UAB
Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **accepted version** of the journal article:

Ion, Georgeta; Díaz Vicario, Anna; Mercader, Cristina. «Making steps towards improved fairness in group work assessment : The role of students' self- and peer-assessment». *Active Learning in Higher Education*, Vol. 25, Num. 3 (2023), p. 425-437 DOI 10.1177/14697874231154826

This version is available at https://ddd.uab.cat/record/337465

under the terms of the IN COPYRIGHT license.

Making steps towards fair assessment through students' self and peer - assessment

Georgeta Ion, Anna Díaz-Vicario and Cristina Mercader

Universitat Autònoma de Barcelona, Spain

Accepted version of the manuscript:

Ion, G., Díaz-Vicario, A., & Mercader, C. (2024). Making steps towards improved fairness in group work assessment: The role of students' self- and peer-assessment. *Active Learning in Higher Education*, *25*(3), 425-437. https://doi.org/10.1177/14697874231154826

Making steps towards improved fairness in group work assessment: The role of students' self and peer – assessment

Abstract

Group work benefits student learning in many ways, but group work assessment may be challenging for academics and sometimes is perceived as less fair than individually set assessment tasks. Peer assessment (PA) and self assessment (SA) represents a possible approach supporting lecturers and students to better differentiate individual contributions to group-based assignments.

In this study, we explore the students' perception of SA and PA as strategies to facilitate the assessment of the process of group work and to contribute to increasing the sense of fairness in assessment.

With these aims, we designed a study of an assessment practice in which students were involved as partners in the assessment and grading process of a group project assignment. At the end of the course, and when the SA and PA experience concluded, we administered a survey. The findings show that 1) including SA and PA strategies in student tasks enhances the students' perceived sense of fairness in the assessment process, 2) the specifics of the instructional design do not have a significant effect on student perceptions of this form of assessment, and 3) sociodemographic factors such as the academic year, the degree major and the age of the students have a moderate effect. We conclude that SA and PA improve the sense of democracy of the assessment process, as this is not only in the teachers' hands but also shared and agreed upon within the group.

Keywords: fairness, self-assessment, peer-assessment, group work, active assessment

Introduction

Assessment is one of the educational processes perceived as unfair by students (Sambell et al., 1997). However, in recent years, the concern for a fairer and ethical assessment practice has been seen as the primary target of "new generation assessment environments" (Nieminen & Tuohilampi, 2020, p.7). The perception of fairness in assessment is even more critical when the task involves group work. Despite the research acknowledging the benefits of group work for student learning (Gibbs, 2009; Hannaford, 2017), there are a wide range of studies that indicate concern that group work assessment may be less fair than individually set assessment tasks (Barfield, 2003; Nordberg, 2008).

The development of effective group work assessment strategies, designed to engage students in authentic experiences and to induce the best learning and competencies, raises a number of important issues. For example, how should we assess group work, and where does the balance lie between product and process and between group and individual?

This article reports on an assessment practice aiming to promote a fair method of group work assessment for the Faculty of Educational Studies at the Universitat Autònoma de Barcelona (UAB), which was conceived to allow individualized marks for each student. A rubric of self (SA) and peer assessment (PA) of the group work process conducted during an assignment was designed. Once the practice concluded, a study was carried out exploring students' perceptions of the role of PA and SA in increasing the sense of fairness in group work contexts. The perceived benefits and challenges associated with PA and SA in group work assessment were discussed.

Group work learning as a form of assessment: Benefits and challenges

Collaborative learning through group work is considered a valuable form of learning in many contexts. Group work promotes deep and active learning (Entwistle & Waterston, 1988; Ruel et al., 2003) and provides students with experiential and authentic learning. It is linked to future students' employability skills and as training in collaboration, represents a means to one of the most valuable skills demanded by work places (Maguire & Edmondson, 2001).

