

Anonymity as an instructional scaffold in peer assessment: its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills

Tijs Rotsaert¹  · Ernesto Panadero² · Tammy Schellens¹

Received: 28 September 2016 / Revised: 21 June 2017 / Accepted: 24 July 2017 /
Published online: 4 August 2017

© Instituto Superior de Psicologia Aplicada, Lisboa, Portugal and Springer Science+Business Media B.V. 2017

Abstract Although previous research has indicated that providing anonymity is an effective way to create a safe peer assessment setting, continuously ensuring anonymity prevents students from experiencing genuine two-way interactive feedback dialogues. The present study investigated how installing a transitional approach from an anonymous to a non-anonymous peer assessment setting can overcome this problem. A total of 46 bachelor's degree students in Educational Studies participated in multiple peer assessment cycles in which groups of students assessed each other's work. Both students' evolution in peer feedback quality as well as their perceptions were measured. The content analysis of the peer feedback messages revealed that the quality of peer feedback increased in the anonymous phase, and that over time, the feedback in the consecutive non-anonymous sessions was of similar quality. The results also indicate that the transitional approach does not hinder the perceived growth in peer feedback skills, nor does it have a negative impact on their general conceptions towards peer assessment. Furthermore, students clearly differentiated between their attributed importance of anonymity and their view on the usefulness of a transitional approach. The findings suggest that anonymity can be a valuable scaffold to ease students' importance level towards anonymity and their associated need for practice.

Keywords Anonymity · Peer assessment · Peer feedback

✉ Tijs Rotsaert
Tijs.Rotsaert@UGent.be

¹ Department of Educational Studies, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

² Departamento de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid, Madrid, Spain

Introduction

Peer assessment is an assessment method in which students are actively involved in the assessment process. In peer assessment activities, students are required to reflect on the quality of peers' work and discuss how well it corresponds with the explicitly stated goals or criteria of the work (Strijbos and Sluijsmans 2010). One of the central components of peer assessment is peer feedback, which is the information that a student provides to a peer (Topping 1998). Students are engaged in high-level cognitive processing during the peer feedback process because it requires skills such as explaining, identifying mistakes and gaps, and providing suggestions for improvement (King 2002). For this reason, feedback is seen as an essential aspect of the learning process (Hattie and Timperley 2007). Furthermore, peer feedback exemplifies professional practice, in which colleagues collaborate and offer input on how to improve current work (Van der Pol et al. 2008).

However, because peer assessment is a fundamentally social and collaborative learning activity, learners' interpersonal beliefs can negatively impact its outcomes (Panadero 2016; Raes et al. 2013). This includes possible reciprocity effects and negative feelings caused by interpersonal variables such as friendship marking due to friendship bonds, psychological unsafety, fear of disapproval when giving a low score or negative feedback (i.e., recrimination), and distrust in one's own and others' evaluative capabilities (e.g., Harris and Brown 2013; van Gennip et al. 2010; Vanderhoven et al. 2015). The recognition of the presence of interpersonal variables in peer assessment is thus important because the pressure students experience in the process may in turn directly impact how they view the value of peer assessment (Li 2016).

Previous research has shown that providing anonymity to assessors can help relieve the interpersonal burden from students, especially in peer assessment activities in which a score is given (Yu and Liu 2009). Providing anonymity leads to more positive perceptions towards peer assessment (Vanderhoven et al. 2015) and helps students become more willing to give critical feedback (Howard et al. 2010). However, anonymous settings do not reflect daily face-to-face situations in which people give and receive feedback with known identities, and, importantly, a call for non-oversimplified implementations of anonymity has been issued (Panadero 2016).

Accordingly, in order to involve students in reflective criticism of the products of their peers, classroom interventions are needed that (a) recognize peer feedback as an essential component of peer assessment, (b) acknowledge the inherent social nature of the peer assessment process, and (c) guide students towards an open, dialogic, and non-anonymous feedback environment in which they can develop sustainable assessment skills (Boud and Soler 2015; Carless et al. 2011).

Peer feedback quality

The process of assessing and commenting on the strengths and weaknesses of a peer's work can help familiarize assessors with the evaluation criteria and thereby develop knowledge on what constitutes high-quality work (Cho and Cho 2010). In essence, well-formulated feedback should provide an answer to three questions: "Where am I going?" (feed up), "How am I going?" (feedback), and "Where to next?" (feed forward) (Hattie and Timperley 2007). However, it cannot be expected that every student will offer high-quality feedback because this requires high-level cognitive processing (Strijbos et al. 2010); students need to be capable

of dealing with specific assessment criteria to judge a peer's performance (Gielen and De Wever 2015).

Peer feedback quality can be approached in two ways: (a) in terms of accuracy, consistency across assessors, and/or concordance with teacher feedback (Van Steendam et al. 2010) or (b) in terms of content and/or style characteristics (Gielen et al. 2010). The first approach focuses on the numbers of errors and/or holistic scores for correctness of peer comments, in comparison with a yardstick (usually a teacher). This definition originates from the summative view of peer assessment, in which scoring validity and reliability are the primary goals (Dochy, Segers & Sluijms, 1999; Gielen et al. 2010). In our opinion, this view is problematic because peers are inevitably novices and not experts, and therefore certain grades of separation from the scoring of an expert should be expected. The second approach with regard to peer feedback quality, which is used in this study, defines it in terms of feedback content characteristics. Gielen et al. state that "the advantage of this approach is that such characteristics are not domain- or task-specific, thus teaching students to focus on content and style characteristics results in a generic skill transferable to other settings" (Gielen et al. 2010, p. 306). In other words, this approach focuses on the development of students' evaluative expertise, preferably beyond the immediate task (cf. Boud and Soler 2015; Carless et al. 2011). Nicol et al. (2014) add that to fully realize the benefits of peer feedback, students must produce a written explanation for their evaluative judgments.

More specifically, previous research has indicated that qualitative feedback should contain two types of information: verifications and elaborations (Narciss 2008). *Verification* refers to "a dichotomous judgment to indicate that a response is right or wrong"; in other words, it tells the assessee whether a certain criterion was met or not. Gielen and De Wever (2015) found in their study on asynchronous online environments that students tended to give more positive than negative verifications. When offered a more structured environment (i.e., a peer feedback template), students gave more negative verifications. The negative verifications are necessary to expose shortcomings in a peer's performance. *Elaboration* refers to "relevant information to help the learner in error correction" (Hattie and Gan 2011, p. 253). These types of information are thus seen as the structural components of feedback because students require feedback that tells them not only if they performed the task (in)correctly (feedback) but also why and what they should do to improve their work (feedforward) (e.g., Prins et al. 2005). Therefore, offering elaborations that justify the verification (correct vs. incorrect) is presumed to be beneficial for students' learning and as a consequence, a balanced proportion of verifications and elaborations is more valuable than providing verifications alone (Gielen and De Wever 2015). Furthermore, previous research has shown that practice is crucial for the development of peer assessment skills (Sluijmsmans 2002). The more practice in peer assessment processes, the more likely students are to develop expertise in making sound peer assessment judgements (Panadero 2016).

A recent study by Rotsaert, Panadero, Schellens & Raes (2017) in an anonymous synchronous face-to-face peer assessment environment found that peer assessment practice (i.e., repeatedly practicing one's evaluative judgment skills through involvement in multiple peer assessment sessions) improved peer feedback quality in terms of content characteristics; messages contained more negative verifications and informative and suggestive elaborations after the intervention. The feedback thus becomes more descriptive of the actual performance, rather than just pointing out the positive aspects. This finding of a content quality increase in a synchronous peer assessment setting is promising because students get less time to formulate their feedback message compared to an online asynchronous environment, in which students

have more time to reflect on their feedback before providing it (Tsai et al. 2002). However, when anonymity remains preserved in a face-to-face peer assessment setting, this does not reflect authentic feedback dialogues as, for example, in work environments (see further).

Perceived peer feedback skills

The potential of learning and assessment activities depends on the way students perceive them (Boud and Soler 2015). If students think that they are becoming more capable as peer assessors, they will be more motivated to perform peer assessment and believe it is useful (Vanderhoven et al. 2015). Therefore, it is important to consider students' perception of the improvement of their peer feedback skills when studying the development of their peer feedback quality; the alignment of students' perceived peer feedback skills and the actual quality of their generated peer feedback quality is important. A recent study found that after providing peer feedback multiple times, students reported a self-perceived increase in their peer feedback skills, which were measured before, throughout, and after the intervention (Author, 2017). Previous research has shown that this results in a number of benefits, such as students having more control over the feedback process and as a result more control over their own learning (Nicol et al. 2014). This could support the development of students' sustainable assessment skills—that is, increasing their capacity to judge their own future work (Boud and Soler 2015).

