Studies in Educational Evaluation 83 (2024) 101410

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Investigating anonymity in formative and summative peer assessment: Effects on university students' social-affective factors, perceptions and preference

Juan Fraile [a,*,1], Meritxell Monguillot [b,2], Carles González-Arévalo [b,3], Paula Lehane [c,d,4], Ernesto Panadero [c,e,f,5]

[a] Universidad Francisco de Vitoria, Madrid, Spain
[b] Department of Education, Institut Nacional d'Educació Física de Catalunya (INEFC), Universitat de Barcelona, Barcelona, Spain
[c] Centre for Assessment Research, Policy and Practice in Education (CARPE), Institute of Education, Dublin City University, Dublin, Ireland
[d] School of Inclusive and Special Education, Dublin City University, Dublin, Ireland
[e] Facultad de Educación y Deportes, Universidad de Deusto, Bilbao, Spain
[f] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

ARTICLE INFO

Keywords:
Peer assessment
Perceptions
Peer grading
Summative peer assessment
Social variables
Interpersonal variables

ABSTRACT

This study examines the impact of anonymity on students' social-affective processes, as well as their perceptions and preferences regarding peer assessment and grading for both formative and summative purposes. 177 higher education students participated in this quasi-experimental study. They participated in two peer assessment activities after receiving training on the provision of effective feedback. The first peer assessment activity was formative in nature, while the second was summative. The study had two conditions: anonymous peer assessment and non-anonymous peer assessment. Results indicate that although 73.6 % of all students (90.5 % in the anonymous condition and 56.5 % in the non-anonymous condition) preferred anonymous peer assessment, this preference did not significantly impact perceptions of fairness or peer pressure. This challenges assumptions around the effects of anonymity. Interestingly, non-anonymous settings were still favourably viewed, with the open-ended responses indicating the value of face-to-face dialogue. This suggests that non-anonymous, highly formative peer assessment requiring more interaction may lead to better social and interpersonal outcomes.

Peer assessment is defined as an evaluative practice among individuals of similar status, which generally involves students judging each other's work (Topping, 1998). This approach is considered a powerful pedagogical and learning strategy among higher education students, one that promotes performance across a range of outcomes. For example, several meta-analyses have demonstrated the positive relationship between peer assessment activities and academic performance (Double et al., 2020; Li et al., 2020; Yan et al., 2022). Other research has shown how peer assessment can also benefit self-regulated learning and evaluative judgement (Panadero & Broadbent, 2018). However, it is important to acknowledge that peer assessment is a social process. While those peer assessment can facilitate cognitive and emotional growth in line with social development theory (Vygotsky, 1978) several interpersonal and intrapersonal effects can negatively impact learning outcomes (Alqassab et al., 2023; Panadero et al., 2023). These include low trust in the peer as an assessor, tensions with peers, friendship, peer pressure, etc.

Anonymity has been proposed as a solution to some of these negative effects, as it hides much of the social elements to the students, be they in the role of assessor or assessee (Rotsaert et al., 2018). A review of the effects of anonymity in peer assessment shows a slight advantage for anonymity in some circumstances but also the need for a much more

* Correspondence to: Universidad Francisco de Vitoria, Carretera Pozuelo-Majadahonda km. 1.8 – 28223 Pozuelo de Alarcón, Madrid, Spain.
E-mail address: juan.fraile@ufv.es (J. Fraile).
[1] https://orcid.org/0000-0002-7212-7380
[2] https://orcid.org/0000-0002-1114-6529
[3] https://orcid.org/0000-0003-1652-811X
[4] https://orcid.org/0000-0003-0856-3505
[5] https://orcid.org/0000-0003-0859-3616

https://doi.org/10.1016/j.stueduc.2024.101410
Received 24 December 2023; Received in revised form 26 September 2024; Accepted 4 October 2024
Available online 25 October 2024
0191-491X/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

detailed investigation (Panadero & Alqassab, 2019). Importantly, peer assessment can be implemented with formative and summative purposes, and anonymity might affect the effects of these purposes differently. To address this call to action, this quasi-experimental study aims to explore the effects of anonymous and non-anonymous formative and summative peer assessment approaches on university students' social-affective factors, perceptions and preferences for peer assessment.

1. Peer assessment: formative and summative purposes

Formative peer assessment refers to educational practices where students evaluate each other's work with the primary goal of providing constructive feedback to enhance learning and performance during the instructional process (Alqassab & Panadero, 2020). This method fosters a collaborative learning environment, encouraging students to engage in critical thinking, self-regulation, and reflective practices (Topping, 2013). Research has consistently shown that formative peer assessment positively influences academic performance and student motivation (see meta-analysis of Double et al., 2020; Li et al., 2020; Yan et al., 2022). By promoting continuous feedback and active participation, formative peer assessment helps students internalise assessment criteria and improve their understanding of the subject matter (Alqassab et al., 2023).

Summative peer assessment, also known as peer grading or marking, involves students evaluating their peers' work to assign a grade or score that contributes to the final evaluation of performance, typically occurring at the end of a learning period (Panadero, 2016). This form of assessment has also been found to positively impact student performance, as demonstrated by the meta-analysis of Sanchez et al. (2017), who reported that summative peer assessment enhances academic achievement on subsequent tests by promoting deeper engagement with assessment criteria and fostering accountability among students. Additionally, research indicates that the summative approach improves students' evaluative skills, aligning closely with teacher assessments when clear criteria and training are provided (Li et al., 2016).

Both formative and summative peer assessment are usually investigated separately in the literature. Specifically, a review by Alqassab et al. (2023) found that the examination of formative peer assessment occurred in 59.1 % of studies. Approximately 36.1 % of studies investigated summative peer assessment in 36.1 %. Only 4.8 % of cases examined both formative and summative assessment approaches. Those that do combine formative and summative approaches in peer assessment interventions tend to implement formative peer assessment during the learning process while using the outcomes to make summative judgments (e.g., Chinn, 2005). The integrative use of formative and summative peer assessment holds potential benefits, such as enhancing student engagement, accountability, and academic achievement. Unfortunately, research in this area remains limited and, furthermore, the possibility of conducting the peer assessment with a variety of design elements poses a challenge to drawing robust conclusions (Alqassab et al., 2023). In this study, we will explore the integration of both approaches with particular attention to one, specific design element – anonymity.

2. Peer assessment and anonymity

Anonymity in peer assessment has been explored for its potential to enhance the quality and fairness of feedback by reducing social biases and pressures among students. According to Panadero and Alqassab (2019), anonymity can provide several advantages, such as delivering more critical peer feedback and increasing the comfort level of students by mitigating the fear of negative repercussions. This is particularly important in educational settings where the interpersonal dynamics between assessors and assessees can influence the assessment outcomes. However, a review by Panadero and Alqassab (2019) highlighted mixed results regarding the impact of anonymity on academic performance, with some studies showing better results for anonymous conditions and others indicating no significant differences.

Furthermore, anonymity has been shown to affect the accuracy of peer grading, where anonymous assessments often result in lower scores, suggesting that students may be more critical when their identity is concealed (Lu & Bol, 2007; Peterson & Peterson, 2011). However, the overall impact on grading accuracy is inconsistent, with some studies indicating that non-anonymous peer grading correlates more closely with teacher evaluations (Li et al., 2016; Vanderhoven et al., 2015). Therefore, more research is needed.

3. The present study

In our study, we explore students' perceptions and preferences regarding anonymous and non-anonymous peer assessment. Our research integrates the implementation of both formative and summative peer assessment, with participants divided into two groups: one experiencing anonymous peer assessment and the other non-anonymous peer assessment. Particular attention will be given to students' perceptions and preferences on these particular elements of peer assessment.

3.1. Students' perceptions

Formative peer assessment and modality (anonymous vs. non-anonymous). It becomes evident that the use of anonymity in peer assessment presents a complex interplay of intrapersonal and interpersonal outcomes. The effects of anonymity are not straightforward and are influenced by various factors, including the nature of the task and the assessment's overall design and implementation (Panadero, 2016; Panadero & Alqassab, 2019). Panadero and Alqassab's (2019) review note that anonymity in peer assessment may lead to more critical feedback and increased self-perceived social effects, such as comfort and reduced peer pressure, especially in higher education settings. This promotes a safer environment for honest and open feedback (Rotsaert et al., 2018; Vanderhoven et al., 2015). However, this is contrasted by the findings of Yu and Sung (2016) and Vanderhoven et al. (2015), where anonymity did not consistently translate to observable social benefits, particularly when assessed using objective measures like sociograms. This highlights the dichotomy between subjective perceptions and objective social metrics in evaluating the impact of anonymity in peer assessment.