However, while group work is considered a high-potential learning setting, assessment-related issues can arise (Hassanien, 2006). For instance, the group work assessment should consider not only the product but also the process carried out by the group members to reach the final outcome (Glebhill & Smith, 1996). In many cases, teachers decide to apply assessments and evaluation of the final outcome or final learning product resulting from the group effort and

assign the same grade to all group members (Mellor, 2012). Because one of the challenges of the group work assessment is to detect the individual contributions to the final product (Dijkstra et al., 2016; Forsell et al., 2020; Van Aalst, 2013) and evaluate individual effort, assessment could raise ethical issues for both teachers and students. This happens especially when the project's elaboration involves significant autonomous work by the group, out of the instructor's careful supervision or access to continuous assessment procedures. Students prefer an individual assessment (Parsons, 2004) or complain about group marks and disagree with strategies that assess group work by focusing only on the product (Barfield, 2003; Dyrud, 2001).

If groups are very large, personal motivation and effort tend to be lower than in individual work scenarios (Springer et al., 1999). This can expose students to contradictory assessment results which lead them to consider assessment unreliable (Perry, 2010). Both teachers and students perceive this situation as unfair (Barfield, 2003; Forsell, Forslund & Chiriac, 2021), and these types of arrangements rarely lead to positive student learning behaviour.

Although work groups are a widely used practice in higher education settings, the recent study of Meijer et al. (2020) concludes that there is a lack of both practical methods and research literature concerning group work assessment. Forsell et al. (2020, 2021) point out that a common method used to overcome the challenges in group assessment is the use of peer assessment, where students assess each other's contributions in group work. However, there is not much evidence on the efficacy of peer or self-assessment as an approach to help lecturers and students arrive at individual marks in collaborative working environments.

To fill this gap, in this paper, we will present an assessment practice using SA and PA to assess the process in working group settings, and we will explore the perceptions students have regarding whether and how this practice contributed to a higher sense of fairness in assessment.

Peer and self-assessment in collaborative group work tasks: Contributions to the sense of fairness

Fairness in assessment has been an area of study for decades from different perspectives and views. A fair assessment process has been linked to the opportunities given to all students to learn through the assessment outcomes (Linn, 1993) and to transparency in the definition and sharing of criteria (Linn, Baker & Dunbar, 1991). It has also been discussed in relation to the assessment strategies employed since "*as with traditional forms of assessment, questions of fairness arise in the selection of tasks and the grading of responses and the more informal and*

open-ended such assessment becomes, the greater the reliance on the judgment of the teacher/assessor" (Gipps & Stobart, 2009, p.107). In addition, a fair assessment is also associated with terms such as equity, validity, or ethics.

Ways to make assessment fairer come in a wide variety of forms, such as collecting data from multiple sources (including students themselves), using a range of assessment tasks involving a variety of contexts, response formats and styles, and drawing on a comprehensive approach to make decisions (Gipps & Stobart, 2009). This broader approach is likely to offer students opportunities to demonstrate their performance and to engage actively with their own learning and peers' learning.

There are several reasons to consider students' participation in the assessment as key to securing fair assessment procedures. The first is linked to the idea that SA and PA encourage students not to repeat/reproduce patterns of assessment (teacher-driven procedures) but to create their own spaces and to make visible their own voices (Flores et al., 2015) in a context especially designed for and with them. Second, considering students' right to participate in assessment contributes to promoting students' agentic behaviour (Andrade, 2010), understood as a capacity with deep implications for students' learning strategies and self-efficacy (Bandura) (Winstone, Pitt & Nash, 2021) and self-regulation (Panadero et al., 2019; Panadero & Alqassab, 2019). Third, SA and PA benefit student learning in many ways (Brown & Harris, 2013; Cheng & Warren, 2000), producing a positive impact on interpersonal and intrapersonal skills, which are highly valued for future professional development (Simpson & Clifton, 2015).

