students who have little trust in their own or their peers' abilities and who are reluctant to participate in peer feedback activities. All the training material is available upon request.

6. Perspectives for research with high relevance for educational practice

Besides the two implications directly stemming from the main findings of the dissertation discussed in the previous section, possible directions for practice also arise from this

dissertation. As these possible directions are currently understudied, further research is required before they can be translated into clear implications for practice. In this section, I will develop perspectives for further research associated with two main directions for practice: (1) adopting a curriculum perspective for the peer feedback-process and (2) training teaching staff.

6.1 Adopting a curriculum perspective for (peer) feedback process

In each empirical study of this dissertation, peer feedback activities and interventions to support them were implemented in a single course unit, without adopting a curriculum perspective. This is the common way of conducting research on the peer feedback-process. Indeed, all the studies included in Study 1 also adopted a similar perspective by focusing on single course units. However, I believe that adopting a curriculum perspective makes more sense for two reasons: it is a way of dealing with the fact that previous experiences seem to affect students' perceptions of the peer feedback-process and it is coherent with the literature on the development of feedback skills.

First, findings from Study 4 suggest that students' previous experiences with peer feedback activities can shape their current perceptions. Most students were, at first, reluctant to participate in peer feedback activities and saw far more disadvantages than advantages to it. The interviews conducted with sixteen students revealed that their previous experiences may explain this mostly negative view of the peer feedback-process. Expect for Aaron and Eva, all of these students had participated in the same peer feedback activities the year before, and they mostly reported negative experiences with these activities (for example Clémentine explained "*So last year I was always very frustrated, at least with the feedback I got, because on the same text, I could get "Wow, excellent!" and "I didn't understand a word you wrote" or "I haven't got the time".").* This previous experience with the peer feedback-process seems to have affected most students' views on the peer feedback-process (e.g., Amélia said "*When I found out that there would still be feedback, as last year's experience hadn't been particularly useful to me... well, I wasn't too enthusiastic [laughs]*."). These previous experiences seem to have prevented some students from fully benefiting from the intervention. For example, Eden is one of the two students whose perspective did not seem to have evolved even though he received the intervention. When he was asked how he experienced the training, he explained what he learned before adding "*But in the back of my mind, I've not been a big fan of [peer] feedback since last year.*". Moreover, half of the students spontaneously said in the interviews that they thought the training session should have been implemented earlier, and that they did not understand why they did not receive it the year before, the first time they participated in a peer feedback activity. This impact of previous experiences could be dealt with by adopting a curriculum perspective, instead of a course unit perspective. It would mean that each peer feedback activity would be designed in line with the other activities that students have already participated in or will participate in during their bachelor/master and that the scaffold proposed to students would be adapted to their level of experience.

Adopting a curriculum perspective for the design of peer feedback activities is in line with the literature on the development of feedback literacy skills (e.g., Boud & Molloy, 2013; Carless & Winstone, 2023: Malecka et al., 2020). As the ability to use feedback is a complex skill, it cannot sufficiently be developed in a single course unit but needs to be practised in different settings over time to improve. Therefore, activities that support the development of feedback literacy, such as peer feedback activities, must be embedded within the practices of various courses in the curriculum (Malecka et al., 2020). Implementing multiple peer feedback activities can take time, especially if teachers need to coordinate to organise well-designed activities that follow a logical progression from the first course unit to the next. However, as some researchers have argued (e.g., Malecka et al., 2020; Winstone & Nash, 2016), time dedicated to developing students' feedback skills and engagement with feedback is time well spent. Indeed, teaching staff often spend a lot of time and effort providing feedback-information to students, sometimes without it being beneficial as students do not always know how to act upon this feedback (Winstone & Nash, 2016). If the feedback-process is seen as a shared responsibility between students and teachers, it makes more sense to spend time fostering students' engagement with feedback than to spend time focusing on feedback delivery (Winstone & Nash, 2016). Moreover, developing students' feedback literacy can be considered an important outcome of higher education in itself as these skills are important for lifelong learning (Tai et al., 2018) and as providing feedback to colleagues is considered an important skill in some professions (e.g., medical practitioners, Frank et al., 2015). In this sense, the feedback-process should be seen as time dedicated to pursuing important learning objectives rather than time taken away from them (Malecka et al., 2020).

Therefore, further research could be conducted to investigate the impact of design elements of peer feedback activities on interpersonal processes, but with a curriculum perspective. For example, it would be interesting to compare the effectiveness of the intervention developed in Study 4 when it is provided to students who have never participated in peer feedback activities, students who mostly have positive experiences with peer feedback activities, and students who mostly have negative experiences with peer feedback activities. It would also be interesting to investigate how the intervention could be adapted from a curriculum perspective. The comprehensive training session would probably be more relevant to provide before the first peer feedback activity students participated in. It remains to be seen when would be the best time to implement the backward evaluation and how the peer feedback-process could be scaffolded the following times it is used, to help students keep developing their feedback skills. Inspirations could also come from other pedagogical techniques than the peer feedback-process. For example, at Maastricht University, a university that heavily relies on problem-based learning, all first-year students receive a training on problem-based learning when they arrive at the university (Servant-Miklos, 2019). Something similar could be organised for the peer feedback-process if the teaching staff decides that peer feedback activities are an important aspect of students' learning at their university.

6.2 Training teaching staff

Another important perspective for practice requiring further research is the training of teaching staff. Indeed, the findings of this current dissertation highlight the fact that implementing effective peer feedback activities is not a straightforward process: teaching staff require skills to carefully design the activities, adapt them to students' skills and needs, provide training, supervise the activities, etc. This raises the question of whether university teachers have enough feedback literacy to carry it out.

The concept of teacher feedback literacy is an emerging concept that is important given that students' capacities to reap benefits from feedback-process depend on the environments created by teachers (Carless & Winstone, 2023). According to the framework developed by Carless and Winstone (2023), teacher feedback literacy involves a design dimension (teachers must be able to design assessment environments that facilitate effective feedback-process), a relational dimension (teachers must be able to deal with the communication and relational aspects of feedback-process), and a pragmatic dimension (teachers must be able to compromise between what is ideal and what is realistic with the limited resources and time).