Su's (2023) study into anonymous peer feedback demonstrates how the concealment of identities can influence emotional states and group dynamics. Anonymity was found to enhance students' self-confidence and improve peer relations, as evidenced by reduced self-reports of peer pressure and increased positive attitudes (Panadero & Alqassab, 2019; Vanderhoven et al., 2015). This aligns with the deindividuation effect and social identity theory, which suggest that anonymity may lead to less inhibited behaviours and balanced group engagement (Hogg et al., 2012; Jessup et al., 1990;). However, Lin (2018) notes that anonymity doesn't always result in positive emotional expressions, as anonymous feedback might be more neutral and less emotionally charged.

In contrast, Senden et al. (2023) focus on the psychological safety and trust aspects of peer assessment, underlining that these factors are crucial for effective peer interactions and learning outcomes. They assert that psychological safety, defined as the extent to which students feel safe giving and receiving feedback, is a direct predictor of students' conceptions of peer assessment and indirectly influences perceived learning (Cheng & Tsai, 2012; van Gennip et al., 2009). Although anonymity can alleviate certain concerns like peer pressure and fear of disapproval (Su, 2023; Vanderhoven et al., 2015), it doesn't necessarily enhance psychological safety or trust. In fact, Yu and Wu (2011) found that real names in peer assessment fostered more positive views towards assessors, including higher trust.

Summative peer assessment and modality (anonymous vs. non-anonymous). Regarding summative peer assessment and students'

perceptions, Panadero (2016) suggests that it might have adverse effects on these intrapersonal and interpersonal variables. However, a pedagogically richer approach, as concluded by Li et al. (2016), could mitigate these outcomes typically associated with a more traditional implementation of summative peer assessment (Alqassab & Panadero, 2020).

Concerning the additional impact of anonymity, the review by Panadero and Alqassab (2019) echoes the findings of Li et al. (2016), which found that non-anonymous summative peer assessment led to greater accuracy than anonymous evaluations. However, Panadero and Alqassab (2019) emphasize the importance of holding assessors accountable for their grading accuracy to ensure students do not tend towards overscoring; this aspect could explain the results observed by Li et al. (2016). Furthermore, Panadero and Alqassab (2019) reveal that contrary to the assumption that anonymity would enhance the positive impacts of grading, findings from their review suggested the opposite effect. This discrepancy may be due to the prevalent use of the formative approach in studies that incorporated summative peer assessment, which complicates the separation of their individual effects.

It is noteworthy that among the 449 studies reviewed by Alqassab et al. (2023) from 2011 to 2018, none of the 22 studies identified as having both formative and summative purposes included a robust process for supporting any form of peer assessment. This is evidenced by the fact that none of the studies included training students, providing assessment criteria, or an iterative, bidirectional approach. Similarly, we are unaware of any study with these characteristics in the years not covered by Alqassab et al. (2023).

3.2. *Student preference for anonymous or non-anonymous peer assessment*

We identified a few studies in which participants were asked about their preference for anonymous or non-anonymous peer assessment. Curiously, almost all of the studies were conducted in higher education contexts with English as a Foreign Language (EFL) learners in Asia. While this limits the generalisability of their findings, these studies still offer important insights. For example, students' comfort level and ability to provide constructive criticism appeared to be influenced by the modality of peer assessment but for varying reasons. For example, some tended to favour the usefulness of non-anonymous peer assessment in face-to-face interactions (e.g. Kim, 2019, 2023a) with others preferring the psychological safety of anonymity (Coomber & Silver, 2010; Kaya, 2021). The emotional comfort provided by anonymity, as observed by Su (2023), suggests that anonymous peer review can serve as a 'buffer zone', allowing students to express more honest and critical feedback without the apprehension associated with face-to-face interactions. Su's (2023) study, involving 40 Chinese university students over an eight-week period, found that anonymity significantly increased students' perceived comfort levels, particularly at the peer and class levels. This complexity underscores the need to consider individual differences and context-specific variables.

The preference for either anonymous or non-anonymous peer assessment is not uniform and is influenced by factors such as language proficiency and the nature of feedback. Studies by Kim, (2019, 2023a) and Kim and Lan (2021) highlight that students with lower language proficiency levels tend to prefer anonymous peer review, likely due to reduced pressure and fear of direct confrontation. This finding aligns with Allen and Mills (2016), who note that limited proficiency learners are often concerned about the impact of their skills on the effectiveness of peer review. On the other hand, students who are more confident in their language abilities may lean towards face-to-face modality, appreciating the immediate, clarifying dialogues that it facilitates (e.g., Schillings et al., 2021).

4. Framing our study within a feedback model: MISCA

While the impact of the purpose (formative, summative, or both) and the modality (anonymous or non-anonymous) approaches to peer assessment requires further examination, this can only be effectively addressed by considering other elements of the peer assessment process. In this study, we adopted the MISCA feedback model proposed by Panadero and Lipnevich (2022), which encompasses five elements to be considered: Message, Implementation, Student, Context, and Agents (MISCA). This model particularly highlights the implementation aspect and the impact of student attributes on feedback processing. First, as Panadero and Lipnevich (2022) emphasize, *Student characteristics* are pivotal, given that individual differences (e.g., Lipnevich et al., 2016) influence how feedback is processed. These characteristics, therefore, uniquely affect interactions among all model elements. Accordingly, we explored self-regulation skills, content self-efficacy and prior academic achievement in our research as prior research has shown these factors significantly influence feedback processes and outcomes (e.g., Zimmerman, 2000).

While the feedback *Message* was not our primary focus, we trained the students in providing peer assessment using a guided approach (details in Supplementary Material A). Concerning *Implementation*, the primary goals of our peer assessment intervention were to enhance student learning and achievement (Black & Wiliam, 1998; Sadler, 2009), with emphasis on cognitive, self-regulatory, and affective processes. Our instructional *Context* involved designing a robust intervention, informed by prior research, that enabled students to utilise and learn from the feedback in a formative, iterative process. We also considered the classroom climate (e.g., Trzesniewski et al., 2021) in our study, as a key component of the formative context. Lastly, the role of feedback *Agents* in our study was emphasized by integrating peer assessment, student training, and assessment tools, thereby actively involving students in the assessment process and criteria.

5. Aim and research questions

While research has highlighted the cognitive and social-affective outcomes of peer assessment in higher education, the specific influence of anonymity remains unclear. Existing studies offer mixed insights into how anonymity affects student perceptions and preferences, indicating a critical gap in understanding its full potential and nuanced effects. This study addresses this gap by uniquely integrating both formative and summative peer assessment approaches. Appropriate scaffolds for effective peer assessment such as student training, criteria provision, and co-creation were also included. All of this occurred in an iterative, bidirectional process focusing on anonymity.

Our study aims to address these gaps by specifically examining the social-affective variables, perceptions, and preferences of students involving anonymous and non-anonymous approaches to peer assessments that incorporate formative and summative elements. Peer assessment was conducted in pairs on tasks performed by other student pairs. We explored three research questions (RQs):

RQ1. : Do interpersonal and social variables differ between anonymous and non-anonymous conditions when performing formative and summative peer assessments?

RQ2. : What are students' perceptions of social variables in anonymous and non-anonymous formative and summative peer assessment?

RQ3. : What are the predictors for anonymity preference in formative and summative peer assessment?

6. Method

6.1. Research design

This study employs a quasi-experimental design examining the use of formative and summative peer assessment approaches that varied by anonymity status. The control group engaged with non-anonymous peer assessments. The other condition involved pairs of peers assessing their colleagues' work anonymously. Self-regulation, self-efficacy, and grade point average (GPA) were measured prior to the intervention. Social variables, student preferences on peer assessment modality (anonymous vs. non-anonymous), and qualitative perceptions of students were measured after the intervention. This approach did not differentiate between the roles of giving or receiving feedback; but rather focused on the students' overall experiences and perceptions of the formative and summative peer assessment process.