Despite the identified benefits of PA and SA in practice, studies developed worldwide and in Spanish universities show that most assessment practices maintain traditional approaches, and PA and SA are still extremely rare (Panadero et al., 2019). The findings of Panadero et al. (2019) are among the most recent to identify a lack of student involvement in assessment, but previous works reveal the same trend and show nothing has changed in recent years in the Spanish context (Panadero et al., 2014; Panadero & Brown, 2017; Quesada-Serra et al., 2017). The PA and SA procedure is rarely used in practice and is almost never included in the course syllabus; a study developed by Fraile et al. (2018) showed that only 7.55% of course syllabi include self-assessment, 9.43% include self-grading and 1.99% include both.

As mentioned before, in this paper, we will present SA and PA as strategies used to assess the process in working group settings, and we will explore the potential of this practice to ensure a fairer sense of assessment. Concretely we expect to address the following research hypotheses:

H1: Active learning activities through peer and self-assessment tasks contribute to fostering a sense of fairness in the assessment process.

H2: The instructional and assessment design affects the perception of fairness.

H3: Demographic and academic variables affect the perception of fairness in the assessment process.

Method

Description of the assessment practice

We present an assessment practice implemented in the School of Educational Sciences of the UAB (Spain) in six subjects of the bachelor's degree, including primary education, social education, education studies and early childhood education, involving 240 students in total. Students were involved in a PA and SA experience in which they assessed themselves and their peers regarding their involvement in the development of a group work project assignment.

In this practice, students had to develop a collaborative group work assignment about a specific topic in the subject. In each group, there were between 4 and 6 students. In some cases, the group composition was established by the teachers, some had the opportunity to choose their own group, and others were created randomly or using other criteria, such as choosing the topic, or adhering to previously established conditions.

Teachers used different instructional and assessment designs; thus, the group project and PA and SA tasks were adapted to fit the characteristics of the subject, the content taught, and the competencies expected to be developed. Using Topping's constellation of peer feedback in higher education as a reference (1998), Table 1 summarizes the modalities adopted to develop the PA and SA tasks in the subjects in which we have developed the practice.

- insert table 1 here-

Table 1. Modalities of PA and SA implemented in our experience

Subject	School organization, education and educational contexts, organization and groups		
Objective	(1) To allow teachers to individualise the grades of students who developed group tasks, and (2) To involve the students in the assessment process		
Focus	Summative	Formative	Both

Product/Output	Grade		
Relation to staff assessment	Substitutional		
Official Weight	% of the final mark		
Directionality	Reciprocal		
Privacy	Anonymous	Confidential	Public
Contact	Distance	Face-to-face	
Year	Same		
Ability	-		
Constellation assessor	Groups	Individuals	
Constellation assessed	Individuals		
Place	In class	Out of class	
Time	Class time	Free time	
Requirement	Compulsory		
Reward	Without		

At the beginning of the assessment practice, teachers explained to students their role as assessors with the PA and SA rubric and trained the students how to implement it in their group activity. Students participated in the assessment of the group work process using the PA and SA rubric, while teachers assessed the final product (group work project).

To assess the group process, a group of lecturers designed a rubric that includes 21 indicators organized into six main dimensions: implication and participation, acceptance and fulfilment of tasks, responsibility and commitment, contribution to the work environment, analysis and task contribution, and communication (Agud, Alcalde, Armengol et al., 2018). The PA and SA rubric has a four-level Likert scale to measure each indicator. To collect the peer and self-assessments carried out by the students, the CoRubrics add-on (https://www.corubrics.org) was used.

The rubric was administered differently in each subject depending on three factors: who fills in the rubric, when the students fill in the rubric and how the grades are established (see Table 2). Teachers also decided how the results were shared once the process was finished: anonymously, so nobody knows the individual grades; privately, so the student receives an average score of each of their grades; or publicly, so the student knows the details of their assessment and who assessed them.

- Insert table 2 here-

Table 2. Uses of the PA and SA rubric for assessing individual participation in group work (Ion, et al, 2021, p. 164).