Anonymity as an instructional scaffold within peer assessment

The literature on social psychology suggests that non-anonymous and anonymous interactions may produce differential effects on participants' perceptions of their interaction counterparts, the interaction space, and the experience itself.

The theoretical foundations with regard to the possible impact of anonymity imply that students will enact different feedback behavior depending on whether their identity as assessors is revealed (Yu and Sung 2015). It is this approach that the current study adopts, and Yu and Sung (2015) provide two perspectives from social psychology that help to support and contextualize this adoption. Social Identity Theory (SIT) suggests that in addition to his/her unique personal identity as an individual, a person also forms a social identity according to the groups with which he/she affiliates (Pearce 2013). SIT has been used to explain and predict certain personal behaviors on the basis of, among other factors, interpersonal relationships in group situations (Hogg et al. 2006). More specifically, SIT proposes that a person with a more well-received social identity and greater charisma would be perceived as a more reliable source of normative information (e.g., peer feedback) and thus have more influence over the behavior of other group members (Hogg et al. 2012). Furthermore, adolescents and young adults are known to be particularly influenced by the views of their peers, compared to younger children (Brown 2004). In consideration of these effects, it has been suggested that offering anonymity to participants during group interactions could foster higher participation and more balanced engagement among individuals (Chester and Gwynne 2006; Hosack 2004).

Anonymity is also believed to offer a sense of psychological safety (Miyazoe and Anderson 2011; Yu and Liu 2009), which is defined as a shared belief denoting one's emotional ability to take an interpersonal risk without fearing negative consequences with regard to one's well-being, self-image, and status (Kahn 1990; Zhang et al. 2010). In general, individuals who feel psychologically safe are more likely to perceive differences in opinions as opportunities rather

than conflicts and to provide candid and critical peer feedback that can lead to higher quality learning outcomes (Lu and Bol 2007; van Gennip et al. 2009). Moreover, the negative influences that self-consciousness can exert on the assessor in peer assessment activities can also be relieved (Roberts and Rajah-Kanagasabai 2013; Zhang et al. 2010). With these liberating effects, anonymity can help relax the social customs and conventional roles that are usually expected of students (Miyazoe and Anderson 2011) and is thus preferable to a situation in which participants know each other (e.g., Hosack 2004).

In the case of peer feedback, anonymity should allow students to express feedback that may differ from a prevailing group norm or the views of a dominant individual. Recent studies that examine this topic are scarce and highly context-dependent, given that anonymity can be operationalized in many formats. For example, in a series of studies by Yu et al. (e.g., Sung et al. 2010), no significant difference was found between the actual interaction behavior of participants in non-anonymous and anonymous online peer assessment conditions. Other studies have indicated that providing anonymity to assessors (i.e., when a single-blind environment is created) can help relieve the interpersonal burden on students (Yu and Liu 2009). Cheng and Tsai (2012) found that anonymity is preferable in order to avoid the pressure of friendships.

The existing research suggests that retaining anonymity for assessors is one of several factors that encourages student participation (Ballantyne et al. 2002; Vickerman 2009). Additionally, due to its role in diminishing reciprocity effects, anonymity for the assessor might result in fairer assessment (Freeman and McKenzie 2000). Vanderhoven et al. (2015) found that in a synchronous peer assessment setting, anonymity for assessors helped in decreasing their fear of disapproval when giving a low score or negative feedback. Furthermore, students seemed to experience less peer pressure (Vanderhoven et al. 2015).

In essence, when students interpret assessments and feedback they have received from peers, the social context can be critical because peer assessment “does not happen in a vacuum; rather it produces thoughts, actions, and emotions as a consequence of the interaction of assesseees and assessors” (Panadero 2016, p. 2). Although current research tends to favor the use of anonymity, in a review study, Panadero (2016) pointed out a tension between implementing anonymous peer assessment and the formative use of peer assessment; anonymity might lessen the impact of interpersonal processes while simultaneously hinder the creation of a rich and interactive feedback environment. When implementing peer assessment, teachers are thus challenged to find a balance between the creation of a safe learning environment provided through anonymity and the creation of a rich peer feedback setting, which will consequently take up more class time (Panadero 2016). To date, no clear guidelines are available for teachers to cope with this tension.

Furthermore, the relationship between the attributed importance of anonymity for students and the quality of feedback has, to date, not been sufficiently explored. Howard et al. (2010) found in an online asynchronous peer assessment setting that students whose anonymity was preserved were approximately five times more likely to provide critical feedback than those whose identities were known to their recipients. Furthermore, it has been suggested that the relevance of anonymity should be explored in real face-to-face peer assessment settings because current studies mostly focus on online peer assessment settings (Ainsworth et al. 2011). Building on the suggestion of Howard et al. (2010), who state that “anonymity might provide a scaffold toward a protected social dynamic for novice feedback” (p. 90) and in order to acknowledge students’ possible feelings of unsafety and psychological discomfort (Vanderhoven et al. 2015), in this study, assessors’ anonymity was used as a facilitator to generate critical feedback without social repercussions. Our intention was to explore this hypothesis by gradually evolving from an

anonymous peer assessment towards dialogic feedback in a face-to-face setting. In dialogic feedback, interpretations are shared, meanings negotiated, and expectations clarified between assessor and assessee (Carless et al. 2011). In order to be able to study this practice phase, we created a face-to-face peer assessment activity in which students experienced the transition from an anonymous to a non-anonymous setting (see “[Procedure](#)”).

Research questions and hypotheses

The aim of this study is to explore the effects of faded anonymity on peer feedback quality, while also exploring students’ perceptions about the process. The specific research questions and hypotheses are:

RQ1: How does peer feedback quality change over time when students consecutively practice peer assessment in anonymous and non-anonymous settings?

(H1) There will be an overall increase in negative verifications and informative and suggestive elaborations because the faded anonymity effect should help students to become willing to point out/elaborate on weaknesses in peers’ work and formulate suggestions for improvement.

RQ2: How do students’ perceived peer feedback skills change over time in a peer assessment setting with a transition from anonymous to non-anonymous?

(H2) There will be an increase in perceived peer feedback skills due to students’ appreciation for the transition towards non-anonymity in the peer assessment setting.

RQ3: How does the transition from an anonymous to a non-anonymous peer assessment affect students’ perceptions regarding (a) the importance of anonymity, (b) their perceptions towards interpersonal variables, and (c) their general conceptions towards peer assessment?

(H3a) The importance given to anonymity will be lower after the intervention because participants will by then appreciate the authentic and personalized style of peer assessment and peer feedback.

(H3b) Regarding the positive interpersonal variables (psychological safety, trust, and value congruency), an overall increase is hypothesized. Regarding the negative interpersonal variables (fear of disapproval and friendship marking), the opposite evolution is expected.

(H3c) An overall increase in students’ general conceptions towards peer assessment is expected because students will appreciate the imposed transition between anonymity modes and its effects on their peer feedback skills.

Method

Participants

The participants in this study were 46 third-year bachelor’s students in Educational Studies who were enrolled in the course Instructional Design. Participants’ mean age was 21 years, and

the majority were female (84.44%). A total of 88.6% had prior experience with peer assessment.

Procedure

Students received a group assignment to prepare and present a workshop (max. 30 min) on one of the provided topics (e.g., The Jigsaw Classroom). Students worked in small groups (of which there were 16). The learning goal was to organize a short workshop in a team and choose an appropriate format to transfer new learning content. Participants received an introductory lecture to the pedagogical principles, in which examples of expert presenting performance (i.e., modeling) were shown and assessed by the students via an example rubric that was used in a previous peer assessment project. Students were asked to apply the pedagogical principles they had learned in the lectures (e.g., the principle of gradualism). In order to avoid excessively lengthy workshop sessions, the class was divided (two times eight groups, called A and B).

For the peer assessment task, students assessed their peers on the content (group level) and presentation (individual level) of the workshop in terms of rubric scores and criteria-related feedback. Assessors were told that their peer feedback would not affect their grade in the course to avoid possible apprehension. The function of the peer assessment activity was formative in nature; the university teachers' intention was to promote the idea that students can learn from their peers' use of didactic principles and presentation styles. However, to stimulate effort and justify the investment of time in the workshop, the mean peer assessment score received as assessee was taken into account for 15% of the course grade.