6.2. Participants

177 second-year higher education students from two intact class groups with the same teachers (second and third authors) were involved in this study. One teacher for the practical part and one for the theoretical part, with full coordination and coherence between them. At the start of the program, students were randomly assigned to one of these two groups. Each of these sections was randomly associated with one of the two conditions of this study: anonymity (experimental group) vs. non-anonymity (control group). For tasks where peer assessment was implemented, students worked in pairs. The sample consisted of 39 pairs in the anonymous condition and 29 pairs in the non-anonymous condition. Those students who formed trios (there were one trios in total due to practical reasons in teaching and specific situations) or who did not complete the relevant questionnaires or activities were excluded from the analyses. Therefore, 93 participants were in the anonymous condition (52.5 %) and 84 students in the non-anonymous condition. The majority of participants were male (76.8 %). The students were enrolled in a course about teaching physical activity and sport in the second year of the Degree in Physical Activity and Sports Science.

The research was approved by the Ethics Committee of the first author's University (Code 13/2023). Participation in the study was voluntary, although it was integrated into the instructional design of the course. That is, considering that the implementation of formative and summative peer assessment is a beneficial teaching practice, all students participated in it. The only difference was the voluntary participation in the research with the completion of the questionnaires and the use of student data. In the second session, prior to data collection through the questionnaires, the students were informed about the research and given the information sheet and the informed consent. All the students accepted.

6.3. Materials

6.3.1. Dependent variables assessment instruments

Interpersonal and social variables. The questionnaire by Vanderhoven et al. (2015) was used since it was considered the most complete tool used in the reviewed literature. We selected all the scales used in their study except the first, 'anonymity', as it was very specific to their study. Therefore, five scales were used in relation to how students experienced peer assessment: perceived rating fairness (3 items, reliability index α =.52), positive attitudes (4 items, α =.81), experienced peer pressure (4 items, α =.75), experienced fear of disapproval (4 items, α =.95), feelings of discomfort when evaluated (3 items, α =.54). In this last scale, the second item was not used as it was entirely specific to the research context of Vanderhoven et al. (2015). Consequently, the questionnaire consisted of 18 items, with slight adaptations to the context in some items compared to Vanderhoven et al. (2015). All were answered using a five-point Likert scale ("totally disagree"- "totally agree"). Items were recoded as necessary. In relation to the *Perceived Rating Fairness* scale, a higher score on this scale indicates that the peer assessment was considered by the participant to be honest/correct/objective. On the *Positive Attitudes* scale, high scores suggested a positive attitude towards peer assessment. Higher scores on the *Experienced Peer Pressure* suggested that participants felt low levels of pressure or inhibition and felt that they had good levels of independence and freedom. For the *Experienced Fear of Disapproval* scale, a higher score indicates less fear of peer disapproval. The final scale was *Feelings of Discomfort when Evaluated*. A higher score on this scale indicates greater comfort in assessing and being assessed by peers.

Peer assessment modality (anonymous or non-anonymous) preference. We asked students, after the intervention, about their preference for one modality or another, regardless of whether students had experienced peer assessment anonymously or non-anonymously.

Qualitative students' perceptions of social variables in anonymous and non-anonymous peer assessment. Additionally, a qualitative questionnaire with five open questions was used to explore (1) how students felt when reading the peer feedback received, (2) how they felt at the moment they knew their peers were reading the feedback they provided, (3) how the received and (4) how the provided peer feedback affected their learning and tasks, and, finally, (5) whether they wished to add anything else of interest that they had not previously commented on.

6.3.2. Instruments for measuring control variables

6.3.2.1. Self-regulation of learning. In order to reach an appropriate estimation of self-regulation, two different instruments were used following prior suggestions (Boekaerts & Corno, 2005; Samuelstuen & Bråten, 2007):

First, the 'Emotion and Motivation Self-Regulation Questionnaire (EMSR-Q)' (Alonso-Tapia et al., 2014). Participants rate their agreement to 20 items on this questionnaire using a five-point Likert scale ("strongly disagree" to "strongly agree"). This instrument is structured based on five first-order scales. It also has two second-order factors: (1) learning self- regulation style, with 12 items and a reliability index (Cronbach's α) of.76; and (2) avoidance self-regulation style, with 12 items and a reliability of α =.88 (the first-order scale 'Performance oriented self-regulation' computes on both second-order scales). The second-order scale 'learning self- regulation style' includes self-messages or mental verbalizations that affect students' motivation, goals and learning. The higher the value on this scale, the greater the positive effect of emotional and motivational strategies on student learning. The second-order scale 'avoidance self-regulation style' includes self-messages and actions that show a lack of regulation or are geared toward task avoidance. The greater the value on this scale, the greater the negative effect on learning of the emotional strategies and motivations implemented by the student.

Second, the 'Deep Learning Self-Regulation Strategies (DSL-Q)' (Panadero et al., 2021). It is an evaluation of the application of regulatory and learning strategies. It is designed around common, situational, and context-specific learning scenarios. It involves respondents rating their level of agreement to 30 statements using a five-point Likert scale. This tool groups four first-order scales around which the items are arranged: (1) Basic learning self-regulation strategies (8 items, reliability index (Cronbach's α) of.75), (2) Visual elaboration and summarizing strategies (8 items, α =.84), (3) Deep information processing strategies (8 items, α =.80), and (4) Social learning self-regulation strategies (6 items, α =.66).

6.3.2.2. Self-efficacy questionnaire. Created *ad hoc* to measure students' self-efficacy in providing feedback, partly developed from previous research. It includes 9 items with a five-point Likert scale ("strongly disagree"–"strongly agree"). Examples of items are "I believe I am capable of giving good feedback to a colleague" and "When I give

feedback to a colleague, I believe that the person I am giving it to does not influence me." Self-efficacy was measured before the intervention. The reliability index was α =.40.

6.3.2.3. Students' first year grade point average (GPA). GPA is the average of a student's grades over a period of time. We accessed participants' first-year university records to extract this information, after obtaining ethics committee approval.

6.3.2.4. Academic performance. Students, working in pairs, developed a didactic unit (DU) related to some area of physical activity and sport, which accounted for 40 % of the final grade of the subject. This task involved planning learning activities in relation to the other elements of the teaching program. For grading, teachers provided an assessment/grading tool (Supplementary Material D) with 10 sections and quality assessment criteria. The two teachers divided the task of grading the DUs. Inter-rater reliability (Cronbach's α) was used between the two subject teachers. For the calculation, both teachers graded the same 20 % of DUs. Thus, for each of the 10 criteria, a minimum of α =.62 (criterion 1) to a maximum of α = 1.00 (criterion 9) was obtained, with 80 % of the criteria α >.80 and 70 % of the criteria α >.90. In total, for the overall numerical grade obtained from the 10 items, the resulting agreement level between both teachers was α =.95.

6.3.2.5. Accuracy in peer grading. To explore accuracy, the DU development task was also used. Each pair of students, jointly and consensually, randomly graded (organized by the teachers) the DU of another pair (anonymously in the experimental group and non-anonymously in the control group) using the DU assessment and grading tool (Supplementary Material D). This task accounted for 10 % of the final grade of the subject. This is because, as Panadero and Alqassab (2019) pointed out, it is important to contemplate the implications of ensuring assessors are responsible for the accuracy of their summative peer assessments. Without consequences for inaccuracies, such as scoring too generously, an assessor might be inclined to give higher scores, seeing no benefit in being precise at the risk of appearing strict to the individual being assessed.

Accuracy was calculated by comparing the grade of the student pair with the grade given by the teacher on that same DU of another pair. The degree of accuracy, and the corresponding grade, was awarded according to the table in Supplementary Material E. Given this paper's research questions, this data will be examined at a later date.

6.3.3. Assessment instruments used in the intervention

As detailed in the *Procedure* section, teachers initially provided two checklists that established the goals for two of the three parts of the didactic unit (DU). Subsequently, the final DU assessment instrument (Supplementary Material D) was co-created with the students, which also included these parts, but in greater depth and detail. This final instrument was used by the teachers to grade the students' UDs, and the data from this grading was used in this study.

6.4. Data analysis

This study sought to explore higher education students' perceptions of peer assessment when anonymity is considered. Descriptive analysis was first conducted to provide a comprehensive picture of the sample's characteristics in relation to the key variables under examination. Inferential statistics were calculated to the data to examine the relationship between the variables with relevant outcomes variables. Where appropriate, bias-corrected and accelerated (BCa) bootstrapped confidence intervals set to 95 % (BCa 95 % CI) were applied to reduce bias. A *p*-value of <.05 was the threshold for statistical significance used in this study, unless specific p-values are reported. Bonferroni corrections were applied where necessary. It should be noted that not all participants provided relevant data; therefore, the *n* for the results presented will vary.