To the best of my knowledge, there are no studies investigating if higher education teaching staff have enough feedback literacy to optimally organise peer feedback activities. However, it is admitted that teacher feedback literacy requires professional development opportunities (Panadero & Lipnevich, 2022; Winstone & Carless, 2023). Given that, in Belgium, as in most countries, no formal pedagogical training is required to become a university professor (Demougeot-Lebel et al., 2012) and that the emphasis on research leaves little time for professors to participate in professional development courses on the pedagogical aspect of their profession (Rege-Colet & Berthiaume, 2015), it can be hypothesized that most of them have low levels of feedback literacy. For example, a study conducted on Spanish teachers revealed that only a minority of higher education teachers in the sample had followed formative assessment courses or even any pedagogical general courses, which, according to the author of the study, could explain why they used peer feedback-process less frequently than primary and secondary teachers (Panadero & Brown, 2017). Future studies should be conducted to identify higher education teachers' needs concerning (peer) feedback and to investigate how these needs could be supported.

7. To conclude

Despite its important advantages, the use of the peer feedback-process in higher education is not without challenges. Interpersonal processes, such as trust or psychological safety play an important role and can hinder the effectiveness of the peer feedback-process if they are not taken into account. Therefore, this dissertation strives to better understand how peer feedback activities can be designed to ensure that students feel psychologically safe and have trust in their and their peers' abilities. Study 1 presented a literature review examining the impact of instructional design elements on interpersonal processes of peer feedback-process. It revealed that the two design elements for which there are enough quasi-experimental studies to conclude anything are privacy, whose effects are mixed, and training, which has

positive effects on several interpersonal processes. Therefore, we decided to create a training session targeting trust and psychological safety (Study 2) and investigated its efficacy through a pre-and post-test design with two cohorts of students (Study 3). No significant effect of the intervention was found, neither on students' levels of psychological safety, trust in the self, and trust in the other, nor on the type of feedback they provided or the performance improvement following the peer feedback-process. This absence of effects could be explained by the impact that the COVID-19 pandemic had on the study, making the control and experimental group non-comparable. An improved version of the intervention was created (with a training session targeting only trust and a backward evaluation) and its efficacy was investigated in Study 4. Even though no significant impact was found on students' perceptions of trust in the self and in the other, the intervention positively affected students' rate of participation in a following peer feedback activity. More research is needed to understand what exactly the intervention affected (e.g., it may be that it increased students' feeling of legitimacy or the meaning they attributed to the peer feedback-process).

I hope that this dissertation will be of use to both researchers and practitioners. For research, the literature review identified important gaps and perspectives for future research, while the experimental studies provided insights into the role of backward evaluation and training for dealing with interpersonal processes. For practice, an intervention that efficiently increases students' participation in peer feedback activities was created and key recommendations for the implementation of peer feedback in higher education were provided.

8. Appendices

8.1 Appendix A: Correlation matrix for Study 3

Table 1. Correlation matrix for Study 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Psychological safety pre-test																	
2. Psychological safety post-test	.76**																
3. Trust self pre-test	.40**	.37**															
4. Trust self post-test	.53**	.68**	.63**														
5. Trust other pre-test	.12	.05	-.06	-.02													
6. Trust other post-test	.29*	.18	-.10	.13	.65**												
7. Grade first draft	.07	.06	.01	.02	.01	.18											
8. Grade final version	.02	-.05	-.08	-.05	.10	.15	.74**										
9. Improvement global grade	-.09	-.18	-.02	-.12	.06	-.12	-.39**	.34**									
10. Improvement grade "diagram"	-.18	-.04	-.08	.00	-.08	-.15	-.24*	.22*	.67**								
11. Improvement grade "relevance of the situation"	.04	.03	.12	.09	.40**	.20	-.18	-.01	.19	.01							
12. Improvement grade "instructions"	-.05	-.16	-.12	-.16	-.04	-.26	-.14	.28**	.51**	.15	-.11						
13. Improvement grade "success criteria"	.13	-.12	.13	-.10	.06	.01	-.26*	.14	.53**	.01	.04	.11					
14. Verification feedback	.00	-.05	-.05	-.08	-.03	.04	.16	.10	-.08	-.04	-.11	.02	-.14				
15. Correction feedback	-.03	-.07	-.11	-.15	-.10	.00	.23*	.17	-.06	-.01	-.19	-.01	-.03	.57**			
16. Justification feedback	.05	.03	.00	-.02	.18	.24*	.17	.12	-.02	.03	-.02	-.06	.04	.36**	.29**		
17. Question feedback	-.12	.03	.02	.10	-.11	.13	.25*	.09	-.24*	-.14	-.15	-.07	-.20	.10	.21*	.11	
18. Suggestion feedback	.04	.00	-.14	-.12	-.02	.18	.15	.16	.03	.09	-.17	.09	.02	.30**	.19	.37**	.19

Note. * = $p < .05$; ** = $p < .01$

8.2 Appendix B: Correlation matrix for study 4

Table 2. Correlation matrix for Study 4

	Trust self T1	Trust self T2	Trust self T3	Trust other T1	Trust other T2	Trust other T3	Impression of the training	Perceived impact of the training
1. Trust self T1								
2. Trust self T2	.45**							
3. Trust self T3	.53**	.71**						
4. Trust other T1	.55**	.18	.31*					
5. Trust other T2	.36**	.49**	.38**	.37**				
6. Trust other T3	.27*	.22	.48**	.62**	.45**			
7. Impression of the training	-.17	.25**	.15	-.06	.22*	.02		
8. Perceived impact of the training	-.13	.28*	.25*	.08	.07	.12	.75**	
9. Final exam grade	.09	.11	.01	.03	.10	.01	.11	-.15