As for the peer assessment procedure, each student acted as an assessor seven times (i.e., a total of eight peer assessment sessions per class over 4 weeks) and once as an assessee. In order to test the hypotheses, the feedback provided by the assessors remained anonymous during the first 2 weeks but was given non-anonymously during the final 2 weeks of the intervention. Because the students were physically present in face-to-face classroom situations, guaranteeing anonymity for the assessor during peer feedback provision was facilitated through the use of Mobile Response Technology (MRT), with which assessors had the opportunity to give immediate (non)anonymous peer assessment scores and peer feedback via web-enabled devices such as smartphones, tablets, or laptops (Magaña & Marzano, 2014). In this study, the free MRT tool Socrative™ was used. Based on findings from qualitative data in previous similar studies (e.g., Author et al. 2017; Cartney 2010), in both phases, the teacher had the possibility of identifying the assessors in case unfriendly or hostile messages were given. This paper thus created a single-blind anonymous environment in which anonymity for the assessors towards the assessees—but not the teacher—is guaranteed during peer feedback provision in an initial phase.

Every peer assessment session included three steps, as depicted in Fig. 1. First, all the assessors evaluated the presenting group; second, the results (i.e., both rubric scores and feedback messages) were projected in the classroom; and third, the projected results were orally discussed. In order to move on to dialogic feedback, the teacher moderated this discussion phase by asking reflective questions. This included both content-related input to enforce a shared understanding of the criteria (e.g., What is the reason for the high number of remarks on the presentation structure?) and social-affective input, which involved acts that build up trust and scale up mutual support between assessors and assessees (e.g., Do not worry too much if your timing was criticized; we all know it takes a lot of experience to get this right.

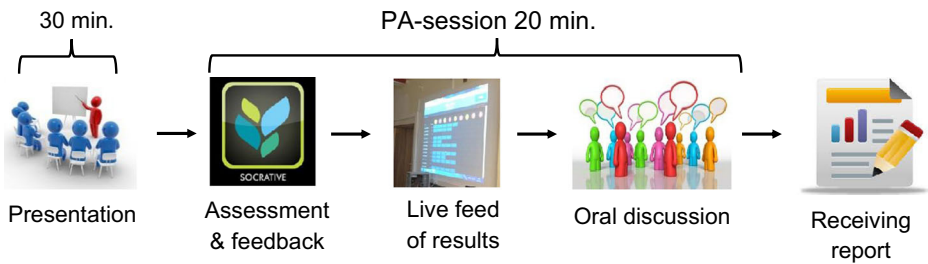


Fig. 1 Peer assessment session

Envision this task as an opportunity for practice) (Xu and Carless 2016). During the first 2 weeks (anonymous phase), the reflective questions remain broad so students can answer teachers' question, without explicitly referring to their personal feedback message in the system. In the final 2 weeks (non-anonymous phase), however, we aimed for two-way dialogic feedback in which there was a possibility to refer directly to students' input in the Socrative system. The role of the teacher was to facilitate discussions on the strengths and weaknesses of the workshops. Third, the Socrative reports were sent to the assessors.

In order to implement the best possible conditions for peer assessment to promote students' learning, two additional scaffolds were used. First, students developed their own five-level rubric with three didactic principles and three presentation-related criteria (Panadero et al. 2013) (see Appendix Tables 6 and 7). Second, as suggested by Reinholz (2015), during the workshop sessions, the assessors received three guidelines to support them while giving feedback: (1) make sure your feedback is specific and linked to the matching rubric criteria, (2) give suggestions for future improved performance, and (3) draw attention to strengths but do not be hesitant to indicate weaknesses.

Measurements

Content analysis (RQ1)

To measure the evolution of peer feedback quality, the feedback content was analyzed on four occasions (henceforth, session 1, session 2, session 3, and session 4) on a random subsample of eight out of a total 16 workshops (two groups were analyzed per session). This resulted in a database of 4390 coded segments.

The first two levels (i.e., peer feedback style and peer feedback type) of the hierarchical content analysis scheme by Gielen and De Wever (2015) have been used, with a slight modification. Peer feedback style here consists of three categories, verification, elaboration, and general, which refers to general statements that can be labeled neither verification nor elaboration and do not contain explicit evaluative information (e.g., I will give you feedback on two things). Regarding peer feedback type, there are five categories: positive verification, neutral verification, negative verification, informative elaboration, and suggestive elaboration (see Table 1). Additionally, because this study drew a distinction between two types of criteria, we added another level to our data: whether the peer feedback related to a content-related or a presentation-related criterion in the rubric.

Data was coded by the first author and an external coder who was trained for the task. For the segmentation process, a random subsample of feedback messages of five of the eight coded sessions was segmented ($N = 659$) and double-coded by both. For the segmentation process,

Table 1 Coding scheme for analyzing peer feedback content quality (modification based on Gielen and De Wever 2015)

Category	Subcategory	Description	Examples
Peer feedback style	Verification	The feedback segment is an evaluative statement expressed as a positive, neutral, or negative remark on past performance.	<i>Content-related: There were no concrete examples given.</i> <i>Presentation-related: The interaction was very well-planned in the session.</i>
	Elaboration	The feedback segment is an informative statement that builds further on a verification or remark expressed as, e.g., a question, confirmation, suggestion, or justification.	<i>Content-related: [...] because of the drawing we were able to make ourselves</i> <i>Presentation-related: Try to look a bit more at the public.</i>
	Positive	The feedback segment is a positive evaluative statement.	<i>Content-related: The theory part was very well structured.</i> <i>Presentation-related: She speaks in a relaxed way.</i>
	Negative	The feedback segment is a negative evaluative statement.	<i>Content-related: The new theory was mentioned too briefly.</i> <i>Presentation-related: A lot of content was just read out loud.</i>
	Neutral	The feedback segment is a neutral evaluative statement.	<i>Content-related: It is a pity that your response rate was that low.</i> <i>Presentation-related: The presentation had to be given in a short time.</i>
Elaboration type	Informative	The feedback segment is an informative statement that gives more details about a previous evaluative statement, without prompting the student to adapt his work.	<i>Content-related: It is a pity that you didn't show your own website (neg. verification); that made it a bit boring.</i> <i>Presentation-related: It was good that you used Prezi (pos. verification); this made it all more dynamic.</i>
	Suggestive	The feedback segment is a suggestive statement that provides more details about a previous evaluative statement with the purpose of prompting the student to adapt his work.	<i>Content-related: For future presentations, try to give some more examples.</i> <i>Presentation-related: Try to divide the text fragments more among all participating speakers.</i>

Text rendered in italics present examples of content-related or 3 presentation-related peer feedback messages.

Krippendorff's alpha was .99 for the content-related criteria and .98 for the presentation-related criteria. The hierarchical double-coding of 1977 segments resulted in the following Krippendorff's alpha values: content-related peer feedback style (.97), content-related peer feedback type of verification (.98), content-related peer feedback type of elaboration (.96), presentation-related peer feedback style (.98), presentation-related peer feedback type of verification (.98), and presentation-related peer feedback type of elaboration (.98). These were above or equal to the popular benchmark of .80 (De Swert, 2012; Landis & Koch, 1977).

Students' perceived peer feedback skills (RQ2)

Participants reported their peer feedback capability using a 10-point slider scale (0 *totally not capable*–10 *totally capable*, rounded to one decimal place) for three items (*Rate your capability of being able to formulate suggestions for improvement regarding a peer's work*; *Rate your capability of being able indicate weaknesses in a peer's work*; *Rate your capability of giving a substantiated opinion on a peer's work*). This scale was measured before the start of the intervention (Cronbach's $\alpha = .88$), after the first anonymous session (Cronbach's $\alpha = .92$), after the first non-anonymous session (Cronbach's $\alpha = .95$), and after the final session (Cronbach's $\alpha = .93$).

Students' perceptions towards anonymity, interpersonal variables, and conceptions towards peer assessment (RQ3)

These variables were measured before the intervention (henceforth, measurement time 1), after the anonymous sessions (henceforth, measurement time 2), and at the end of the non-anonymous sessions (henceforth, measurement time 3) (see Fig. 2 and Table 2), except *fear of disapproval* and *friendship marking* because students first needed to experience the specific peer assessment setting before their opinion was sounded. All items were measured using a seven-point Likert scale anchored by 1 (*totally disagree*) and 7 (*totally agree*).