	Who fills in the rubric?	When do students fill in the rubric?	How is the score established for each indicator?
Self-assessment	Each student about themselves	Individually, during virtual or face-to-face sessions	Each student decides the score according to their assessment
Peer-assessment	Each member of the group individually fills in the rubric for each peer of the group	Individually, during virtual or face-to-face sessions	Each student decides the score according to their personal assessment
	The group fills in the rubric collaboratively for each member	During virtual or face-to-face class sessions or outside the class session	Each indicator score is agreed on collaboratively among peers after group decision-making

At the end of the experience, each group had completed a project that was assessed and marked by the teacher, and each student assessed their contribution and their peers' contribution to the group project using the rubric. The ratio of the two final percentages were established by the lecturer and were formed by the mark on the project (percentage between 50 and 90% of the total score) and the PA and SA mark given by each of the members of the group (percentage between 10 and 50%).

The study

When the PA and SA experience concluded, a survey was administered to collect students' perceptions about the assessment and their views on how their involvement in the assessment process contributed to a sense of fairness. The survey included 20 variables organized into four dimensions adapted by Ion, Muñoz & Rodriguez (2022) from the scale proposed by Valentine et al (2021): (1) fairness in human judgement and credibility; (2) multiplicity of opportunities and assessors; (3) procedural fairness and transparency; and (4) fitness for purpose (Table 3). Each dimension was measured with a 7-point Likert scale, and the responses were collected online through the SurveyMonkey platform.

At the same time, the survey included general questions about the instructional design, the instructions given to developing the PA and SA activity, previous experiences in similar activities, and sociometric information.

- Insert table 3 here-

Table 3. Dimensions and variables of the survey

Dimensions	Variables
Fairness in human judgement and credibility	1. It changed my way of understanding assessment
	2. I have the feeling that I assessed my friends better.
	3. I have the feeling that my friends assessed me better.
	4. I trust my classmates' judgement.
	5. I am satisfied with the peer-assessment experience developed
Multiplicity of opportunities and assessors	6. It helped me to know various assessment methods, instruments, and resources.
	7. It helped me to improve my participation within the group
	8. It made me aware of how I can improve my gaps regarding relations with classmates
	9. It caused me tension or conflicts with classmates
Procedural fairness and transparency	10. It helped me to use assessment criteria
	11. It helped me to assimilate assessment criteria
	12. It made the assessment process fairer
	13. It has contributed to knowing my classmates better
	14. It has increased my commitment to the group.
Fitness for purpose	15. It helped me to develop my assessment skills
	16. It has contributed to improving my relationship with my classmates
	17. It made me feel that my judgement is taken into consideration
	18. It caused me stress
	19. It helped me to understand better my strong and weak spots when I work with classmates

Sample

The sample included 73.75% of the students involved in the experience (n = 177), 82.7% of whom identified themselves as female and 17.3% as male. The rubric was included in a first-year subject (72.2%) and in a second-year subject (25.6%) of the bachelor's degree of primary education (42.9%), social education (38.3%), education studies (10.3%) and early childhood education (8.6%). A total of 61.4% affirmed that they had previous experiences assessing the teamwork process.

The organization of the groups was created mostly by students (43.8%), but also through random assignment (23.3%), by lecturers (2.3%), or by other criteria (30.7%). Seventy-nine percent of the sample assessed their peers individually, and 21% did so together with the

colleague who was being assessed. The results of the assessment were given either anonymously (48.6%), privately (23.4%) or publicly (28%). Most of the students received training about PA and SA practice before carrying out the activity (78.6%), but in some cases, students were trained during the process (21.4%). Finally, the PA and SA were developed either outside of class (65.9%) or in class (34.1%) and mostly (85.9%) before they received the teachers' mark on the project.

Data analysis

Descriptive analysis and correlational analysis were applied using IBM Statistical Package for the Social Sciences (SPSS v.20). Specifically, means and standard deviations were analysed for descriptive analysis, T tests were used to compare dependent variables and the ANOVA test was used (with Bonferroni for a post hoc test) to compare the differences between the aspects of the instructional design and sociodemographic differences. In cases where there was no variance, the Kruskal–Wallis test was applied to look for significant differences.