Note. * = p < .05; ** = p < .01

References

References

Aben, J. E. J., Dingyloudi, F., Timmermans, A. C., & Strijbos, J. (2019). Embracing errors for learning: Intrapersonal and interpersonal factors in feedback provision and processing in dyadic interactions. *The impact of feedback in higher education: Improving assessment outcomes for learners* (pp. 107-125). https://doi.org/10.1007/978-3-030-25112-3_7

Aben, J., Timmermans, A. C., Dingyloudi, F., Mascareño Lara, M., & 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. *Learning and Individual Differences, 97*, 102175. https://doi.org/10.1016/j.lindif.2022.102175

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment and Evaluation in Higher Education, 42*(2), 252-265. https://doi.org/10.1080/02602938.2015.1102863

*Akçay, A. O., Güven, U., & Karahan, E. (2021). The Views of Pre-Service Elementary Teachers About Online and Traditional Peer Assessment. *International Journal of Assessment Tools in Education, 8*(2), 409-422. https://doi.org/10.21449/ijate.762104

Al Qunayeer, H. S. (2020). Supporting postgraduates in research proposals through peer feedback in a malaysian university. *Journal of further and Higher Education, 44*(7), 956-970. https://doi.org/10.1080/0309877X.2019.1627299

Alger, M., & 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*, 102716-102716. https://doi.org/10.1016/j.system.2021.102716

Alharbi, M. A., & Alqefari, A. N. (2022). Students' uptake and perspectives on teacher and peer feedback on written assignments. *Learning & Teaching in Higher Education Gulf Perspectives, 18*(2), 107-118. https://doi.org/10.1108/LTHE-02-2021-0015

Alqassab, M., Strijbos, J., Panadero, E., Ruiz, J. F., Warrens, M., & To, J. (2023). A systematic review of peer assessment design elements. *Educational Psychology Review, 35*(1), 18. https://doi.org/10.1007/s10648-023-09723-7

Alqassab, M., Strijbos, J., & Ufer, S. (2018a). Training peer-feedback skills on geometric construction tasks: Role of domain knowledge and peer-feedback levels. *European Journal of Psychology of Education, 33*(1), 11-30. https://doi.org/10.1007/s10212-017-0342-0

Alqassab, M., Strijbos, J., & Ufer, S. (2018b). The impact of peer solution quality on peer-feedback provision on geometry proofs: Evidence from eye-movement analysis. *Learning and Instruction, 58*, 182-192. https://doi.org/10.1016/j.learninstruc.2018.07.003

Berndt, M., Strijbos, J.-W., & Fischer, F. (2018). Effects of written peer-feedback content and sender's competence on perceptions, performance, and mindful cognitive processing. *European Journal of Psychology of Education, 33*(1), 31-49. doi:10.1007/s10212-017-0343-z

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347-364. https://doi.org/10.1007/BF00138871

Black, N., Mullan, B., & Sharpe, L. (2016). Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. *Health Psychology Review, 10*(3), 341-357. https://doi.org/10.1080/17437199.2016.1168268

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Blanchard, P. N., & Thacker, J. W. (2012). *Effective training: Systems, strategies, and practices* (5th ed.). Pearson/Prentice Hall.

Bloxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets. *Assessment and Evaluation in Higher Education, 35*(3), 291-300. https://doi.org/10.1080/02602931003650045

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149-149. https://doi.org/10.3389/fpubh.2018.00149

Bolinger, A. R., & Stanton, J. V. (2020). *Role-play simulations*. Elgar

Booth, A., Sutton, A., & Clowes, M. (2016). *Systematic approaches to a successful literature review* (Second ed.). SAGE.

Boud, D., Dawson, P., Yan, Z., Lipnevich, A., Tai, J., & Mahoney, P. (2022). *Measuring learners' capabilities to engage in feedback: the Feedback Literacy Scale* [Paper presentation]. EARLI SIG 1 & 4 joint conference 2022, Cadiz, Spain.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education, 38*(6), 698-712. https://doi.org/10.1080/02602938.2012.691462

Bouwer, R., & Dirkx, K. (2023). The eye-mind of processing written feedback: Unraveling how students read and use feedback for revision. *Learning and Instruction, 85*. https://doi.org/10.1016/j.learninstruc.2023.101745

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews, 6*(1), 245-245. https://doi.org/10.1186/s13643-017-0644-y

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Carless, D. (2013). Trust and its role in facilitatic dialogic feedback. In Boud, D., & Molloy, E. (Eds.). *Feedback in higher and professional education*. Routledge.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education, 43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395-407. https://doi.org/10.1080/03075071003642449

Carless, D., & Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education, 28*(1), 150-163. https://doi.org/10.1080/13562517.2020.1782372

Carvalho, A. (2013). Students' perceptions of fairness in peer assessment: Evidence from a problem-based learning course. *Teaching in Higher Education, 18*(5), 491-505. https://doi.org/10.1080/13562517.2012.753051

Çevik, Y. D. (2015). Assessor or assessee? Investigating the differential effects of online peer assessment roles in the development of students' problem-solving skills. *Computers in Human Behavior, 52*, 250-258. https://doi.org/10.1016/j.chb.2015.05.056

Cheng, K., & Tsai, C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology, 28*(4), 599-618. https://doi.org/10.14742/ajet.830

*Cheng, L., Li, Y., Su, Y., & Gao, L. (2023). Effect of regulation scripts for dialogic peer assessment on feedback quality, critical thinking and climate of trust. *Assessment and Evaluation in Higher Education, 48*(4), 451-463. https://doi.org/10.1080/02602938.2022.2092068

*Ching, Y., & Hsu, Y. (2016). Learners' interpersonal beliefs and generated feedback in an online role-playing peer- feedback activity: An exploratory study. *International Review of Research in Open and Distance Learning, 17*(2), 105-122. https://doi.org/10.19173/irrodl.v17i2.2221

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge/Taylor & Francis Group.