In order to fully capture students' experience with the imposed transition between anonymous and non-anonymous peer assessment, the quantitative data were triangulated with qualitative data from open-ended questionnaire questions and focus groups. Participants' opinions on the transition from anonymous to non-anonymous were captured via an open-ended question after the first non-anonymous session ("How did you experience the non-anonymous setting?") and again after the second non-anonymous session ("Does your opinion about the importance of anonymity remain the same, or has it changed?"). Additionally, in the fifth week of the intervention, all students were involved in focus groups moderated by the first author. In order to enable rich group discussions, the students were split in four groups of approximately 11 students. The focus groups were organized around statements that were discussed with the participants. For the aims of this paper, the results of the statement about the

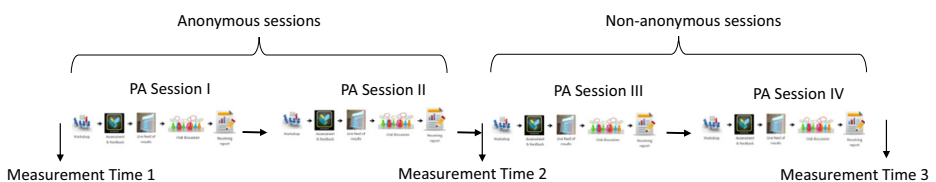


Fig. 2 Intervention overview

Table 2 Students' perceptions towards anonymity, interpersonal variables, and conceptions towards peer assessment

	Number of items	Cronbach's α measurement time 1	Cronbach's α measurement time 2	Cronbach's α measurement time 3
Importance of anonymity Example item: I think it is important that my peer feedback is given anonymously.	4	.85	.91	.87
General conceptions of peer assessment Example item: Peer assessment is useful.	6	.87	.86	.85
Trust in own evaluative capabilities Example item: I have confidence in my capacity to give feedback to peers.	4	.83	.87	.89
Trust in peers' evaluative capabilities Example item: My peers are capable of giving me feedback.	4	.68	.87	.89
Value congruency Example item: My peers have the same expectations about the goals of this peer feedback activity.	4	.65	.84	.81
Psychological safety Example item: I feel comfortable giving feedback to peers in this group.	4	.90	.84	.86
Friendship marking Example item: I gave friends higher rubric scores and more positive feedback.	4	–	.92	.85
Fear of disapproval Example item: I was afraid that when I gave low rubric scores or critical feedback, I would no longer be accepted by my peers.	4	–	.93	.94

transition approach from an anonymous to a non-anonymous peer assessment setting were analyzed (i.e., “The transition from an anonymous setting to a non-anonymous setting is a good approach to evolve towards direct interactive non-anonymous feedback setting.” All focus groups were filmed.

Data analysis

Regarding RQ1, the qualitative content data were treated quantitatively. Repeated measures ANOVAs were performed for all content categories with estimable amounts of feedback messages. The mean number of segments of a specific category was entered as a dependent variable, and the group (A or B), gender, and presentation mode (whether the student him/herself gave his/her workshop in an anonymous or non-anonymous session) were entered as between-subjects variables. For reasons of clarity, only when significant differences or interaction effects occurred are they discussed in the results section. The category “general” was not included because it was not identified among our data. Furthermore, as only small amounts of neutral verifications were found, and not during each session, these were not presented in the analyses.

Likewise, the results of RQ2 and partly RQ3a, 3b, and 3c were also analyzed via repeated measures ANOVA analyses. The qualitative data from the open-ended question (RQ3a) was brought together and organized in a five-column report with: (1) the respondent's ID, (2) the

response to open-ended question 1, (3) the response to open-ended question 2, (4) the development in perceptions towards anonymity (i.e., consequently pro or contra anonymity or a change in opinion towards anonymity), and (5) thematic coding of students' arguments for stances towards anonymity. Through a deductive analysis approach, the thematic codes used were based on the concepts discussed in the theoretical framework: peer feedback quality, perceived peer feedback skills, and interpersonal variables. This approach is particularly useful when one has specific research questions that already identify the main themes or categories used to group the data and then look for similarities and differences (Braun and Clarke 2006).

The video recordings of the four focus groups were also thematically analyzed. As mentioned earlier, only participants' responses on how they appreciate the transition approach were analyzed. The moderator used a hand-raising approach (pro/con) to identify students' opinions; this made it possible to not only gain insight into a number of arguments but also a clear, "final" individual opinion on the discussed topic.

Results

RQ1: How does peer feedback quality change over time when students consecutively practice peer assessment in anonymous and non-anonymous settings?

First, the results about the verifications. In line with hypothesis (H1), the number of negative verifications of the content-related criteria significantly increased over time [$F(2.36, 99.02) = 3.18, p = .038, \eta_G^2 = .07$] (Greenhouse-Geisser estimates of sphericity) (see Table 3). This means that after multiple sessions, students gradually dared to indicate more weaknesses in peers' work regarding the application of didactic principles. More specifically, contrast analyses revealed a significant increase between peer assessment session 1 and session 2 [$F(1, 42) = 4.94, p = .032, r = .32$] and sessions 1 and 4 [$F(1, 42) = 10.83, p = .000, r = .45$]. There was no significant difference between peer assessment sessions 2 and 4 [$F(1, 42) = 1.37, p = .249$]. In relation to the frequency of positive verifications for the content-related criteria, a similar development was found: a significant increase over time [$F(2.49, 104.47) = 7.21, p = .00, \eta_G^2 = .14$] (Greenhouse-Geisser estimates of sphericity), with significant contrasts between peer assessment session 1 and session 2 [$F(1, 42) = 18.29, p = .000, r = .56$] and between peer assessment session 1 and session 4 [$F(1, 42) = 16.21, p = .000, r = .53$]. This

Table 3 Verification type: mean amount of positive and negative verifications per student per session for content- and presentation-related criteria

Verification type		Anonymous		Non-anonymous	
		Session 1 <i>M</i> (SD)	Session 2 <i>M</i> (SD)	Session 3 <i>M</i> (SD)	Session 4 <i>M</i> (SD)
Content-related criteria	Positive	.82 (.66) ^{h, o}	2.00 (1.61) ⁿ	1.75 (1.35)	1.64 (1.30) ^o
	Negative	.16 (.37) ^{p, q}	.39 (.54) ^p	.34 (.65)	.57 (.85) ^q
Presentation-related criteria	Positive	3.32 (2.90) ^f	4.32 (2.48) ^s	5.75 (3.10) ^s	5.02 (3.46) ^f
	Negative	.93 (.97) ^{v, y}	1.66 (1.33) ^{v, w}	.89 (1.06) ^{w, x}	1.66 (1.16) ^{x, y}

The same superscripts indicate significant differences between peer assessment sessions at $p < .05$. For example, the superscript *n* means that there are significant differences between the *positive verifications* for the *content-related criteria* in peer assessment sessions 1 and 2

means that students' evolution in giving negative and positive verifications for the content-related criteria is comparable.

Similar to the content-related criteria, the negative verifications for the presentation-related criteria increase over time [$F(3, 126) = 6.32, p = .000, \eta_G^2 = .08$]. Between peer assessment session 1 and session 2, there is a significant increase [$F(1, 42) = 8.78, p = .005, r = .42$]. It is noted that the results show a significant decrease after the first non-anonymous session [$F(1, 42) = 10.14, p = .003, r = .44$] but significantly increase between sessions 3 and 4 [$F(1, 42) = 9.30, p = .004, r = .43$], resulting in the same quantity as the second anonymous session. Regarding the positive verifications for the presentation-related criteria, there was a mean effect of time [$F(3, 126) = 5.19, p = .002, \eta_G^2 = .10$]; more specifically, there was a significant increase between sessions 1 and 4 [$F(1, 42) = 5.24, p = .027, r = .33$] and between sessions 2 and 3 [$F(1, 42) = 6.27, p = .016, r = .36$].

Regarding elaborations (Table 4), the number of informative elaborations of the content-related criteria significantly increased over time [$F(2.36, 99.02) = 7.67, p = .001, \eta_G^2 = .14$] (Greenhouse-Geisser estimates of sphericity) (see Table 4). This means that after multiple sessions, students gradually gave more relevant information to help their peers in error correction (i.e., the application of a certain didactic principle). Additionally, contrast analyses revealed a significant increase between session 1 and session 2 [$F(1, 42) = 21.95, p = .003, r = .59$] and sessions 1 and 4 [$F(1, 42) = 7.93, p = .007, r = .40$], while there was a significant decrease between sessions 2 and 4 [$F(1, 42) = 4.86, p = .033, r = .32$].

For the presentation-related criteria, again, students' informative elaborations increase over time [$F(3, 126) = 7.90, p = .000, \eta_G^2 = .14$]. More specifically, there is an increase between session 1 and session 2 [$F(1, 42) = 18.95, p = .000, r = .56$] and between session 1 and session 4 [$F(1, 42) = 17.76, p = .000, r = .55$].