Findings

The results are presented following the two objectives of the paper and through the demonstration of the three hypotheses formulated.

Contribution of PA and SA to the sense of fairness in the assessment process.

The first hypothesis we expected to test referred to whether active learning activities with peer and self-assessment tasks contribute to fostering the sense of fairness in the assessment process.

The findings shown in Table 4 show high scores on all the dimensions and variables (min = 4.32, max = 5.67). Except for those items that are reversed (whose means are, in fact, the low ones), the means of each item or variable are above 4, and the dimensions are close to having a mean of 5 points (scale of 1 to 7).

- Insert table 4 here-

Table 4. Means and SD by dimension and variable

Dimensions and variables	Mean	SD
Fairness in Human Judgement and Credibility	4.96	1.01
1. It changed my way of understanding assessment.	4.42	1.64
2. I have the feeling that I assessed my friends better*.	3.25	1.98
3. I have the feeling that my friends assessed me better*.	3.19	1.80

4. I trust my classmates' judgement.	5.67	1.35
5. I am satisfied with the peer-assessment experience developed.	5.15	1.60
Multiplicity of Opportunities and Assessors	4.85	1.05
6. It helped me to know various assessment methods, instruments and resources.	4.75	1.57
7. It helped me to improve my participation within the group.	4.61	1.82
8. It made me aware of how I can improve my gaps regarding my relations with classmates.	4.55	1.75
9. It caused me tension or conflicts with classmates*.	2.50	1.85
Procedural Fairness and Transparency	4.90	1.21
10. It helped me to use assessment criteria.	5.28	1.43
11. It helped me to assimilate assessment criteria.	4.96	1.43
12. It made the assessment process fairer.	4.97	1.73
13. It has contributed to knowing my classmates better.	4.54	1.79
14. It has increased my commitment to the group.	4.76	1.61
Fitness for Purpose	4.93	1.10
15. It helped me to develop my assessment skills.	5.14	1.44
16. It has contributed to improving my relationship with my classmates.	4.32	1.72
17. It made me feel that my judgement is taken into consideration.	5.34	1.60
18. It caused me stress*.	3.17	1.98
19. It helped me to understand better my strong and weak spots when I work with classmates.	4.99	1.69

*To calculate the mean of the dimension, reversed variables marked with (*) have been transformed.

The results show that the dimension with the highest mean is "Fairness in Human Judgement and Credibility" (M = 4.96; SD = 1.01), and the lowest mean is "Multiplicity of Opportunities and Assessors" (M = 4.85; SD = 1.05), although there are no statistically significant differences according to the T test results. Regarding the specific variables, the highest are "I trust my classmates' judgement" (M = 5.67; SD = 1.35), "It made me feel that my judgement is taken into consideration" (M = 5.34; SD = 1.60), "It helped me to use assessment criteria" (M = 5.28; SD = 1.43), "I am satisfied with the peer-assessment experience developed" (M = 5.15; SD = 1.60) and "It helped me to develop my assessment skills" (M = 5.14; SD = 1.44). On the other hand, the variables with the lowest scores are "It caused me tension or conflicts with classmates" (M = 2.50; SD = 1.85), "It caused me stress" (M = 3.17; SD = 1.98), "I have the feeling that my friends assessed me better" (M = 3.19; SD = 1.80), and "I have the feeling that I assessed my friends better" (M = 3.25; SD = 1.98), which correspond to the reversed items, describing negative outcomes that could arise from the experience.

Therefore, because dimensions and variables show high means and the reversed variables show low means, we accept the first hypothesis "Active learning activities through peer and self-assessment tasks contribute to fostering the sense of fairness in the assessment process".