Colognesi, S., & Deschepper, C. (2019). La relecture collaborative comme levier de réécriture et de soutien aux corrections des textes. *Le Français Aujourd'Hui, 203*(4), 63-72. https://doi.org/10.3917/lfa.203.0063

Creswell, J. W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th ed.). Pearson education.

De Ketele, J., Chastrette, M., & Cros, D. (2007). *Guide du formateur* (3e éd. ed.). De Boeck.

De Stobbeleir, K., Ashford, S., & Zhang, C. (2020). Shifting focus: Antecedents and outcomes of proactive feedback seeking from peers. *Human Relations, 73*(3), 303-325. https://doi.org/10.1177/0018726719828448

Demougeot-Lebel, J., Ricci, J., & Romainville, M. (2012). Regards croisés sur des pratiques de formation à l'enseignement universitaire. *Savoirs : Revue Internationale De Recherches En Éducation Et Formation Des Adultes, 28*(1), 113-133. https://doi.org/10.3917/savo.028.0113

Di Pietro, G. (2023). The impact of covid-19 on student achievement: Evidence from a recent meta-analysis. *Educational Research Review,* https://doi.org/10.1016/j.edurev.2023.100530

Domínguez, C., Jaime, A., Sánchez, A., Blanco, J. M., & Heras, J. (2016). A comparative analysis of the consistency and difference among online self-, peer-, external- and instructor-assessments: The competitive effect. *Computers in Human Behavior, 60*, 112-120. https://doi.org/10.1016/j.chb.2016.02.061

Double, K., McGrane, J., & Hopfenbeck, T. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*(2), 481-509. https://doi.org/10.1007/s10648-019-09510-3

Dusenberry, L., & Robinson, J. (2020). Building psychological safety through training interventions: Manage the team, not just the project. *IEEE Transactions on Professional Communication, 63*(3), 207-226. https://doi.org/10.1109/TPC.2020.3014483

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350-383. http://dx.doi.org/10.2307/2666999

Edmondson, A. (2003). Managing the risk of learning: Psychological safety in work teams. In West, M., Tjosvold, D., & Smith, K. (Eds). *International Handbook of Organizational Teamwork and Cooperative Working*. (pp. 255-275). Wiley

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 23-43. https://doi.org/10.1146/annurev-orgpsych-031413-091305

Educational Endowment Foundation. (2023). Teaching & Learning Toolkit. An accessible summary of education evidence. https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit

Efron, S. E., & Ravid, R. (2019). *Writing the literature review: A practical guide*. Guilford Publications.

Eisinga, R. N., Grotenhuis, H. F. t., & Pelzer, B. J. (2013). The reliability of a two-item scale: Pearson, cronbach or spearman-brown? *International Journal of Public Health, 58*(4), 637-642. https://doi.org/10.1007/s00038-012-0416-3

Elliott, S. A., & Brown, J. S. L. (2002). What are we doing to waiting list controls? *Behaviour Research and Therapy, 40*(9), 1047-1052. https://doi.org/10.1016/S0005-7967(01)00082-1

Enders, C. K. (2010). *Applied missing data analysis*. The Guilford press.

Er, E., Dimitriadis, Y., & Gašević, D. (2021). Collaborative peer feedback and learning analytics: theory oriented design for supporting class-wide interventions. *Assessment & Evaluation in Higher Education, 46*(2), 169–190. https://doi.org/10.1080/02602938.2020.1764490

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70-120. https://doi.org/10.3102/0034654312474350

Ewald, H., Klerings, I., Wagner, G., Heise, T. L., Stratil, J. M., Lhachimi, S. K., Hemkens, L. G., Gartlehner, G., Armijo-Olivo, S., & Nussbaumer-Streit, B. (2022). Searching two or more databases decreased the risk of missing relevant studies: A metaresearch study. *Journal of Clinical Epidemiology, 149*, 154-164. https://doi.org/10.1016/j.jclinepi.2022.05.022

Falchikov, N. (1995). Peer Feedback Marking: Developing Peer Assessment. *Innovations in Education and Training International, 32, 175-187.* https://doi.org/10.1080/1355800950320212

Fithriani, R. (2018). Cultural influences on students' perceptions of written feedback in L2 writing. *Journal of Foreign Language Teaching & Learning, 3*(1), 1-13. https://doi.org/10.18196/ftl.3124

Frank, J., Snell, L., & Sherbino, J. (2015). *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada.

Gabelica, C., Bossche, P. V. d., Segers, M., & Gijselaers, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review, 7*(2), 123-144. https://doi.org/10.1016/j.edurev.2011.11.003

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment and Evaluation in Higher Education, 36*(2), 137-155. https://doi.org/10.1080/02602930903221444

Grohmann, A., & Kauffeld, S. (2013). Evaluating training programs: Development and correlates of the questionnaire for professional training evaluation. *International Journal of Training and Development, 17*(2), 135-155. https://doi.org/10.1111/ijtd.12005

*Güler, Ç. (2017). Use of WhatsApp in Higher Education: What's Up With Assessing Peers Anonymously? *Journal of Educational Computing Research*, *55*(2), 272-289. https://doi.org/10.1177/0735633116667359

Guo, X., & Lei, P. (2020). Effect of quality characteristics of peer raters on rating errors in peer assessment. *International Journal of Testing, 20*(3), 206-230. https://doi.org/10.1080/15305058.2020.1720216

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Multiple Raters* (4th edition ed.). Gaithersburg, MD: Advanced Analytics, LLC.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research and Development, 20*(1), 53-70. https://doi.org/10.1080/07294360123776

Harland, T., Wald, N., & Randhawa, H. (2017). Student peer review: Enhancing formative feedback with a rebuttal. *Assessment and Evaluation in Higher Education, 42*(5), 801-811. https://doi.org/10.1080/02602938.2016.1194368

Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education, 36*, 101-111. https://doi.org/10.1016/j.tate.2013.07.008

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review, 17*, 50-62. https://doi.org/10.1016/j.edurev.2015.12.002

Hew, K. F. (2018). Unpacking the strategies of ten highly rated MOOCs: Implications for engaging students in large online courses. *Teachers College Record, 120*(1), 1-40. https://doi.org/10.1177/016146811812000107

Hofstede, G. (1989). National Cultures Revisited. *Behavior Science Research*, *18*(4), 285-305.