Concerning suggestive elaborations for the content-related criteria, a significant main effect of time was found [$F(1.32, 55.52) = 15.65, p = .00, \eta_G^2 = .26$] (Greenhouse-Geisser estimates of sphericity). However, no suggestions for improvement on the content-related criteria were given during the first session. Students gave significantly more suggestions for improvement on peers' work in session 3, compared to the anonymous session 2 [$F(1, 42) = 4.64, p = .037, r = .32$], as well in session 3, compared to session 4 [$F(1, 42) = 10.35, p = .002, r = .45$].

For the presentation-related criteria, a main effect of time was found [$F(3, 126) = 8.89, p = .000, \eta_G^2 = .16$], a significant increase was found between session 1 and session 2 [$F(1,$

Table 4 Elaboration type: mean amount of informative and suggestive elaborations per student per session for content- and presentation-related criteria

Elaboration type		Anonymous		Non-anonymous	
		Session 1 <i>M</i> (SD)	Session 2 <i>M</i> (SD)	Session 3 <i>M</i> (SD)	Session 4 <i>M</i> (SD)
Content-related criteria	Informative	.30 (.51) ^{p, q}	1.36 (1.56) ^{q, r}	1.09 (1.25)	.75 (1.06) ^{p, r}
	Suggestive	n.a.	.02 (.15) ^s	.18 (.45) ^{s, t}	.75 (1.06) ^t
Presentation-related criteria	Informative	.64 (1.18) ^{u, v}	2.09 (2.07) ^u	1.41 (1.62)	1.86 (1.82) ^v
	Suggestive	.14 (.55) ^{w, x}	.68 (.80) ^w	1.11 (1.30)	.68 (.93) ^x

The same superscripts indicate significant differences between peer assessment sessions at $p < .05$. For example, the superscript *p* means that there are significant differences between the amount of *informative elaborations* for the *content-related criteria* in peer assessment sessions 1 and 4

42) = 11.62, $p = .001$, $r = .47$], as well between session 1 and session 4 [$F(1, 42) = 12.79$, $p = .001$, $r = .48$].

To summarize, it was expected that negative verifications and informative and suggestive elaborations would increase, as was the case between peer assessment session 1 and session 2 (anonymous phase). The difference in content quality between the anonymous phase (session 2) and the end of the non-anonymous phase (session 4) is marginal. Remarkably, there was a decrease in the amount of informative elaborations for the content-related criteria between the second anonymous setting and the fourth non-anonymous session.

RQ2: How do students' perceived peer feedback skills change over time in a peer assessment setting with a transition from anonymous to non-anonymous?

When assessing students' perceived peer feedback skills, before, during (twice), and after the peer assessment sessions, the means (standard deviations in parentheses) were 6.52 (.91), 6.99 (.91), 7.08 (1.00), and 7.11 (.92), respectively. A repeated measures analysis indicates a significant main effect of time [$F(2.32, 78.84) = 10.63$, $p = .000$, $\eta_G^2 = .15$] (Greenhouse-Geisser estimates of sphericity). Contrast analyses reveal that the students perceived a significant improvement in their feedback skills from peer assessment session 1 to session 2 [$F(1, 34) = 17.03$, $p = .000$, $r = .58$], but after session 2, no further significant increases were reported [$F(1, 34) = .98$, $p = .329$].

RQ3a: How does the transition from an anonymous to a non-anonymous peer assessment affect students' perceptions regarding the attributed importance of anonymity?

For the importance level the participants attached to anonymity (mean effect time: [$F(2, 78) = 6.98$, $p = .002$, $\eta_G^2 = .12$]), the pre-test results (measurement time 1; 4.95 (1.31)) indicate that the students initially strongly preferred an anonymous peer assessment environment. There was no significant increase after the anonymous sessions (measurement time 2; 5.25 (1.31)), and there was a significant decrease after the non-anonymous sessions compared to both time 1 [$F(1, 39) = 4.74$, $p = .036$, $r = .33$] and measurement time 3 4.33 (1.47) [$F(1, 39) = 17.75$, $p = .000$, $r = .56$]. One could state that the students' importance level evolves towards a more neutral stance.

Because the quantitative results showed that all students strongly preferred anonymity in the pre-test and that there was a significant decrease in students' attributed importance towards anonymity towards a more neutral stance, students' responses to the qualitative data will provide to a detailed picture of students' arguments for this reported decrease. When evaluating students' responses to both the open-ended questions on this issue, which were asked after the first non-anonymous session and the second non-anonymous session, four different kinds of experienced evolutions were found. One group of students ($N = 23$) preferred a continued anonymous peer assessment setting, both after the first and second non-anonymous settings. The most important reasons that were given, if not in combination, are the fact that (a) they felt more comfortable doing it anonymously ($N = 10$) and (b) they felt more hesitant to speak freely in the non-anonymous session, but the content of their peer feedback messages was the same; because they can be more straightforward in an anonymous setting, they prefer to keep it anonymous ($N = 9$). A third reason that was stated frequently was the fear of negative consequences in a non-

anonymous setting ($N = 9$). A smaller group of students also indicated that they felt that they were more honest in an anonymous setting and therefore would maintain the anonymity. A second group of students ($N = 10$) indicated after the first non-anonymous session that anonymity was not so important any longer and confirmed this in their responses after the second non-anonymous session. The majority ($N = 6$) stated that the content of their feedback messages was the same, but they spent a bit more time on word choice and nuance. For that reason, students stated that for future similar peer feedback settings, non-anonymous participation would not be seen as a burden. Four students mentioned that they came to an understanding after the non-anonymous sessions that it was important to give non-anonymous and honest feedback because it otherwise loses its relevance. A third small group ($N = 9$) stated after the first non-anonymous session that anonymity was important, but their opinion changed after the final session. These students thus experienced an evolution through the sessions. The reasons for this change were that over time, the students felt they were giving better-argued feedback in the non-anonymous session ($N = 5$), felt more comfortable giving non-anonymous feedback ($N = 2$), and due to the second non-anonymous session, experienced an increase of trust in their own evaluative capabilities ($N = 2$). One student attributed less importance to anonymity after the first non-anonymous session, but this importance level increased again after the second non-anonymous session. The main reason for this was that she felt uncomfortable when she was not able to formulate a suggestion for improvement.

Regarding the results of the focus groups, with special attention to the experience as a whole, students' opinions about the transition from an anonymous to a non-anonymous peer assessment reveal that only three of the 46 participants did not agree that it helps to gradually evolve towards the aimed-for interactive peer feedback setting, with the knowledge they will experience this in real-world (work) contexts. Students' most important motives supporting the approach were related to the effects of negative interpersonal variables (e.g., friendship marking), which were less present than initially expected by the students (see also RQ3b). Another reason was that the anonymous sessions gave them time to practice their critical feedback skills, which they recognized as a skill they needed to learn. Finally, students mentioned that after experiencing both anonymous and non-anonymous sessions, the non-anonymous input was a good starting point to guide the teacher discussions. The following statements illustrate students' opinions on the transition:

Pro transition:

"It is a good thing that we learn how to cope with giving and receiving feedback. The transition makes you conscious of the fact that you need to learn how to be specific in your feedback and there is no need to put a gloss on it; in that case, it becomes useless." (ID 07)

"[The anonymous sessions] were good because we didn't know our peers that well in the beginning and because of that, we could still give honest feedback. This also allowed me to formulate critical feedback. In the third session, our feedback was non-anonymous, which was bit of shock in the beginning but in the end, everyone gave their honest opinions about each other's workshop." (ID 31)

Against transition:

"I think the transition is disadvantageous. Not in a way that I gave a different kind of feedback, but due to this I really hoped that I would not hurt my peers with my feedback, to the point that they would no longer like me." (ID01)

RQ3b: How does the transition from an anonymous to a non-anonymous peer assessment affect students' perceptions regarding their perceptions towards interpersonal variables?

Students' perceptions towards *psychological safety* indicate that students felt initially moderately comfortable giving their opinions on their peers' work (Table 5). Overall, no significant increase over time was found [$F(2, 80) = 1.72, p = .186$], and surprisingly, students who gave their workshop in the non-anonymous setting reported an overall significantly higher level of psychological safety [$F(1, 40) = 8.40, p = .006, r = .42$] (mean difference = .54).

Regarding students' *trust in their own evaluative capabilities* (main effect time [$F(1.70, 66.22) = 13.30, p = .000, \eta_G^2 = .06$ —Greenhouse-Geisser estimates of sphericity]), a significant increase was found between measurement times 1 and 2 [$F(1, 39) = 4.91, p = .033, r = .33$], although the initial trust was already high. No further change was noted between the anonymous and non-anonymous settings. Regarding students' *trust in peers' evaluative capabilities*, no significant changes over time were found [$F(2, 78) = .864, p = .426$].