Instructional and assessment design aspects influencing the perception of fairness in the assessment

The second hypothesis we wanted to test was whether the instructional and assessment design affects the perception of fairness. The different assessment design aspects applied were privacy (anonymous, confidential, public), constellation assessor (person assessed is also an assessor or not) and place (in class, out of class). ANOVA was applied for each variable and for the mean of the dimension. The results show that only the constellation assessor affects some of the results. Specifically, there are significant differences in the variable "I trust my classmate's judgement" [$F(1, 174) = 7.60$, $p = 0.006$]. When the person assessed does not participate in their assessment, the students trust their classmate's judgement more (without the assessed, $M = 5.81$, $SD = 1.34$, together with the assessed, $M = 5.14$, $SD = 1.27$), although the mean is still fairly high and confirms that they generally trust their classmate's judgement. The other significant difference is found in the variable "It made the assessment process fairer" [$F(1, 174) = 6.60$, $p = 0.011$]. In the same vein, when the assessment is developed together with the peer that has been assessed, the perception of fairness is lower ($M = 4.32$; $SD = 1.87$) than when they develop the assessment individually ($M = 5.14$; $SD = 1.67$). Finally, the mean of the dimension of fairness of human judgement and credibility follows the same path [$F(1, 174) = 4.90$, $p = 0.028$], since when the assessment is developed individually, it is more trustworthy ($M = 5.05$; $SD = 1.01$) than when the person assessed is part of the process ($M = 4.64$; $SD = 0.96$).

The level of privacy when sharing the assessment results and the place where the assessment is developed does not affect the perception of fairness, since not a single variable shows statistical significance in the ANOVA test. However, in general, the lowest scores across dimensions are when the assessment has been shared publicly, except in the dimension of fairness in human judgement and credibility, which is, in fact, the highest ($M = 5.03$; $SD = 1.00$). The other aspects of the experience, such as the way teams were created (randomly, by teachers, by students, other criteria), the moments students received training about PA and SA (previously or during the assessment process) and the moment when students developed their PA (before receiving the teachers' grades or afterwards), also did not show statistically significant differences.

Therefore, since the significant differences are exceptional when comparing the modalities of the instructional designs, we reject H2: "Instructional and assessment design affect the perception of fairness in the assessment process".

Demographic and academic variables influencing the perception of fairness in the assessment

The third hypothesis we studied refers to the demographic variables and their effect on fairness. The sociodemographic variables considered in the survey were academic year, final grade obtained in the project, gender, age and the major in which students were enrolled. The academic year is the factor that more often shows significant differences in the ANOVA test (and Bonferroni post hoc test). The bachelor's degree specialisation also makes significant differences in four variables and grade obtained in the project and age in only two variables. On the other hand, gender does not affect any of the variables.

As described in Table 4, regarding the academic year, 2nd-year students perceived the fairness of the process to be lower than the 1st-year students in half of the variables. In the dimension "fairness in human judgement and credibility", two variables are affected by the academic year: "It changed my way of understanding assessment" [$F(2, 173) = 4.33$, $p = 0.012$], where the mean of 2nd-year students was 3.80 (SD = 1.69) and 1st-year students was 4.61 (SD = 1.58), and "I am satisfied with peer-assessment experience developed" [$F(2, 173) = 6.50$, $p = 0.002$], where the mean of 2nd-year students was 4.49 (SD = 1.75) and 1st-year students 5.41 (SD = 1.47). In the "multiplicity of opportunities and assessors" dimension, the mean of the dimension is significantly lower for 2nd-year students (M = 4.48, SD = 1.01) than for 1st-year students (M = 5.00, SD = 1.04) [$F(2, 172) = 5.20$, $p = 0.011$]. Moreover, there were significant differences in the variable "It helped me to know various assessment methods, instruments and resources" [$F(2, 173) = 5.60$, $p = 0.004$]. In the same vein, 2nd-year students showed less agreement with the affirmation (M = 4.09, SD = 1.74) than 1st-year students (M = 4.98, SD = 1.45).