Holgado Tello, F. P., Chacón Moscoso, S., Barbero García, I., & Sanduvete Chaves, S. (2006). Training satisfaction rating scale: Development of a measurement model using polychoric correlations. *European Journal of Psychological Assessment : Official Organ of the European Association of Psychological Assessment, 22*(4), 268-279. https://doi.org/10.1027/1015-5759.22.4.268

Howell, D. C. (2007). *Statistical methods for psychology* (6e éd. ed.). Thomson.

Hu, K., Godfrey, K., Ren, Q., Wang, S., Yang, X., & Li, Q. (2022). The impact of the COVID-19 pandemic on college students in USA: Two years later. *Psychiatry Research, 315*, 114685-114685. https://doi.org/10.1016/j.psychres.2022.114685

Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: A meta-analysis. *Assessment and Evaluation in Higher Education, 44*(6), 863-880. https://doi.org/10.1080/02602938.2018.1545896

Hunt, D., Bailey, J., Lennox, B., Crofts, M., & Vincent, C. (2021). Enhancing psychological safety in mental health services. *International Journal of Mental Health Systems, 15*(1), 33-33. https://doi.org/10.1186/s13033-021-00439-1

*Kaya, F. (2021). Emotions related to identifiable/anonymous peer feedback: A case study with turkish pre-service english teachers. *Issues in Educational Research, 31*(4), 1088-1100.

Kim, M. (2009). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education*, 34(1), 105–114. https://doi.org/10.1080/02602930801955960

*Kim, S. (2019). Japanese student writers' perspectives on anonymous peer review, *ELT Journal*, *73(3*), 296–305. https://doi.org/10.1093/elt/ccy061

*Kim, S., & Lan, Y. (2021). L2 writers' perspectives on face-to-face and anonymous peer review: Voices from China. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras, 35*, 149-164. https://doi.org/10.30827/portalin.vi35.13887

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344-348. https://doi.org/10.1016/j.learninstruc.2009.08.005

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. https://doi.org/10.1037/0022-3514.77.6.1121

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lee, M.-K. (2015). Peer feedback in second language writing: Investigating junior secondary students' perspectives on inter-feedback and intra-feedback. *System*, 55, 1–10. https://doi.org/10.1016/j.system.2015.08.003

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education, 45*(2), 193-211. https://doi.org/10.1080/02602938.2019.1620679

*Li, L. (2017). The role of anonymity in peer assessment. *Assessment and Evaluation in Higher Education, 42*(4), 645-656. https://doi.org/10.1080/02602938.2016.1174766

Li, L., Steckelberg, A.L. (2005). The instructional design portfolio: Peer assessment support system (Pass). *TECHTRENDS TECH TRENDS 49*, 80–84. https://doi.org/10.1007/BF02824115

Li, L., Liu, X., & Steckelberg, A. L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525-536. https://doi.org/10.1111/j.1467-8535.2009.00968.x

*Lim, W., Son, J., & Kang, S. (2021). How reducing discomfort impacts peer assessments of preservice teachers. *Sustainability (Basel, Switzerland), 13*(11), 6435. https://doi.org/10.3390/su13116435

*Lin, G. (2018). Anonymous versus identified peer assessment via a facebook-based learning application: Effects on quality of peer feedback, perceived learning, perceived fairness, and attitude toward the system. *Computers and Education, 116*, 81-92. https://doi.org/10.1016/j.compedu.2017.08.010

Lindvall, J., & Ryve, A. (2019). Coherence and the positioning of teachers in professional development programs. A systematic review. *Educational Research Review, 27*, 140-154. https://doi.org/10.1016/j.edurev.2019.03.005

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*, 1-29. https://doi.org/10.3389/feduc.2021.720195

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied, 15*(4), 319–333. https://doi.org/10.1037/a0017841

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. Teaching in Higher Education, 11(3), 279–290. https://doi.org/10.1080/13562510600680582

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*(1), 30-43. https://doi.org/10.1016/j.jslw.2008.06.002

Luxton-Reilly, A. (2009). A systematic review of tools that support peer assessment. *Computer Science Education, 19*(4), 209-232. https://doi.org/10.1080/08993400903384844

Malecka, B., Boud, D., & Carless, D. (2020). Eliciting, processing and enacting feedback: Mechanisms for embedding student feedback literacy within the curriculum. *Teaching in Higher Education,* 1-15. https://doi.org/10.1080/13562517.2020.1754784

Martin, B. O., Kolomitro, K., & Lam, T. C. M. (2014). Training methods: A review and analysis. *Human Resource Development Review, 13*(1), 11-35. https://doi.org/10.1177/1534484313497947

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709-734. https://doi.org/10.2307/258792

McConlogue, T. (2012). But is it fair? developing students' understanding of grading complex written work through peer assessment. *Assessment and Evaluation in Higher Education, 37*(1), 113-123. https://doi.org/10.1080/02602938.2010.515010

McGarr, O., & Clifford, A. M. (2013). Just enough to make you take it seriously: Exploring students attitudes towards peer assessment. *Higher Education, 65*(6), 677-693. https://doi.org/10.1007/s10734-012-9570-z

Min, H. (2005). Training students to become successful peer reviewers. *System, 33*(2), 293-308. https://doi.org/10.1016/j.system.2004.11.003

Misiejuk, K., & Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers and Education, 175*, 104319. https://doi.org/10.1016/j.compedu.2021.104319

Misiejuk, K., Wasson, B., & Egelandsdal, K. (2021). Using learning analytics to understand student perceptions of peer feedback. *Computers in Human Behavior, 117*, 106658. https://doi.org/10.1016/j.chb.2020.106658

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment and Evaluation in Higher Education, 45*(4), 527-540. https://doi.org/10.1080/02602938.2019.1667955

Mostert, M., & Snowball, J. D. (2013). Where angels fear to tread: Online peer-assessment in a large first-year class. *Assessment and Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2012.683770

Mulder, R. A., Pearce, J. M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education, 15*(2), 157-171. https://doi.org/10.1177/1469787414527391

Narciss, S. (2008). Feedback Strategies for Interactive Learning Tasks. In Spector, M., Merrill, J., Van Merrienboer, J., & Driscoll, M. (Eds), *Handbook of Research on Educational Communications and Technology* (pp. 125-143). Lawrence Erlbaum Associates.