Regarding students' *value congruency about the peer assessment criteria*, a positive evolution over time was found [$F(2, 78) = 14.13, p = .000, \eta_G^2 = .20$]. More specifically, a significant increase was found from measurement time 1 to time 2 [$F(1, 39) = 22.19, p = .000, r = .60$].

Although the level of perceived *friendship marking* was low after the non-anonymous sessions, contrary to our hypothesis, a significant increase was found between the anonymous and non-anonymous sessions [$t(41) = 2.24, p = .031, \text{Cohen's } d = .34$].

As expected, students' level of *fear of disapproval* significantly diminished after the non-anonymous (measurement time 3) sessions, compared to the preceding anonymous sessions (measurement time 2) [$t(43) = -3.734, p = .001, \text{Cohen's } d = .56$].

RQ3c: How does the transition from an anonymous to a non-anonymous peer assessment affect students' perceptions regarding their general conceptions towards peer assessment?

Concerning students' conceptions towards peer assessment, a significant main effect of time was found [$F(2, 70) = 13.30, p = .000, \eta_G^2 = .17$]. The means (with standard deviations in parentheses) were 4.90 (1.02), 5.39 (.80), and 5.53 (.71), respectively. As expected, a significant increase was found between measurement time 1 and 2 [$F(1, 35) = 13.57, p = .001, r = .53$] and times 1 and 3 [$F(1, 35) = 21.29, p = .00, r = .61$]. Furthermore, the expected increase between measurement times 2 and 3 was non-significant [$F(1, 35) = 1.44,$

Table 5 Descriptive data on interpersonal variables

	Measurement Time 1	Measurement Time 2	Measurement Time 3
Trust in own evaluative capabilities	5.02 (.77)	5.28 (.57)	5.28 (.52)
Trust in peers' evaluative capabilities	5.12 (.79)	5.29 (.79)	5.21 (.67)
Psychological safety	4.61 (.95)	4.91 (.86)	4.77 (.89)
Value congruency	4.63 (.74)	5.23 (.67)	5.16 (.69)
Friendship marking	/	2.22 (.81)	2.82 (1.17)
Fear of disapproval	/	3.60 (1.28)	3.24 (1.38)

$p = .238$]. This means that during the intervention with two phases, there was no decrease in students' conceptions towards peer assessment.

Discussion

The purpose of this study was to explore the effects of disappearing anonymity on peer feedback quality, while also exploring students' conceptions about the process. For the first hypothesis (H1) about the evolution of peer feedback content quality over time, the results regarding verifications and elaborations will be discussed separately. It was found regarding the verifications that the peer feedback content quality increased because students offered more positive and negative verifications over time, which supports the posited hypothesis of the effect of faded anonymity. Interestingly, and against H1, the amount of negative verifications of the presentation-related criteria significantly decreased in the session following the transition from anonymous to non-anonymous sessions. This finding suggests that students might benefit from experiencing both types of settings in order to non-anonymously point out negative aspects of their peers' work. Furthermore, the amount of negative verifications in the second non-anonymous session was equal to or higher than the negative verifications in the second anonymous session. This finding again favors the paper's transition approach because the feedback became more descriptive of the actual performance, rather than just pointing out positive aspects (Author, 2017). The fact that the amount of positive verifications remained stable (content-related criteria) or increased (presentation-related criteria) during the non-anonymous sessions points to the fact that the students felt the need to pinpoint both positive and negative aspects of their peers' work, which corroborates the findings of Gielen and De Wever (2015).

Regarding the elaborations, the hypothesis of an overall increase was also confirmed. The significant decrease in the informative elaborations between the second anonymous session and the second non-anonymous session might be related to the fact that due to multiple experiences with the workshops, students gave significantly more elaborative suggestions to improve peers' work, rather than only informing them why certain aspects of their work were positive or negative. This points to an improvement in students' evaluative expertise over time (Sadler 2010) as they related their evaluations to the evaluations of others; they reflected whether their judgements were appropriate or not, looking for ways to improve future feedback content and wondering what they had missed in making their judgements that others had noticed (Boud et al. 2013). In this study, peer feedback quality was defined based on the presence of structural components in a peer feedback message, as defined by Gielen and De Wever (2015). Because no impact of the peer feedback from either the perspective of the assessor or the assessee on similar future performances was taken into account (Evans 2013), the current quality measurement can only be interpreted in terms of its potential impact for future performances. Researchers are encouraged to include an explicit performance measure in future research.

Furthermore, analyzing the assessee's mindful processing of the content of the Socratic feedback reports would be a valuable topic for future research in order to explicitly close the feedback loop. Mindful cognitive processing was recently explored by Bolzer et al. (2015) and refers to "how deeply the peer feedback has been cognitively processed and understood" (p. 425). Their study showed that eye-tracking methodologies provide valid measures to deduce mindful cognitive processing (e.g., during the reading phase when processing peer feedback, for example, when being confronted with contrasting feedback in the Socratic report). In that

sense, this study focused on the supply side of peer assessment (i.e., the development of the evaluative expertise of assessors by offering practice opportunities), rather than the receiver side (i.e., focusing on performance improvement of the assessee).

This focus was mirrored in the nature of this paper's peer assessment task, as well as the nature of the used assessment criteria, and should be taken into account when interpreting these results. Previous research has established that peer assessment is a complex learning task that requires high-level cognitive processing, and studies have shown that if students do not master domain-specific knowledge, having to perform a peer assessment of these domain-specific tasks may hinder their learning and performance (van Zundert, Könings, Sluijsmans & van Merriënboer, 2012). For this reason, students were acknowledged as novices within the discipline and expected to assess a peer's work in terms of applying didactical principles, rather than assessing the peers' knowledge on the workshop subject, and in terms of presentation. Having different opinions is enriching, as long as peers are transparent about how they relate their judgment to the mutually discussed criteria (Gielen et al., 2011). As such, students are key consumers and producers of the formative assessment information (Andrade, 2010), in which the accuracy of peer feedback messages is defined in terms of their appropriateness to the assessment criteria, rather than being determined by the subject-related expertise of the teacher (or another expert).

Of course, this approach narrows the generalizability of this paper's results and its application in other disciplines. For example, in science education, the provision of feedback with scientifically correct content is crucial, and when analyzing peer feedback content quality, this accuracy component should also be included in the coding process (e.g., Hovardas, Tsivitanidou & Zacharia, 2014). Furthermore, the value of our approach is supported by the findings of Gielen et al. (2010), who found that justification was superior to the accuracy of comments in having a positive impact on performance. Thus, peer feedback goes beyond a corrective function to one that promotes critical discourse (Gan & Hattie, 2014). The specific framing of the paper's peer assessment activities implies that peer feedback needs to be seen in the context of negotiating meaning and connecting ideas, rather than providing the "right" answers.

The second hypothesis (H2) regarding students' perceived evolution in peer feedback skills can be partially maintained. Because students already rated themselves highly positive in session 1, there was only a significant increase in session 2. More importantly, there was no decrease in perceived improvement after sessions 3 and 4, suggesting that students did not feel hindered by the non-anonymous setting that was created in sessions 3 and 4.

As expected, the questionnaire data show that students' importance level significantly decreased after session 2 (anonymous) (H3a). This means that overall, students' opinions on the importance of anonymity resulted in a rather neutral stance. The qualitative data, however, showed a more diverse picture: half of the students preferred a continued anonymous peer assessment setting, both after the first and the second non-anonymous settings. The second half of the studied population found anonymity less important after the first or second non-anonymous session. The data of the focus groups clearly show that students differentiated between the amount of importance they attribute towards anonymity, and whether a transition approach from an anonymous to a non-anonymous setting is seen as a good approach, to evolve towards direct interactive non-anonymous feedback settings. Only three out of the 46 students did not agree about this. Although the initial hypothesis of an overall increase of students' peer assessment conceptions over time was not confirmed (H3c), the fact that

students' peer assessment conceptions did not decrease after the non-anonymous phase was initiated again favors the implementation of a transition approach.