For the open-ended qualitative data, coding of the data into deductive and inductive categories was carried out by the first and second author. They first prepared a set of categories based on the review by Panadero and Alqassab (2019). They then independently coded the qualitative responses of the first 10 participants in the database. After pooling them, they made several decisions about the coding and the inductive categories that emerged. They then co-coded the responses of a further 5 participants, and finally the remainder of the coding was done by the second author. Upon completion of the analysis, which involved identifying 414 meaning units, the two authors systematically categorized these units into 95 detailed subcategories. Recognising the need to meaningfully interpret the large dataset, these were further grouped into 18 broader subcategories — 10 reflecting student-perceived positive aspects and 8 indicating perceived negative aspects of the intervention.

6.5. Procedure

The study was conducted in a course related to didactics in teaching processes of the 2nd year of the Degree in Physical Activity and Sports Sciences at a Spanish university. One of the course objectives was to develop a didactic unit (DU) in pairs. The students organised themselves in pairs of their choice, as the course teachers aimed to maximise learning through collaborative interaction. The course was taught by two teachers who collaborated closely throughout the process. One handled the theoretical sessions and the other the practical sessions. Although the DUs were developed primarily during the practical sessions, they drew extensively on the knowledge imparted in the theoretical sessions. The DUs, which integrated both theoretical insights and practical applications, were graded by both teachers.

Students were randomly divided into four groups upon entering the university. Two of these groups were randomly assigned to the anonymous condition and two to the non-anonymous condition. Students completed self-regulation and self-efficacy questionnaires at the beginning of the course.

Bidirectional anonymity was ensured by establishing a code for each pair that remained consistent throughout all activities. In this way, teachers made pairings for each activity ensuring that each time there were different feedback pairs. Thus, students could receive observations from other peers and determine their relevance, minimizing the risk if feedback always came from the same pair and was not appropriate (van den Berg, 2006).

The course sessions, focused on educational methods and their application in the development of a DU, were structured in a sequence of three phases with different activities: initial, development, and synthesis.

6.5.1. Initial Phase

The objective of this phase was to activate students' prior knowledge about how to generate effective feedback (e.g., Hattie & Timperley, 2007; Henderson, et al., 2019) and to clarify the role of criteria to do so. Exemplars with DUs of varying quality were used, asking students to generate assessment criteria and identify what constitutes quality. Additionally, students received feedback from the teachers on simple initial assessment tasks. All of this was discussed with the students for their learning.

6.5.2. Development phase – formative peer assessment

The main task of this research, as previously mentioned, was the development of didactic unit (DU) in pairs. This was broken down into three sections: (1) objectives of the DU (accounting for 30 % of the DU grade), (2) assessment criteria for the activities outlined in the DU (35 %), and (3) other sections (35 %). From the beginning, the teachers provided a control list for the objectives (Supplementary Material B) and

another for the assessment criteria (Supplementary Material C). The instructional approach to these two sections was structured as follows: firstly, the teacher provided an explanation; secondly, the pairs developed the section; thirdly, pairs provided feedback on their peers' work, either anonymously (experimental group) or non-anonymously (control group). Fourth, students analysed the feedback received and refined their work accordingly.

6.5.3. Synthesis phase – summative peer assessment

Teachers and students co-constructed the final assessment instrument, intended to assess/grade the DU (Supplementary Material D). This tool, based on the assessment criteria and guidance from the teachers, was employed by the pairs for self-assessment. Over a period of three weeks, students used it to refine and enhance their DU projects. After submission, two actions were implemented. First, with the same instrument, student pairs graded another pair's DU (accounting for 10 % of their final course grade). Second, an additional 10 % of their course grade was determined by how accurately their grading of another pair's DU matched the teacher's grade for the same DU, as outlined in Supplementary Material E.

At this final point, students completed the questionnaire on social variables perception and self-efficacy.

Finally, we want to note that we used the instrument to report the characteristics of peer assessment designs (Panadero et al., 2023. See Supplementary Material F) to provide a detailed context of the intervention and facilitate the replicability of the study.

7. Results

To test whether the conditions were comparable in terms of their performance within the programme, we compared students' first-year grade point average (GPA) according to the condition they were in. While there were some differences noted in GPAs between the anonymous (n=91; M=1.67, SD=0.33) and non-anonymous (n=83; M=1.75, SD=0.39) conditions, the mean difference −0.090, BCa 95 % CI [-0.21, 0.03] was not significant; t(160.5)=-1.636, p=.113.

7.1. RQ1: Do interpersonal and social variables differ between anonymous and non-anonymous conditions when performing formative and summative peer assessments?

Participants' perceptions of interpersonal and social variables were gathered using an adapted version of Vanderhoven et al.'s (2015) questionnaire and involved five scales. Note that, as explained in detail in the *Methods* section, this questionnaire had a mixture of positive and negative item wordings on each scale. Consequently, some items were reversed, with a focus on higher scores indicating more positive feelings. Thus, a higher score on *Perceived Fairness* indicates that the peer assessment was more honest/correct/objective; on *Positive Attitude* a more positive disposition; on *Experienced Peer Pressure* less pressure or inhibition, i.e. more independence and freedom; on *Experienced Fear of Disapproval* less worry or fear of peer disapproval; and on *Feelings of Discomfort when Evaluated* more comfort in assessing and being assessed by peers. To determine if students' perceptions of interpersonal and social variables significantly varied according to condition, a series of bootstrapped independent samples t-tests were conducted with a Bonferroni correction applied to reduce the risk of Type I errors. The significance level was consequently set at p=.01. Table 1 summarises these test statistics.

As shown in Table 1, differences in the social and interpersonal variables experienced by participants in the anonymous and non-anonymous conditions were not significant. However, it is interesting to note that the effect size representing the difference in participants' reported positive attitudes between conditions was d=0.27. This is a relatively robust effect size given the size of the sample and suggests that those in the non-anonymous condition had a more positive attitude to peer assessment.

Additionally, participants' *Perceived Fairness, Positive Attitude, Experienced Peer Pressure*, and *Experienced Fear of Disapproval* were examined as dependent variables using a one-way between-groups analysis of covariance where the experimental condition was the type of grading experienced (anonymous vs. non-anonymous). The independent variable was condition (anonymous, non-anonymous). Participants' scores on the researcher-designed measure of self-efficacy, the DSL-Q (Panadero et al., 2021) as well as performance on the two second-order scales from the EMSR-Q (Alonso-Tapia et al., 2014) were used as covariates. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, etc. *Feelings of Discomfort when Evaluated* was not examined as this variable violated assumptions related to the homogeneity of regression slopes with one covariate (DSL-Q; Panadero et al., 2021).

Concerning *Perceived Fairness*, there was no significant difference between the two groups (Anonymous: n=56; M=12.00, SD=2.44; Non-Anonymous: n=53, M=11.72, SD=2.13); F(1, 103) =0.780, p =0.379, *partial eta squared* =.01). With regard to *Positive Attitude*, there was no significant difference between the two groups in their reports (Anonymous: n=56; M=14.36, SD=3.24; Non-Anonymous: n=53, M=15.40, SD=3.52); F(1, 103) =1.877, p =0.174, *partial eta squared* =.02). Relating to *Experienced Peer Pressure*, there was no significant difference between the two groups (Anonymous: n=56; M=16.88, SD=3.04; Non-Anonymous: n=53, M=17.04, SD=2.97); F(1, 103) =0.030, p =0.863, *partial eta squared* =.00). Finally, with reference to *Experienced Fear of Disapproval*, there was no significant difference between the two groups (Anonymous: n=56; M=17.36, SD=3.41; Non-Anonymous: n=53, M=17.77, SD=3.18); F(1, 103) =1.729, p =0.697, *partial eta squared* =.00). None of the covariates included in the model were significantly related to these four variables. These results have been summarised in Table 2.

7.2. RQ2: What are students' perceptions of social variables in anonymous and non-anonymous formative and summative peer assessment?

Table 3 shows the frequencies of the grouped categories derived from the open-ended questions. Of the 414 coded units of meaning, 79.5 %

Table 1
Interpersonal and Social Variables by Condition.

	Anonymous (n=64)		Non-Anonymous (n=63)		t	BCa 95 % CI**	p*	d
	M	SD	M	SD				
Perceived Fairness	12.05	2.4	11.70	2.2	0.846	−0.47, 1.14	.40	0.15
Positive Attitude	14.36	3.4	15.29	3.5	−1.500	−2.12, 0.27	.14	0.27
Experienced Peer Pressure	16.83	3.4	17.10	2.9	−0.477	−1.45, 0.92	.63	0.09
Experienced Fear of Disapproval	17.22	3.8	17.51	3.2	−0.465	−1.54, 0.88	.64	0.08
Feelings of Discomfort when Evaluated	11.50	2.2	11.73	2.2	−0.562	−1.00, 0.54	.58	0.10

* where p is two-sided
** Bias Corrected and Accelerated 95 % Confidence Interval

Table 2
Summary Table of ANCOVA results by Dependent Variable.