Neugebauer, M., Patzina, A., Dietrich, H., & Sandner, M. (2023). Two pandemic years greatly reduced young people's life satisfaction: evidence from a comparison with pre-COVID-19 panel data, *European Sociological Review*, 1-15. https://doi.org/10.1093/esr/jcad077

Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education, 27*(4), 405-417. https://doi.org/10.1080/0307507022000011525

Ng, W.S., & Yu, G. (2023). Students' attitude to peer assessment process: A critical factor for success. *Interactive Learning Environments, 31*(5), 2967-2985. https://doi.org/10.1080/10494820.2021.1916762

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self- regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment and Evaluation in Higher Education, 39*(1), 102-122. https://doi.org/10.1080/02602938.2013.795518

*Ocampo, J. C. G., Panadero, E., & Díez, F. (2023). Are men and women really different? The effects of gender and training on peer scoring and perceptions of peer assessment. *Assessment and Evaluation in Higher Education, 48*(6), 760-776. https://doi.org/10.1080/02602938.2022.2130167

OECD (2021), *Education at a Glance 2021: OECD Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/b35a14e5-en.

Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilization of tutors' formative feedback: A qualitative interview study. *Assessment and Evaluation in Higher Education, 30*(4), 369-386. https://doi.org/10.1080/02602930500099177

Paakkari, L., Tynjälä, P., & Kannas, L. (2011). Critical aspects of student teachers' conceptions of learning. *Learning and Instruction, 21*(6), 705-714. https://doi.org/10.1016/j.learninstruc.2011.03.003

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology, 10*(1), 178-89. https://doi.org/10.1186/s13643-021-01626-4

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment. In Brown, G. T. L., & Harris, L. R. (Eds), *Handbook of human and Social Conditions in Assessment* (pp. 247–265). Routledge/Taylor & Francis.

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment and Evaluation in Higher Education, 44*(8), 1253-1278. https://doi.org/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. *Assessment and Evaluation in Higher Education, 48*(8), 1053-1075. https://doi.org/10.1080/02602938.2023.2164884

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education, 32*(1), 133-156. https://doi.org/10.1007/s10212-015-0282-5

Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know? In A. A. Lipnevich & J. K. Smith (Eds.), The Cambridge handbook of

instructional feedback (pp. 409–431). Cambridge University Press. https://doi.org/10.1017/9781316832134.020

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

*Parikh, B. T., & Sheehan, A. H. (2016). Student pharmacists' perceptions of the implementation of peer assessment instruction and activities within a didactic course. *Currents in Pharmacy Teaching and Learning, 8*(6), 821-826. https://doi.org/10.1016/j.cptl.2016.08.012

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Wiley. https://doi.org/10.1002/9780470754887

Planas Lladó, A., Soley, L. F., Fraguell Sansbelló, R. M., Pujolras, G. A., Planella, J. P., Roura-Pascual, N., Suñol Martínez, J. J., & Moreno, L. M. (2014). Student perceptions of peer assessment: An interdisciplinary study. *Assessment and Evaluation in Higher Education, 39*(5), 592-610. https://doi.org/10.1080/02602938.2013.860077

Popov, V., Brinkman, D., Biemans, H. J. A., Mulder, M., Kuznetsov, A., & Noroozi, O. (2012). Multicultural student group work in higher education. an explorative case study on challenges as perceived by students. *International Journal of Intercultural Relations, 36*(2), 302-317. https://doi.org/10.1016/j.ijintrel.2011.09.004

Poverjuc, O., Brooks, V., & Wray, D. (2012). Using peer feedback in a master's programme: A multiple case study. *Teaching in Higher Education, 17*(4), 465-477. https://doi.org/10.1080/13562517.2011.641008

*Praver, M., Rouault, G., & Eidswick, J. (2011). Attitudes and affect toward peer evaluation in EFL reading circles. *The Reading Matrix, 11*(2), 89-101.

*Raes, A., Vanderhoven, E., & Schellens, T. (2015). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education, 40*(1), 178-193. https://doi.org/10.1080/03075079.2013.823930

Rege-Colet, N., & Berthiaume, D. (2015). La pédagogie de l'enseignement supérieur : Repères théoriques et applications pratiques : Tome 2 : Se développer au titre d'enseignant. P.I.E-Peter Lang S.A., Éditions Scientifiques Internationales. https://doi.org/10.3726/978-3-0352-0299-1

Reina, D. S., & Reina, M. L. (2015). *Trust and betrayal in the workplace: Building effective relationships in your organization* (Third). Berrett-Koehler Publishers.

Ritter, F.E., Schooler, L.J. (2001). The learning curve. In: Smelser, N.J., Baltes, P.B. *International Encyclopedia of Social and Behavioral Sciences*. Amsterdam, Netherlands: Pergamon.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology, 52*(1), 453-470. https://doi.org/10.1146/annurev.psych.52.1.453

Rotsaert, T. (2017). *The social nature of peer assessment in secondary and higher education : examining students' perceptions on interpersonal processes and peer feedback quality*

in anonymous face-to-face settings using Mobile Response Technology [Unpublished doctoral dissertation]. Ghent University.