Although previous research has found that anonymity might help to reduce interpersonal burdens, the results of this study cannot confirm the initial hypothesis (H3b). That is, students' perceptions towards positive interpersonal variables (psychological safety, trust in own and others' evaluative capabilities, and value congruence) were initially already—moderately—positive; only for trust in one's own evaluative capabilities and value congruency was an increase found after the anonymous sessions, although these significant differences are practically irrelevant. Regarding the negative interpersonal variables, the opinions about friendship marking and fear of disapproval were low after the anonymous sessions and remained low after the non-anonymous sessions. These findings might be explained by the implemented peer assessment scaffolds (i.e., active involvement in rubric criteria development and guiding questions) and the fact that the sample already had already positive attitudes towards peer assessment (as confirmed by the pre-test results on peer assessment conceptions) and these positive attitudes helped to overcome interpersonal burdens in all phases of the intervention. The fact that almost all students strongly appreciated the transition approach suggests that it met their need for practice in a safe environment. Moreover, because the students did not really expect an influence of interpersonal variables, it can be expected that in settings in which this interpersonal burden is more present, the application of this transition approach might be even more valuable. Finally, the results seem to suggest that the claim by Panadero (2016), that anonymity should be “considered carefully in terms of the learning benefits (it might produce or mitigate),” is crucial.

Implications

This paper's findings are important for educational practice because they add to the field's much-needed understanding of peer assessment as a powerful pedagogical practice (Panadero and Brown 2017). This study also confirms earlier findings that practice is an important component in peer assessment implementations (Gielen and De Wever 2015; Liu and Carless 2006). Second, when a transition from an anonymous to a non-anonymous peer assessment environment is facilitated, students' peer feedback quality in the anonymous phase increases over time and the peer feedback quality in the non-anonymous sessions eventually becomes comparable. As stated by Panadero and Brown (2017), it is important for teachers themselves to practice peer assessment with other teachers for their professional development in order to give them greater awareness of the interpersonal dynamics within peer assessment. Exploring our transition approach with anonymity in different contexts and with different groups is certainly part of that. This would help teachers to decide whether they will opt for anonymous modes of feedback, depending on the time available, and the specific characteristics of the student population.

Limitations and directions for future research

Given the sample size, the gender bias of the sample (mostly female), and the fact that this was a peer group assessment setting, the findings of this research should be interpreted with caution. First, because of the peer group assessment setting, it is possible that the feedback effects were diluted at the group level. This study did not examine how students coped with the received feedback within their group (e.g., who took responsibility for

possible negative remarks). This should be included in future lines of research in order to be able to disentangle these effects at the individual level. Second, due to the relatively short duration of the study (two times two sessions), the effect of the peer feedback content quality increase might be confounded with the effects of anonymity in the first 2 weeks. However, installing a control condition with a reverse sequence would negate the theoretical claim of offering anonymity. Finally, facilitating a non-anonymous setting from the start would go against findings by Vanderhoven et al. (2015) and Raes et al. (2013) in similar settings in which students' attributed importance towards anonymity for the assessors was proven.

With regard to the provision of anonymity, based on the theoretical arguments of previous work, this study focused on offering anonymity to the assessor when providing feedback to his/her peers in face-to-face synchronous peer assessment settings. The students in this study clearly differentiated between the attributed importance of anonymity and whether a transition approach from anonymous to non-anonymous is seen as a good approach towards direct interactive non-anonymous feedback settings. Furthermore, a decrease in attributed importance of anonymity was noted. As a consequence, a future exploration could introduce a dynamic self-choice anonymity mode in which students themselves decide when to give their MRT input (non-)anonymously. In itself, the MRT software does not have this option, but one could offer students the opportunity to use an unidentifiable nickname or number, instead of their real name. As such, within the whole peer assessment procedure, students could get the opportunity to decide to give up their anonymity twice: when giving feedback through the MRT system and when participating in the oral discussion phase. It would also be interesting to explore the impact of anonymity towards the assessee as a scaffold that gradually evolves towards an authentic feedback setting. For example, within asynchronous settings, students could be offered fictitious products, then a real product by a peer from another institution, and finally a product by a "real" peer.

Additionally, the interactive exchanges during the oral discussions with the teacher were not actively monitored. A profound and detailed qualitative analysis could be valuable for providing insight into the development of dialogic feedback processes and the role of the teacher in this phase. During this paper's intervention, the involved teachers moderated the oral discussion phase by asking reflective questions; this included both content-related input, to enforce a shared understanding of the criteria, and social-affective input involving acts that build up trust and scale up mutual support between assessors and assessees. However—and this can be considered a limitation of the intervention—the effect of these moderating actions was neither measured nor analyzed, even though the contribution of these actions might be substantive. For example, in a study by van Ginkel et al. (2015), the added value of a teacher in questioning, intervening, and guiding students in the feedback process led to higher quality feedback. These authors concluded that implementing peer assessment requires both training students and ongoing monitoring and feedback regarding the efficacy of their evaluative efforts. In a previous study exploring the role of the teacher in a peer assessment activity, Xu and Carless (2016) outlined specific teacher-enabling processes that should be measured in future qualitative studies: (a) cognitive scaffolding, which involves strategies to promote students' disciplinary understanding (e.g., rephrasing and modeling), improve self-regulated capacities (e.g., helping students use the criteria), and provide feedback regarding content quality (e.g., providing suggestions for improvement), and (b) social-affective support involving

practices that build up students' trust in teachers and peers (e.g., showing interpersonal caring—Murdock, Stephens & Grotewiel, 2016), while cultivating students' rational attitudes towards critical feedback. These essential processes aim to enhance teacher and student feedback literacy, which is a worthwhile path for future research.

Conclusion

The creation of a safe and supportive learning environment in which students feel comfortable and confident to assess their peers is essential for the quality of peer assessment activities. Although offering anonymity to assessors in these activities has been recommended in previous research due to its positive effect on students' peer assessment conceptions and interpersonal variables (e.g., van Gennip et al. 2009), currently, research is lacking on the actual feedback behavior of students in anonymous peer assessment settings. Moreover, the use of anonymity as a temporary scaffold to gradually evolve towards a dialogic feedback environment (e.g., Howard et al. 2010) has not yet been explored.

The content analysis of the peer feedback messages revealed that the quality of the peer feedback increases in the anonymous phase, and that over time, the feedback quality in the non-anonymous sessions was comparable. The focus group results confirm that students appreciate the anonymous phase to practice their peer feedback skills in order to produce high-quality feedback. Our findings suggest that anonymity can be used as a valuable scaffold to ease students' importance level towards anonymity and their associated need for practice in a safe environment. Consequently, in a synchronous setting, this study has found that depending on the available time to organize peer assessment and the specific characteristics of the student population (i.e., in groups in which interpersonal burdens might hinder the peer assessment activity), teachers could choose between only anonymous peer assessment (short) or a sequence of anonymous and non-anonymous peer assessment (long), depending whether they intend to focus on the creation of a safe learning environment or work towards a dialogic non-anonymous peer assessment environment.

Acknowledgments The first author's research was funded by Ghent University BOF fund number BOF13/24J/115. The second author's research was funded by the Spanish Ramón y Cajal program number RYC-2013-13469.

Appendix

Table 6 Rubric criteria

Presentation-related criteria

1. Body language
 2. Use of voice and language
 3. Interaction with audience
- Content-related criteria
4. Principle of illustration and specificity
 5. Principle of activity
 6. Principle of gradualism
-

Table 7 Example: principle of gradualism

Score 1	Score 2	Score 3	Score 4	Score 5
There is a total lack of structure, which makes it impossible for the audience to understand the content. The group lacked time management.	There was an attempt to structure the workshop, but they failed. Time management was poor.	The material was logically built up, though its presentation lacked structure. The session ended on time, though the use of time was not efficient.	There was a good build up and structure. The group presented the material in an effective way. Time management was good.	The workshop had a clearly structured build up. Time management was optimal.

References

- Ainsworth, S., Gelmini-Hornsby, G., Threapleton, K., Crook, C., O'Malley, C., & Buda, M. (2011). Anonymity in classroom voting and debating. *Learning and Instruction, 21*, 365–378. doi:10.1016/j.learninstruc.2010.05.001.
- Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic selfassessment and the self-regulation of learning. *Handbook of Formative Assessment*, 1–18.
- Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education, 27*(5), 427–441. doi:10.1080/0260293022000009302.
- Bolzer, M., Strijbos, J. W., & Fischer, F. (2015). Inferring mindful cognitive-processing of peerfeedback via eye-tracking: role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning, 31*(5), 422–434. doi:10.1111/jcal.12091.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. doi:10.1080/1713695728.
- Boud, D., & Soler, R. (2015). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 1–14* doi:10.1080/02602938.2015.1018133.
- Boud, D., Lawson, R., & Thompson, D. G. (2013). Assessment & evaluation in higher education does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education, 37–41*. doi:10.1080/02602938.2013.769198.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(May 2015), 77–101. doi:10.1191/1478088706qp063oa.
- Brown, S. (2004). Assessment for learning. *Learning and Teaching in Higher Education, 1*, 2004–2005.
- Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(1), 1–14. doi:10.1080/03075071003642449.
- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education, 35*(5), 551–564. doi:10.1080/02602931003632381.
- Chester, A., & Gwynne, G. (2006). Online teaching: encouraging collaboration through anonymity. *Journal of Computer-Mediated Communication, 4*(2), 0–0. doi:10.1111/j.1083-6101.1998.tb00096.x.
- Cheng, K. H., & Tsai, C. C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology, 28*, 599–618.
- Cho, Y. H., & Cho, K. (2010). Peer reviewers learn from giving comments. *Instructional Science, 39*(5), 629–643. doi:10.1007/s11251-010-9146-1.
- De Swert, K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. Retrieved from <http://www.polcomm.org/wpcontent/uploads/ICR01022012.pdf>
- Dochy, F., Segers, M., & Sluismans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education, 24*, 331–350. doi: 10.1080/0307507991233179935.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*, 70–120.
- Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education & Training International, 32*(2), 175–187. doi:10.1080/1355800950320212.

- Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. *Assessment & Evaluation in Higher Education*, 38(5), 611–630. doi:10.1080/02602938.2012.674484.
- Freeman, M., & McKenzie, J. (2000). Self and peer assessment of student teamwork: designing, implementing and evaluating SPARK, a confidential, web based system. In *Flexible learning for a flexible society*. Retrieved from <http://ascilite.org.au/aset-archives/confs/aset-herdsa2000/procs/freeman.html>
- Gielen, M., & De Wever, B. (2015). Structuring peer assessment: comparing the impact of the degree of structure on peer feedback content. *Computers in Human Behavior*, 52, 315–325. doi:10.1016/j.chb.2015.06.019.
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304–315. doi:10.1016/j.learninstruc.2009.08.007.
- 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. doi:10.1016/j.tate.2013.07.008.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction* (pp. 249–270). New York: Routledge. doi:10.4324/9780203839089.ch13.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. doi:10.3102/003465430298487.
- Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: making learning visible. *Studies in Educational Evaluation*, 38, 21–27. doi:10.1016/j.stueduc.2012.04.001.
- Hogg, M. A., Fielding, K. S., Johnson, D., Masser, B., Russell, E., & Svensson, A. (2006). Demographic category membership and leadership in small groups: a social identity analysis. *The Leadership Quarterly*, 17(4), 335–350. doi:10.1016/j.leaqua.2006.04.007.
- Hogg, M. A., van Knippenberg, D., & Rast, D. E. (2012). The social identity theory of leadership: theoretical origins, research findings, and conceptual developments. *European Review of Social Psychology*, 23(1), 258–304. doi:10.1080/10463283.2012.741134.
- Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers and Education*, 71, 133–152. doi:10.1016/j.compedu.2013.09.019.
- Hosack, I. (2004). The effects of anonymous feedback on Japanese university students' attitudes towards peer review. In R. Hogaku (Ed.), *Language and its universe* (pp. 297–322 3). Kyoto: Ritsumeikan Hogaku.
- Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to promote peer feedback: pre-service teachers' comments in asynchronous computer-mediated communication. *Journal of Educational Computing Research*, 43(1), 89–112. doi:10.2190/EC.43.1.f.
- Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, 33(4), 692–724. doi:10.2307/256287.
- King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory Into Practice*, 41(1), 33–39. doi:10.1207/s15430421tip4101_6.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310.
- Li, L. (2016). The role of anonymity in peer assessment. *Assessment & Evaluation in Higher Education*, 2938(April), 1–12. doi:10.1080/02602938.2016.1174766.
- Liu, N.-F., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education*, 11, 279–290.
- Lu, L., & Bol, R. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning*, 6(2), 100–115. doi:10.1017/CBO9781107415324.004.
- Magana, S., & Marzano, R. J. (2014). Using Polling Technologies to Close Feedback Gaps. *Educational Leadership*, 82–83.
- Miyazoe, T., & Anderson, T. (2011). Anonymity in blended learning: who would you like to be? *Educational Technology & Society*, 14(2), 175–187.
- Murdock, T. B., Stephens, J. M., & Grotewiel, M. M. (2016). Students Dishonesty in Face of Assessment: Who, Why and What We Can Do About It. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment*. New York, NY: Routledge.
- Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125–143). Mahwah, NJ: Erlbaum.
- Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122. doi:10.1080/02602938.2013.795518.

- Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: a review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Human factors and social conditions of assessment*. New York: Routledge (pp. 1–39). New York, NY: Routledge.
- 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. doi:10.1007/s10212-015-0282-5.
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195–203. doi:10.1016/j.stueduc.2013.10.005.
- Pearce, J. A. (2013). Using social identity theory to predict managers' emphases on ethical and legal values in judging business issues. *Journal of Business Ethics*, 112(3), 497–514. doi:10.1007/s10551-012-1274-x.
- Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. (2005). Formative peer assessment in a CSCL environment: a case study. *Assessment & Evaluation in Higher Education*, 30(4), 417–444. doi:10.1080/02602930500099219.
- Raes, A., Vanderhoven, E., & Schellens, T. (2013). Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Studies in Higher Education*, 0(0), 1–16. doi:10.1080/03075079.2013.823930.
- Reinholz, D. (2015). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*. 1–15. doi:10.1080/02602938.2015.1008982.
- Roberts, L. D., & Rajah-Kanagasabai, C. J. (2013). "I'd be so much more comfortable posting anonymously": identified versus anonymous participation in student discussion boards. *Australasian Journal of Educational Technology*, 29(5). doi:10.14742/ajet.452.
- Rotsaert, T., Panadero, E., Schellens, T., & Raes, A. (2017). "Now you know what you're doing right and wrong!" Peer feedback quality in synchronous peer assessment in secondary education. *European Journal of Psychology of Education*. doi: 10.1007/s10212-017-0329-x.
- Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35, 535–550.
- Sluijsmans, D. M. A. (2002). *Student involvement in assessment: the training of peer assessment skills*. (Unpublished doctoral dissertation). Open University of the Netherlands, Heerlen.
- Strijbos, J.-W., & Sluijsmans, D. (2010). Unravelling peer assessment: methodological, functional, and conceptual developments. *Learning and Instruction*, 20, 265–269.
- Strijbos, J.-W., Narciss, S., & Dünneber, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20, 291–303.
- Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence*, 33, 135–145. doi:10.1016/j.adolescence.2009.04.004.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249. doi:10.2307/1170598.
- Topping, K. (2009). Peer assessment. *Theory Into Practice*, 48(1), 20–27. doi:10.1080/00405840802577569.
- Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction*, 20(4), 339–343. doi:10.1016/j.learninstruc.2009.08.003.
- Tsai, C.-C., Lin, S. S., & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education*, 38(1–3), 241–252. doi:10.1016/S0360-1315(01)00069-0.
- Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education*, 51, 1804–1817.
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: the influence of interpersonal variables and structural features. *Educational Research Review*, 4, 41–54. doi:10.1016/j.edurev.2008.11.002.
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learning and Instruction*, 20, 280–290. doi:10.1016/j.learninstruc.2009.08.010.
- van Ginkel, S., Gulikers, J., Biemans, J., & Mulder, M. (2015). The impact of the feedback source on developing oral presentation competence. *Studies in Higher Education*, 1-15, doi:10.1080/03075079.2015.1117064.
- Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20, 316–327.
- Vanderhoven, E., Raes, A., Montrieux, H., Rotsaert, T., & Schellens, T. (2015). What if pupils can assess their peers anonymously? A quasi-experimental study. *Computers & Education*, 81, 123–132. doi:10.1016/j.compedu.2014.10.001.

- van Zundert, M. J., Konings, K. D., Sluijsmans, D. M. A., & van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies*, 38(5), 541–557. doi:10.1080/03055698.2012.654920.
- Vickerman, P. (2009). Student perspectives on formative peer assessment: an attempt to deepen learning? *Assessment & Evaluation in Higher Education*, 34, 221–230. doi:10.1080/02602930801955986.
- Xu, Y., & Carless, D. (2016). “Only true friends could be cruelly honest”: cognitive scaffolding and social-affective support in teacher feedback literacy. *Assessment & Evaluation in Higher Education*. 1–13. doi:10.1080/02602938.2016.1226759.
- Yu, F.-Y., & Liu, Y.-H. (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. doi:10.1111/j.1467-8535.2008.00905.x.
- Yu, F.-Y., & Sung, S. (2015). A mixed methods approach to the assessor’s targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments*. 1–18. doi:10.1080/10494820.2015.1041405.
- Zhang, Y., Fang, Y., Wei, K.-K., & Chen, H. (2010). Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities. *International Journal of Information Management*, 30(5), 425–436. doi:10.1016/j.ijinfomgt.2010.02.003.