	Anonymous (n=56)	Non-Anonymous (n=53)	*df*	*F*	*p*	*Partial eta squared*
	M(SD)	*M(SD)*				
Perceived Fairness	12.00(2.4)	11.72(2.1)	1, 103	0.780	0.379	.01
Positive Attitude	14.36(3.2)	15.40(3.5)	1, 103	1.877	0.174	.02
Experienced Peer Pressure	16.88(3.0)	17.04(3.0)	1, 103	0.030	0.863	.00
Experienced Fear of Disapproval	17.36(3.4)	17.77(3.2)	1, 103	1.729	0.697	.00

were positive and 20.5 % negative. All percentages presented are calculated based on the total number of meaning units. Among the positive ones, and with similar values between the two conditions, the most important were the usefulness of the peer assessment for their learning (22 %) and the feeling it generated (19 %), as well as helping their peers (15 %) and the calmness related to the feeling that the peer assessment was done objectively and fairly due to the assessment tools and criteria (11 %). On the negative side, there was a general feeling of nervousness and discomfort (7 %), disagreement with the grade received by their peers (5 %) and discomfort at the thought of how their peers would have received the feedback provided (4 %) and at having to grade them (2 %). In terms of specific perceptions of the peer assessment modality, some students highlighted the usefulness of communication between assessor and assessee (1.2 %), as well as a preference for the non-anonymous modality to be more honest and not self-conscious (0.7 %), or for there to be no agreements regarding peer gradings (0.5 %).

To more effectively illustrate the above findings, the most representative responses to the open-ended questions on students' perceptions of peer assessment from the anonymous (A) and non-anonymous (N-A) peer assessment conditions are presented below, grouped into three directions.

First, positive comments regarding the didactic approach and sequence of the peer assessment were relatively consistent in this data: "I find it very positive because it highlights things that you have missed, perhaps unconsciously, helping you to improve and complement your work". #8N-A). Participants also praised the assessment instruments and criteria: "the feedback was always in line with the criteria to be assessed, so there was no subjectivity" (#4N-A). Positive feelings of usefulness and feeling valued were also recorded: "I have felt useful because I am influencing the learning of my peers" (#48 A) or "my ultimate goal was to help my peers to improve. 'If they improve, I improve'" (#8N-A).

Students also commented on the quality of the feedback. On the one hand, the questioning of students' ability to give feedback to their peers and, consequently, the quality of the feedback: "I think feedback is always a help, although I think it is merely a help because nobody masters the subject and they are really things that we think, but we don't master" (#116N-A). Along these lines, some students also doubt their own ability to give feedback or that of their peers ("sometimes you don't know if their feedback is good or not". #6N-A) or the feelings that this may generate ("I have felt the doubt that I may have offended someone". #77 A). However, on the other hand, and as shown in the content analysis, several students noted the usefulness of the feedback received ("it's nice to see your peers giving you advice to improve, it makes you more confident with them" (#25N-A), "It has helped me to understand things that were not entirely clear to me and helped me to ask myself questions that I would not have realised otherwise" (#45 A) or "I have realised our mistakes because the others had them too" (#22N-A)). Students also acknowledged the efficiency of peer assessment: "if you get feedback from a peer, as long as they are clear what the criteria are, this I think will be more comprehensive than if one teacher has to give feedback to all their students. […] Each student is given more individualised and therefore better feedback" (#101N-A). They also point out that "sometimes we feel attacked by teacher feedback, but among peers we are not so defensive" (#138 A).

Finally, students' preferred modality of peer assessment was a frequent idea in the open-ended responses. In line with the quantitative results of RQ3, participant #138 A asserted that: "Being anonymous you can observe the feedback much more clearly as such, and not think about why you have better or worse feedback, and focus on external aspects such as the relationship you may have with the peers giving you the feedback. You just focus on what you can get out of it". This seems especially important in the context of peer grading, provoking negative feelings related to the grade received ("I don't think the grade I received on my work is fair". #93 A) or not wanting to give low grades to peers ("I don't like to give bad grades to any of my peers". #91 A). However, having the opportunity to comment on the feedback received and provided is also identified as an excellent learning opportunity: "I had a dialogue with them to see what went wrong and I understood it" (#75N-A). This non-anonymous setting allowed for direct communication, which is not possible in anonymous peer assessments.

7.3. RQ3: What are the predictors for anonymity preference in formative and summative peer assessment?

Out of the 125 participants who reported which peer assessment modality they preferred, 92 (73.6 %) reported that they preferred the anonymous condition while 33 (26.4 %) opted for non-anonymous. Importantly, these preference percentages varied depending on the condition. In the anonymous condition, 90.5 % of the participants indicated a preference for anonymous peer assessment and 9.5 % for non-anonymous peer assessment. In the non-anonymous condition, 56.5 % of students preferred anonymity and 43.5 % preferred non-anonymity.

A direct logistic regression was performed to assess the impact of a set of predictor variables on the odds that respondents would prefer anonymous (n=92) or non-anonymous (n=33) approaches to peer assessment. The model contained 5 independent variables involving the interpersonal and social variables that summarised how students experienced peer assessment: *Perceived Rating Fairness, Positive Attitudes, Experienced Peer Pressure, Experienced Fear of Disapproval, Feelings of Discomfort* when evaluated. The full model containing all of these predictors was statistically significant, χ^2 (5, 125)=16.171, p=.006, indicating that the model was able to distinguish between participants who reported a preference for anonymous and non-anonymous peer assessment. It correctly predicted the preference of 75.2 % of participants including 77 % of those who preferred anonymous forms of peer assessment and 58.3 % of those who preferred non-anonymous forms. However, while the model was able to correctly classify many cases, it must also be acknowledged that the model was only able to explain 17.7 % (Naagelkerke R^2) of the variance in preference. As illustrated by Table 4, only one of the independent variables made a uniquely statistically significant contribution to the model: Positive Attitudes. Having a positive attitude was the strongest predictor of reporting a preference for the non-anonymous condition, recording an odds ratio of 1.29. This suggests that the odds are 1.29 times greater that respondents who had a positive attitude would report a preference for non-anonymous approaches to peer assessment.

Direct logistic regression was performed to assess the impact of a set of different set of predictor variables on the odds that respondents would prefer anonymous or non-anonymous approaches to peer assessment.

Table 3
Detailed Exploration of the Two Categories and the 18 Subcategories and Frequencies.

	Anonymous		Non-anonymous		Total	
	n	%	*n*	%	*n*	%
Positive	152	36.7	177	42.8	329	79.5
Useful for learning (knowing where I am failing, how I can improve, reviewing content and assessment criteria to give feedback to my peers, seeing examples of higher quality, etc.)	44	10.6	48	11.6	92	22.2
Feeling good (for various reasons: satisfaction, gratitude, honesty, learning, peace of mind when comparing one's work with that of peers, etc.)	37	8.9	41	9.9	78	18.8
Helping my colleagues (improving their work, feeling useful, self-motivation to improve one's own work and learn, highlighting the positive, etc.)	28	6.8	33	8.0	61	14.7
Fairness thanks to assessment tools and criteria (not being guided by friendships, not thinking about which colleague did the task, being able to focus on positive and negative aspects, more rigorous, constructive criticism, etc.)	20	4.8	26	6.3	46	11.1
Accepting criticism (different points of view, it is right to point out my mistakes as the aim is to learn, try to understand that it is an opportunity to improve despite disagreements).	8	1.9	9	2.2	17	4.1
Put yourself in the teacher's place, learn to give feedback, work on assertiveness in being critical	4	1.0	9	2.2	13	3.1
Joy and consideration of the deservedness of the grade received	4	1.0	5	1.2	9	2.2
Valuing peers (trusting them, the teacher with so many students cannot give the high-quality feedback that can be obtained from peers, good classroom climate)	4	1.0	2	0.5	6	1.4
Importance of communication between assessor and assessee (understanding the feedback received, leading to improvement for both, more help, motivation, classroom climate, teamwork, opportunity to change opinions, correcting mistakes in basic concepts)	3	0.7	2	0.5	5	1.2
Joy of giving an excellent grade to peers	0	0.0	2	0.5	2	0.5
Negative	49	11.8	36	8.7	85	20.5
Negative feelings in general (discomfort, nervousness, etc.)	16	3.9	12	2.9	28	6.8
Disagreement with peer grading (unfairness, expectation of higher grades, etc.)	10	2.4	9	2.2	19	4.6
Uncomfortable and nervous about not knowing how peers receive feedback and doubting one's own ability to give feedback	13	3.1	3	0.7	16	3.9
Discomfort in grading peers (not being fair, worrying about offending peers, not liking giving a low grade, knowing who gets a low grade)	4	1.0	5	1.2	9	2.2
Doubt about the quality of feedback received (whether or not peer is right or wrong, confusion)	2	0.5	4	1.0	6	1.4
Preference for non-anonymous modality to avoid being inhibited by the feedback or grade given	3	0.7	0	0.0	3	0.7
Unfairness detected by peer grading pacts between students in the non-anonymised group	0	0.0	2	0.5	2	0.5
Doubt about the quality of feedback given to other students, sense of responsibility	1	0.2	1	0.2	2	0.5