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in flanders. *Studies in Educational Evaluation, 53*, 29-40. https://doi.org/10.1016/j.stueduc.2017.02.003

*Rotsaert, T., Panadero, E., & 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. *European Journal of Psychology of Education, 33*(1), 75-99. https://doi.org/10.1007/s10212-017-0339-8

Roy, A., Guay, F., & Valois, P. (2013). Teaching to address diverse learning needs: Development and validation of a differentiated instruction scale. *International Journal of Inclusive Education, 17*(11), 1186-1204. https://doi.org/10.1080/13603116.2012.743604

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. https://doi.org/10.1007/BF00117714

Sarasso, P., Ronga, I., Del Fante, E., Barbieri, P., Lozzi, I., Rosaia, N., Cicerale, A., Neppi-Modona, M., & Sacco, K. (2022). Physical but not virtual presence of others potentiates implicit and explicit learning. *Scientific Reports, 12*(1), 21205-21205. https://doi.org/10.1038/s41598-022-25273-4

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *The Academy of Management Review, 32*(2), 344-354. https://doi.org/10.5465/AMR.2007.24348410

*Senden, M., De Jaeger, D., & Coertjens, L. (2023). Safe and sound: Examining the effect of a training targeting psychological safety and trust in peer assessment. *Frontiers in Education, 8*, 1-13. https://doi.org/10.3389/feduc.2023.1198011

Servant-Miklos, V. F. C. (2019). A revolution in its own right: How maastricht university reinvented problem-based learning. *Health Professions Education, 5*(4), 283-293. https://doi.org/10.1016/j.hpe.2018.12.005

Shah, D. (2020). *By the numbers: MOOCs during the pandemic*. Retrieved from https://www.classcentral.com/report/mooc-stats-pandemic

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. https://doi.org/10.3102/0034654307313795

Sinclair, H. K., & Cleland, J. A. (2007). Undergraduate medical students: Who seeks formative feedback? *Medical Education, 41*(6), 580-582. https://doi.org/10.1111/j.1365-2923.2007.02768.x

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*(1), 23-42. https://doi.org/10.1016/S0191-491X(03)90003-4

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J J. G, & Martens, R. L. (2004). Training teachers in peer-assessment skills: Effects on performance and perceptions.

Innovations in Education and Teaching International, 41(1), 59-78. https://doi.org/10.1080/1470329032000172720

Soares, A. E., & 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. *Active Learning in Higher Education, 21*(1), 65-78. https://doi.org/10.1177/1469787417742023

Stanier, L. (1997). Peer assessment and group work as vehicles for student empowerment: A module evaluation. *Journal of Geography in Higher Education*. https://doi.org/10.1080/03098269708725413

Strijbos, J., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers and Education, 46*(1), 29-48. https://doi.org/10.1016/j.compedu.2005.04.002

Strijbos, J. W., Ochoa, T. A., Sluijsmans, D. M. A., Segers, M. S. R., & Tillema, H. H. (2009). Fostering interactivity through formative peer assessment in web-based collaborative learning environments. In C. Mourlas, N. Tsianos, & P. Germanakos (Eds.), *Cognitive and emotional processes in webbased education: Integrating human factors and personalization* (pp. 375–395). IGI Global.

Strijbos, J., & Sluijsmans, D. (2010). Unravelling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*(4), 265-269. https://doi.org/10.1016/j.learninstruc.2009.08.002

*Su, W. (2023). Masked ball for all: How anonymity affects students' perceived comfort levels in peer feedback. *Assessment and Evaluation in Higher Education, 48*(4), 502-512. https://doi.org/10.1080/02602938.2022.2089348

Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *International Review of Research in Open and Distance Learning, 15*(3), 312-327. https://doi.org/10.19173/irrodl.v15i3.1680

Sung, Y., Chang, K., Chang, T., & Yu, W. (2010). How many heads are better than one? the reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence, 33*(1), 135-145. https://doi.org/10.1016/j.adolescence.2009.04.004

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467-481. https://doi.org/10.1007/s10734-017-0220-3

*Thanh, P. T. H., & Gillies, R. (2010). Designing a Culturally Appropriate Format of Formative Peer Assessment for Asian Students: The Case of Vietnamese Students. International Journal of Educational Reform, 19(2), 72-85. https://doi.org/10.1177/105678791001900201

Theureau, J. (2010). Les entretiens d'autoconfrontation et de remise en situation par les traces matérielles et le programme de recherche « cours d'action. *Revue d'Anthropologie Des Connaissances, 4-2*(2), 287-322. https://doi.org/10.3917/rac.010.0287

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276. https://doi.org/10.3102/00346543068003249

Topping, K. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology, 1*(1), 7. https://doi.org/10.31532/InterdiscipEducPsych ol.1.1.007

Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia, 11*(Suppl 1), S80-S89. https://doi.org/10.4103/sja.SJA_203_17

*Tzeng, A., Bruno, B., Cooperrider, J., Dinardo, P. B., Baird, R., Swetlik, C., Goldstein, B. N., Rastogi, R., Roth, A. J., Gilligan, T. D., & Rish, J. M. (2021). A structured peer assessment method with regular reinforcement promotes longitudinal self-perceived development of medical students' feedback skills. Medical Science Educator, 31(2), 655-663. https://doi.org/10.1007/s40670-021-01242-w

UNESCO (2021). *COVID-19: reopening and reimagining universities, survey on higher education through the UNESCO National Comissions*. Retrieved from: https://unesdoc.unesco.org/ark:/48223/pf0000378174

van Gennip, N., Gijbels, D., Segers, M. S. R., & Tillema, H. (2010). Reactions to 360 degree feedback: the role of trust and trust-related variables. *International Journal of Human Resources Development and Management*, *10*(4), 362-379. https://doi.org/10.1504/IJHRDM.2010.036088

van Gennip, N., Segers, M., & Tillema, H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*. https://doi.org/10.1016/j.edurev.2008.11.002

van Gennip, N., Segers, M., & Tillema, H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280-290. https://doi.org/10.1016/j.learninstruc.2009.08.010