Table 4
Logistic Regression Predicting Likelihood of Reporting a Preference for Non-Anonymous Approaches to Peer Assessment according to Interpersonal and Social Factors.

	B(SE)	Odds Ratio	*95 % CI for Odds Ratio*	
			Lower	Upper
Perceived Fairness Rating	−0.05 (0.11)	0.95	0.78	1.17
Positive Attitudes	0.26 (0.10) *	1.29	1.08	1.54
Experienced Peer Pressure	0.05 (0.09)	1.05	0.88	1.25
Experienced Fear of Disapproval	−0.05 (0.08)	0.95	0.81	1.12
Experienced Feelings of Discomfort	0.06 (0.13)	1.07	0.82	1.38

NOTE: $R^2 = 0.121$ (Cox-Snell), 0.177 (Nagelkerke).
* p<.05

The model contained six independent variables involving other *non*-interpersonal variables that have previously been shown to influence how students experience peer assessment. It should be noted that not all participants provided data for each predictor variable. Therefore, the sample sizes between participants who reported a preference for anonymous (n=48) and non-anonymous (n=35) peer assessment was adjusted for this analysis. The full model containing all of these predictors was statistically significant, χ^2 (6, 83)=32.540, p=<.001. The model correctly classified 72.3 % of cases with the explained variation in the dependent variable calculated at 46.8 % (Nagelkerke R^2). Of all cases that were predicted to report a preference for non-anonymous forms of peer assessment, 80.0 % were correctly predicted. For those cases who were predicted to prefer anonymous forms of peer assessment, the model correctly identified 88.9 %.

As illustrated by Table 5, four of the independent variables made a uniquely statistically significant contribution to the model: Condition, Final Grade, DU Grade and EMSR-Q Avoidance Self-Regulation Style.

Table 5
Logistic Regression Predicting Likelihood of Reporting a Preference for Non-Anonymous Approaches to Peer Assessment according to Intrapersonal Factors.

	B(SE)	Odds Ratio	*95 % CI for Odds Ratio*	
			Lower	Upper
Condition	2.45 (0.68)*	11.69	3.09	44.21
Final Grade	−0.99 (0.40)*	0.37	0.17	0.82
DU Grade	0.69 (0.30)*	1.99	1.10	3.60
Self-Efficacy	0.10 (0.11)	1.11	0.90	1.36
EMSR-Q: Learning Self-Regulation Style	0.17 (0.08)*	1.18	1.01	1.38
EMSR-Q: Avoidance Self-Regulation Style	−0.04 (0.05)	0.97	0.88	1.06

NOTE: $R^2 = 0.324$ (Cox-Snell), 0.468 (Nagelkerke).
* p<.05

The DSL-Q self-regulation questionnaire score was not considered in this model because, from a theoretical perspective, it focuses more on the general self-reported actions of self-regulated learners than on their experiences, emotions, motivations, etc., which have a stronger influence on their preferences. Having experienced the non-anonymous condition in this study was the strongest predictor of reporting a preference for the non-anonymous condition, recording an odds ratio of 11.69. A higher performance on the DU grade meant that students were more likely to report a preference for the non-anonymous condition. In contrast, a higher score on their Final Grade made participants less likely to report a preference for non-anonymous peer assessment. Students who scored high on the EMSR-Q Avoidance Self-Regulation Style scale also appeared less likely to prefer non-anonymous conditions. However, given that the confidence interval of the odds ratio contains 1 (0.88, 1.06), this interpretation is somewhat unreliable as the opposite relationship may also be true.

8. Discussion

The purpose of this study was to investigate the effect of anonymous and non-anonymous approaches to peer assessment on higher education students' social-affective experiences, perceptions, and preferences regarding anonymity. Using a quasi-experimental approach, formative and summative approaches to peer assessment were adopted with 177 students. Approximately half of these students engaged in anonymous peer assessment (n=64) and the other half in non-anonymous peer assessment (n=63).

8.1. Social-affective and interpersonal outcomes

Our first two research questions (RQ1 and RQ2) explored the social-affective dimensions of formative and summative peer assessment. Here, we explored whether self-reported interpersonal and social variables differed between anonymous and non-anonymous conditions when performing formative and summative peer assessment. RQ1 examined these variables quantitatively while RQ2 adopted a qualitative approach. In this section, we will consider the results of RQ1, while simultaneously integrating and complementing these findings with the qualitative insights derived from the data gathered for RQ2.

Our findings for RQ1, gathered through a comprehensive questionnaire adapted from Vanderhoven et al. (2015), focused on multiple dimensions: Perceived Fairness, Positive Attitude, Experienced Peer Pressure, Experienced Fear of Disapproval, and Feelings of Discomfort when Evaluated. Although there were no statistically significant differences between the anonymous and non-anonymous groups in these dimensions, it is worth noting a modest effect size (d=0.27) in the Positive Attitude domain was noted. This magnitude of differences between means suggests a slightly more favourable attitude towards peer assessment in the non-anonymous condition. While it cannot be stated with certainty that this was a true effect given its non-significance, it is one variable that may be worth further examination in future research.

Interpretation of these findings in light of the current literature in this field is not straightforward. For example, these findings do not align with Panadero and Alqassab's (2019) review, which noted that anonymity could positively influence learners' perceptions, particularly in terms of critical feedback and self-perceived social effects. However, the design of the current study may have influenced this 'misalignment' with the broader field of literature. Panadero (2016) posited that more formative interventions might counteract the effects of anonymity, as achieving peer assessment's full potential requires high levels of interaction between the assessor and assessee. Our study adopted an atypical approach to peer assessment research where both formative and summative approaches were considered. This currently represents a very small percentage of research in the field (see Alqassab et al., 2023) thus making it difficult to fully understand whether this study affirms or contests what is known about anonymity and peer assessment. Furthermore, given that our intervention included summative peer assessment, and considering the additional challenges posed by this implementation (Panadero, 2016), it could be hypothesised that this would have influenced the results. Therefore, our study's lack of significant differences challenges some common assumptions about the impact of anonymity on peer assessment.

Indeed, in our highly formative assessment context — which included training for students on how to provide effective peer feedback, the use of assessment tools, co-creation of assessment criteria between teachers and students, enhanced self- and co-regulation processes, provision and critique of exemplars, and experience in both formative and summative peer assessments our participants preferred anonymous peer assessment. This preference likely stems from the heightened stakes and potential discomfort associated with summative assessments (Panadero, 2016). Su (2023) noted that, where grades or formal evaluations are involved, the dynamics of peer interactions can be negatively altered and introduce elements of competitiveness or apprehension. Our qualitative results indicated that students perceived greater fairness and honesty when grading their peers anonymously. It should also be that the pedagogically richer approach implemented in the initial stages of this study could have mitigated the negative outcomes typically associated with more traditional implementations of summative peer assessment (see Alqassab & Panadero, 2020). This, combined with the use of anonymous approaches, may have resulted in a preference for anonymity in the data gathered. Consistent with previous studies (e.g., Kim, 2019, 2023a), our participants also manifested the value of face-to-face dialogue for concept clarification and two-way feedback in non-anonymous peer assessments. Therefore, the higher Positive Attitude in non-anonymous condition is consistent with the claim that formative peer assessment interventions, which typically require more interaction, may have better social and interpersonal outcomes (Panadero, 2016).

Additionally, grading peers' work and being graded on the accuracy of their grades, in relation to the teacher's gradings, indeed heightened the sense of responsibility and accountability among students. This aspect was rigorously evaluated by comparing peer-assigned grades with teacher-assigned grades, as recommended by Panadero and Alqassab (2019) and detailed in our method section under 'Accuracy in peer grading.' This approach was implemented to ensure that assessors were held accountable for the accuracy of their summative peer assessments, addressing concerns over potential leniency in scoring without consequences for inaccuracies. This aspect of the intervention may have alleviated the pressure and discomfort students felt not only in assigning grades to their peers but also in receiving them, thereby having a lesser influence on their perceptions towards one modality or the other. Therefore, this reflects the interplay of peer assessment, trust, and psychological safety, as discussed by Senden et al. (2023). These discrepancies underscore the complexity of social dynamics in educational settings, as also observed in studies by Su (2023) and Senden et al. (2023) and pointed out by Panadero and Alqassab (2019). These works highlight that the effects of anonymity are not universally positive or negative, but rather context dependent.

Finally, one aspect to consider is that these students will work in roles requiring public interaction (e.g., coaches or physical trainers) and will also need to assess and grade others (e.g., secondary physical education teachers). Peer assessment and peer grading not only improve academic performance but also develop essential skills for their future professions. Moreover, as university students experience formative assessment, they are more likely to implement it in their future teaching practices (Hamodi et al., 2017; Lorente-Catalán & Kirk, 2016).

8.2. Student preferences for anonymous or non-anonymous peer assessment

In RQ3, 73.6 % of participants (90.5 % in the anonymous condition and 56.5 % in the non-anonymous) expressed a preference for the

anonymous modality. This preference for anonymity resonates with Hosack's (2004) observation that anonymous feedback led Japanese students to worry less about offending the writer, thereby producing more useful feedback. It also aligns with Su's (2023) study in which students clearly preferred the anonymous modality and reported significantly higher levels of comfort on the peer and class-level dimensions. Furthermore, to the best of our knowledge, most studies have reported a preference for anonymous peer assessment (e.g. Kaya, 2021; Kim, 2023a, Ting, 2024). Interestingly, *Positive Attitude* emerged as a strong predictor for preferring non-anonymous peer assessment. This preference aligns with findings from Kim and Lan's (2021) study and Ting (2024), where the option for face-to-face interaction was valued for its ability to facilitate direct verbal communication. Participants in these studies highlighted the benefit of non-anonymous peer assessment, as it allows assessors and assesses to discuss and clarify feedback in person, thereby easing the conveyance and reception of potentially negative comments.

In fact, several studies (e.g., Kaya, 2021; Kim, 2019, 2023a) suggest that the preference for a specific peer assessment modality is influenced by various factors, including cultural backgrounds and L2 proficiency. Our findings illustrate the inclination towards non-anonymous peer assessment among high performers in the DU grade. This grade was assigned during tasks where both formative and summative peer assessments were performed. Such a trend suggests that these high-performing students were confident in engaging in direct, constructive dialogue. This observation aligns with studies by Kim (2023b) and Su (2023), which found that comfort levels and the desire for direct interaction significantly influence modality preference. Furthermore, the preference for anonymous peer assessment among lower proficiency learners, as indicated in our study, echoes the findings of Hyland and Hyland (2006) and Kim (2019). These learners may find anonymity a safer environment, which reduces the anxiety associated with direct criticism and less concern about their perceived low ability to give feedback due to their limited expertise. This preference for anonymity could be attributed to the learners' desire to focus more intently on the content of the feedback, free from interpersonal dynamics, a perspective that aligns with Su's (2023) 'masked ball' analogy.

In conclusion, despite our rich and formative intervention, as highlighted by Li et al. (2016), the challenges associated with peer grading—particularly those noted by Panadero (2016)—could not be fully overcome. This was especially evident in the non-anonymous setting, where the discomfort of grading and being graded by peers had a tangible impact. This challenge led to a preference for the anonymous mode of evaluation. Indeed, students who experienced anonymity clearly showed this preference (90.5 %). However, students in the non-anonymous condition, likely because they experienced the benefits of face-to-face interactions as previously mentioned, did not show as strong a preference for anonymity (56.5 %). Interestingly, experiencing non-anonymous peer assessment seems to shift students' perceptions, with more than half of those in the non-anonymous condition feeling confident and showing an openness towards this modality. Consequently, we suggest that educators who weigh the pros and cons and decide to implement summative peer assessment should consider doing so non-anonymously, particularly within a rich formative process, as it is truly the most beneficial approach.

Finally, it is important to note that in our study students worked in pairs. In several studies, mainly thanks to qualitative results (e.g., Kim, 2019; Kim & Lan, 2021), students stressed the importance of the face-to-face encounter (non-anonymous peer assessment) for numerous valuable reasons that we also found in our results. Thus, the preference for the anonymous modality in our results may be because many of these benefits (e.g., reviewing assessment criteria and content, exploring peers' feedback and views, etc.) were also accessible through the paired nature of the assignments and peer assessments, except for exchanging explanations directly with peers.

8.3. Limitations and Future lines of research

First, although peer assessment was conducted in a formative manner and clear guidelines were provided, peer grading may still have influenced participants' final perceptions across both conditions (as noted by Panadero, 2016). While this could be considered a limitation, this research represents a preliminary step towards integrating formative and summative peer assessment, which have not been directly linked before. Additionally, students' responses reflected the joint formative and summative peer assessment process used in this intervention. Second, an inherent limitation of peer grading is that it is done on the submitted assignment, preventing students from improving their work based on the feedback received. However, summative peer assessment has its own advantages, especially in formatively rich contexts (as concluded by Li et al. (2016) and Sanchez et al. (2017)). One noteworthy advantage of peer grading is that students submit their assignments for grading at a point where they have reached the highest possible quality of their work. This contrasts with formative peer assessments, which occur at an intermediate stage where students may not perceive any risks. Consequently, peer grading can foster deep, responsible, and accountable reflection. This was implemented in our study and reflects notes by Panadero and Alqassab (2019). Therefore, although summative peer assessment does not allow for the resubmission or improvement of the assignments, this process can yield additional benefits, such as enhanced performance on subsequent tests, as demonstrated by Sanchez et al. (2017). Third, some questionnaire scales had low reliability indexes which calls into question their suitability as outcomes measures.

A number of avenues can be proposed for future research. First, as reported in Supplementary Material F, we employed a double-anonymous process in the anonymous condition. There are multiple ways to implement anonymity and it would be important to also explore systematically other options. Second, future studies should implement randomized controlled trials to obtain certainty about the causality of the effects. In other words, as we have stated, the authors are not aware of any other studies that have implemented formative and summative peer evaluation in a formative context, so further studies with and without anonymity remain necessary.

Finally, a potential future line of research, suggested by both the quantitative and qualitative results of this study, could involve a bifurcated approach to peer assessment: formative peer assessment conducted non-anonymously and summative peer assessment conducted anonymously. This approach could harness the distinct benefits of both formative and summative peer assessments. Conducting formative assessments non-anonymously could promote face-to-face dialogue among students, thereby facilitating task improvement through direct communication and feedback. On the other hand, conducting summative assessments anonymously—when students no longer have the opportunity to improve the task—could also be beneficial, as suggested by previous research (e.g., Sanchez et al., 2017). Such an arrangement might mitigate some of the challenges associated with the grading aspect of summative peer assessment, by reducing the personal biases and discomfort that can arise when students are aware of each other's identities during the grading process.

9. Conclusions

This study examined the effectiveness of anonymity in peer assessment using an intervention that incorporated summative and formative approaches to assessment. Our results challenge the conventional wisdom that anonymity unequivocally reduces interpersonal tensions and enhances the peer assessment process. Although 73.6 % of participants preferred anonymous peer assessment, the differences in preference patterns (90.5 % in the anonymous condition versus 56.5 % in the non-anonymous) suggest that the benefits of anonymity are not straightforward. Quantitative analysis revealed no significant differences in social

and interpersonal variables between conditions. Interestingly, the data gathered did indicate that there was a positive disposition toward non-anonymous approaches to peer assessment, particularly among high performers. These findings suggest that the type of intervention, combining both formative and summative peer assessments, may alleviate some of the tensions typically associated with peer assessment. Future research should explore a bifurcated approach to the examination of anonymity in formative and summative peer assessment further to better understand its potential.

Funding

This work was supported by the Universidad Francisco de Vitoria, Call for Research Educational Innovation 2023, Reference Number UFV2023-63.