*van Heerden, M., & Bharuthram, S. (2021). Knowing me, knowing you: The effects of peer familiarity on receiving peer feedback for undergraduate student writers. *Assessment and Evaluation in Higher Education, 46*(8), 1191-1201. https://doi.org/10.1080/02602938.2020.1863910

Van Meenen, F., Masson, N., Catrysse, L., & Coertjens, L. (2023). Taking a closer look at how higher education students process and use (discrepant) peer feedback. *Learning and Instruction, 84*, 101711. https://doi.org/10.1016/j.learninstruc.2022.101711

van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*(4), 270-279. https://doi.org/10.1016/j.learninstruc.2009.08.004

*Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers and Education, 81*, 123-132. https://doi.org/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. *Sociological Methods & Research, 49*(3), 699-718. https://doi.org/10.1177/0049124117747303

Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: A literature review. *Journal of Advanced Nursing, 50*(2), 204-211. https://doi.org/10.1111/j.1365-2648.2005.03380.x

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment and Evaluation in Higher Education, 43*(7), 1032-1047. https://doi.org/10.1080/02602938.2018.1427698

Wen, M. L., & Tsai, C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*(1), 27-44. https://doi.org/10.1007/s10734-004-6375-8

Wichmann, A., Funk, A., & Rummel, N. (2018). Leveraging the potential of peer feedback in an academic writing activity through sense-making support. *European Journal of Psychology of Education, 33*(1), 165. https://doi.org/10.1007/s10212-017-0348-7

Wiedemann, A., Stochl, J., Neufeld, S. A. S., Fritz, J., Bhatti, J., Hook, R. W., Bullmore, E., Dolan, R., Goodyer, I., Fonagy, P., Jones, P., Moutoussis, M., Hauser, T., Neufeld, S., Romero-Garcia, R., Clair, M. S., Vértes, P., Whitaker, K., Inkster, B., . . . NSPN Consortium. (2022). The impact of the initial COVID-19 outbreak on young adults' mental health: A longitudinal study of risk and resilience factors. *Scientific Reports, 12*(1), 16659-16659. https://doi.org/10.1038/s41598-022-21053-2

Willems, J., Vanoverberghe, V., Coertjens, L., & Donche, V. (2023). The first-year integration test: A validation study. *Frontiers in Psychology, 14*, 1101234-1101234. https://doi.org/10.3389/fpsyg.2023.1101234

Wilson, M. J., Diao, M. M., & 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. *Assessment and Evaluation in Higher Education, 40*(1), 15-32. https://doi.org/10.1080/02602938.2014.881980

Winstone, N., Boud, D., Dawson, P., & Heron, M. (2022). From feedback-as-information to feedback-as-process: A linguistic analysis of the feedback literature. *Assessment and Evaluation in Higher Education, 47*(2), 213-230. https://doi.org/10.1080/02602938.2021.1902467

Winstone, N. E., and Nash, R. A. (2016). *The Developing Engagement with Feedback Toolkit (DEFT)*. York: Higher Education Academy.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes. *Educational Psychologist, 52*(1), 17-37. https://doi:10.1080/00461520.2016.1207538

Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). 'it'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience. *Studies in Higher Education, 42*(11), 2026-2041. https://doi.org/10.1080/03075079.2015.1130032

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 3087-3087. https://doi.org/10.3389/fpsyg.2019.03087

Wu, Y., & 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. *Journal of Educational Psychology, 115*(1), 118-142. https://doi.org/10.1037/edu0000768

* Xue, S., Yang, Y., Du, J., & Liu, F. (2023). Multi-layered e-feedback anxiety: An action research study among chinese learners using peer feedback activities in an academic writing course. *Frontiers in Psychology, 14*, 1062517-1062517. https://doi.org/10.3389/fpsyg.2023.1062517

Yang, Y., & Tsai, C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction, 20*(1), 72-83. https://doi.org/10.1016/j.learninstruc.2009.01.003

*Yu, F. Y. (2012). Any effects of different levels of online user identity revelation? *Educational Technology & Society, 15*(1), 64-77.

*Yu, F., & Liu, Y. (2009). Creating a psychologically safe online space for a student-generated questions learning activity via different identity revelation modes. *British Journal of Educational Technology, 40*(6), 1109-1123. https://doi.org/10.1111/j.1467-8535.2008.00905.x

*Yu, F., & Wu, C. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors, classroom climate and learning activities. *Computers and Education, 57*(3), 2167-2177. https://doi.org/10.1016/j.compedu.2011.05.012

Yu, S., Lee, I., & Mak, P. (2016). Revisiting chinese cultural issues in peer feedback in EFL writing: Insights from a multiple case study. *The Asia-Pacific Education Researcher, 25*(2), 295-304. https://doi.org/10.1007/s40299-015-0262-1

*Zhang, M., He, Q., Du, J., Liu, F., & Huang, B. (2022). Learners' perceived advantages and social-affective dispositions toward online peer feedback in academic writing. *Frontiers in Psychology, 13*, 973478-973478. https://doi.org/10.3389/fpsyg.2022.973478

Zhang, F., Schunn, C., Chen, S., Li, W., & Li, R. (2023). EFL student engagement with giving peer feedback in academic writing: A longitudinal study. *Journal of English for Academic Purposes, 64*, 101255. https://doi.org/10.1016/j.jeap.2023.101255

Zhao, Y., Zhang, F., Schunn, C. D., He, P., Li, D., & Zhao, Y. (2024). Feedback, feedback-on-feedback and re-feedback: Effects of written dialogic peer feedback on english as a foreign language writing. *Assessment and Evaluation in Higher Education, 49*(4), 523-538. https://doi.org/10.1080/02602938.2023.2278017

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research and Development, 37*(4), 883-897. https://doi.org/10.1080/07294360.2018.1446417

Zou, Y., Schunn, C. D., Wang, Y., & Zhang, F. (2018). Student attitudes that predict participation in peer assessment. *Assessment and Evaluation in Higher Education, 43*(5), 800-811. https://doi.org/10.1080/02602938.2017.1409872