CRediT authorship contribution statement

Juan Fraile: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Meritxell Monguillot:** Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Carles González-Arévalo:** Validation, Supervision, Investigation, Formal analysis, Data curation, Conceptualization. **Paula Lehane:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Data curation. **Ernesto Panadero:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

We have no conflicts of interest to disclose.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.stueduc.2024.101410.

References

Allen, D., & Mills, A. (2016). The impact of second language proficiency in dyadic peer feedback. *Language Teaching Research, 20*(4), 498–513. https://doi.org/10.1177/1362168814561902

Alonso-Tapia, J., Panadero, E., & Ruiz, M. A. (2014). Development and validity of the Emotion and Motivation Self-regulation Questionnaire (EMSR-Q). *The Spanish Journal of Psychology, 17*, Article E55. https://doi.org/10.1017/sjp.2014.41

Alqassab, M., & Panadero, E. (2020). Peer Assessment. *Routledge Encyclopedia of Education*. Routledge. https://doi.org/10.4324/9781138609877-REE15-1

Alqassab, M., Strijbos, J. W., 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

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy Practice, 5*(1), 7–73. https://doi.org/10.1080/0969595980050102

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199–231. https://doi.org/10.1111/j.1464-0597.2005.00205.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*(4), 599–618. https://doi.org/10.14742/ajet.830

Chinn, D. (2005). Peer assessment in the algorithms course. *ACM SIGCSE Bulletin, 37*(3), 69–73. https://doi.org/10.1145/1151954.1067468

Coomber, M., & Silver, R. (2010). The effect of anonymity in peer review. In In. A. M. Stoke (Ed.), *JALT2009 Conference Proceedings* (pp. 621–631). JALT.

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

Hamodi, C., López-Pastor, V. M., & López-Pastor, A. T. (2017). If I experience formative assessment whilst studying at university, will I put it into practice later as a teacher? Formative and shared assessment in Initial Teacher Education (ITE). *European Journal of Teacher Education, 40*(2), 171–190. https://doi.org/10.1080/02619768.2017.1281909

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

Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research and Development, 38*(7), 1401–1416. https://doi.org/10.1080/07294360.2019.1657807

Hogg, M. A., van Knippenberg, D., & Rast, D. E., III (2012). The social identity theory of leadership: Theoretical origins, research findings, and conceptual developments. *European Review of Social Psychology, 23*(1), 258–304. https://doi.org/10.1080/10463283.2012.741134

Hosack, I. (2004). The effects of anonymous feedback on Japanese university students' attitudes towards peer review. *Language and its Universe, 3*, 297–322.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Jessup, L. M., Connolly, T., & Tansik, D. A. (1990). Toward a theory of automated group work: The deindividuating effects of anonymity. *Small Group Research, 21*(3), 333–348. https://doi.org/10.1177/1046496490213003

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, 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. (2023a). 'Who expresses an honest opinion as it stands in Japan?': cultural issues and Japanese L2 students' experiences of face-to-face and anonymous peer review. *Higher Education Research Development, 42*(2), 322–335. https://doi.org/10.1080/07294360.2022.2073983

Kim, S. (2023b). "Sorry, I don't good English": Japanese L2 students' written peer feedback in the face-to-face and anonymous review modes. *Journal of Writing Research, 15*(2), 199–223. https://doi.org/10.17239/jowr-2023.15.02.02

Kim, S., & Lan, Y. (2021). L2 writers' perspectives on face-to-face and anonymous peer review: Voices from China. *Porta Linguarum: Revista Internacional Délelőtt Didáctica Délelőtt las Lenguas Extranjeras*, (*35*), 149–164. https://doi.org/10.30827/portalin.v0i35.13887

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

Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment Evaluation in Higher Education, 41*(2), 245–264. https://doi.org/10.1080/02602938.2014.999746

Lin, G. Y. (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 Education, 116*, 81–92. https://doi.org/10.1016/j.compedu.2017.08.010

Lipnevich, A. A., Berg, D., & Smith, J. K. (2016). Toward a model of student response to feedback. In In. G. T. L. Brown, & L. Harris (Eds.), *Human Factors and Social Conditions in Assessment* (pp. 169–185). Routledge.

Lorente-Catalán, E., & Kirk, D. (2016). Student teachers' understanding and application of assessment for learning during a physical education teacher education course. *European Physical Education Review, 22*(1), 65–81. https://doi.org/10.1177/1356336X1559035

Lu, J., & Bol, L. (2007). Examining the effectiveness of anonymity in a computer-supported peer assessment system. *Journal of Computer Assisted Learning, 23*(5), 420–431. https://doi.org/10.1111/j.1365-2729.2007.00228.x

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.), *Handbook of Human and Social Conditions in Assessment* (pp. 247–266). New York: Routledge.

Panadero, E., Alonso-Tapia, J., García-Pérez, D., Fraile, J., Galán, J. M. S., & Pardo, R. (2021). Deep learning self-regulation strategies: Validation of a situational model and its questionnaire. *Revista Délelőtt Psicodidáctica (English ed), 26*(1), 10–19. https://doi.org/10.1016/j.psicoe.2020.11.003

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment Evaluation in Higher Education, 44*(8), 1253–1278. https://doi.org/10.1080/02602938.2019.1614146

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, 48*(8), 1053–1075. https://doi.org/10.1080/02602938.2023.2164884

Panadero, E., & Broadbent, J. (2018). Developing evaluative judgement: A self-regulated learning perspective. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work* (pp. 81–89). Routledge.

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, Article 100416. https://doi.org/10.1016/j.edurev.2021.100416

Peterson, E. R., & Peterson, T. R. (2011). Feedback and self-regulated learning: Insights from supervisors' and PhD examiners' reports. *Reflective Practice, 12*(4), 467–480. https://doi.org/10.1080/14623943.2011.590346

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*, 75–99. https://doi.org/10.1007/s10212-017-0341-7

Sadler, R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment Evaluation in Higher Education, 34*(2), 159–179. https://doi.org/10.1080/02602930801956059

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *The British Journal of Educational Psychology, 77*(Pt 2), 351–378. https://doi.org/10.1348/000709906X106147

Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: a meta-analysis. *Journal of Educational Psychology, 109*(8), 1049. https://doi.org/10.1037/edu0000190

Schillings, M., Roebertsen, H., Savelberg, H., van Dijk, A., & Dolmans, D. (2021). Improving the understanding of written peer feedback through face-to-face peer dialogue: Students' perspective. *Higher Education Research Development, 40*(5), 1100–1116. https://doi.org/10.1080/07294360.2020.1798889

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*, Article 1198011. https://doi.org/10.3389/feduc.2023.1198011

Su, W. (2023). Masked ball for all: how anonymity affects students' perceived comfort levels in peer feedback. *Assessment Evaluation in Higher Education, 48*(4), 502–512. https://doi.org/10.1080/02602938.2022.2089348

Ting, K. Y. (2024). Students' perspective toward anonymous peer feedback in online writing classes. *Interactive Learning Environments, 32*(8), 4458–4468. https://doi.org/10.1080/10494820.2023.2201322

Topping, K. J. (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. J. (2013). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 395–412). SAGE Publications.

Trzesniewski, K., Yeager, D., Catalán Molina, D., Claro, S., Oberle, C., & Murphy, M. (2021). Global mindset initiative paper 3: Measuring growth mindset classroom cultures. doi:10.2139/ssrn.3911591.

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*(1), 41–54. https://doi.org/10.1016/j.edurev.2008.11.002

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. https://doi.org/10.1016/j.compedu.2014.10.005

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. *Mind in society: The development of higher psychological processes*. Harvard University Press.

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: a pairwise and network meta-analysis. *Educational Research Review, 37*, Article 100484. https://doi.org/10.1016/j.edurev.2022.100484

Yu, F. Y., & Sung, S. (2016). A mixed methods approach to the assessor's targeting behavior during online peer assessment: effects of anonymity and underlying reasons. *Interactive Learning Environments, 24*(7), 1674–1691. https://doi.org/10.1080/10494820.2015.1041405

Yu, F. Y., & Wu, C. P. (2011). Different identity revelation modes in an online peer-assessment learning environment: Effects on perceptions toward assessors and classroom climate. *Computers Education, 57*(3), 2167–2177. https://doi.org/10.1016/j.compedu.2011.05.022

Zimmerman, B. J. (2000). Attaining Self-Regulation: A Social Cognitive Perspective. In In. M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press.