In the third dimension, "procedural fairness and transparency", there are significant differences in the mean between 2nd-year students (M = 4.44, SD = 1.38) and 1st-year students (M = 5.06, SD = 1.08) [$F(2, 171) = 4.82$, $p = 0.008$] as well as in the variable "It made the assessment process fairer" (2nd-year students: M = 4.24, SD = 1.91; 1st-year students: M =5.24, SD = 1.61) [$F(2, 173) = 5.88$, $p = 0.003$]. Finally, the last dimension "fitness for purpose" also shows differences between 1st- and 2nd-year students. The mean of the whole dimension is lower for 2nd-year students (M = 4.50, SD = 1.13) than for 1st-year students (M = 5.10, SD = 1.05) [$F(2, 171) = 6.15$, $p = 0.002$]. Focusing on the specific variables of this dimension, this difference also occurs in each of the following variables: "It helped me to develop my assessment skills" with lower means

in 2nd-year students (M = 4.60, SD = 1.64) than 1st-year students (M = 5.32, SD = 1.33) [F(2, 173) = 4.27, p = 0.012]; "It made me feel that my judgement is taken into consideration" (2nd-year students: M = 4.58, SD = 1.82; 1st-year students: M = 5.62, SD = 1.44) [F(2, 173) = 7.87, p = 0.000] and "It helped me to understand better my strong and weak spots when I work with classmates", with 2nd-year students obtaining a mean of 4.42 (SD = 1.91) and 1st-year students a mean of 5.19 (SD = 1.58) [F(2, 173) = 3.54, p = 0.027].

Even though the sample has four options for tracks in their bachelor's degrees (early childhood education, primary education, education studies and social education), the differences are largest between education studies and social education, with the latter showing the lowest means. In the first dimension, "fairness in human judgement and credibility", the variable "I am satisfied with the peer-assessment experience developed" shows significantly lower means among social education students (M = 4.87, SD = 1.94) than that of those enrolled in education studies (M = 6.11, SD = 1.18) [F(3, 171) = 2.97, p = 0.018]. In the third dimension, "procedural fairness and transparency", the variable "It made the assessment process fairer" shows significantly lower means in social education students (M = 4.55, SD = 1.86) than students in education studies (M = 5.94, SD = 1.56) [F(3, 171) = 4.35, p = 0.007]. The fourth dimension, "fitness for purpose", shows differences in the variable "It made me feel that my judgement is taken into consideration", with a lower mean in social education students (M = 4.93, SD = 1.74) than education studies students (M = 6.17, SD = 1.15) [F(3, 171) = 3.67, p = 0.020]. Moreover, the mean of the whole dimension is significantly different depending on the degree track [F(3, 169) = 2.88, p = 0.038], and although the ANOVA test does not show a particular track as lower or higher, social education students show the lowest mean (M= 4.69, SD = 1.09), and students from education studies the highest (M = 5.40, SD = 0.97).

The grade obtained is a significant factor in two variables. Students who received a C on the task agreed less with "I am satisfied with the peer-assessment experience developed" (M = 4.43, SD = 1.91) than students who had an A (M = 5.9, SD = 1.25) [F(4, 116) = 3.39, p = 0.028]. Additionally, the comparison test shows that the lower the grade obtained, the lower the mean of "It made me feel that my judgement is taken into consideration" [F(4, 116) = 2.57, p = 0.041]. The results do not show any specific age where the means are higher or lower. However, in general, the older the student is, the lower the mean is, and the younger the student is, the higher the mean is. This is statistically significant in the case of "I am satisfied with the peer-assessment experience developed" [F(5, 166) = 3.09, p = 0.041] and the mean of the dimension "fitness for purpose" [F(5, 164) = 2.48, p = 0.034].

Therefore, we accept the third hypothesis formulated: demographic and academic variables affect the perception of fairness in the assessment process, specifically considering the academic year of the students and the track of the bachelor's degree they are enrolled in.

Discussion and conclusion

In our study, we have started from the premise that, in the university context, group work is acknowledged as beneficial for student learning, contributing to the development of the students' professional skills but representing a challenge for the students' assessment. Students tend to find this practice unfair, especially when the task mark is applied equally to all the group members (Barfield, 2003; Forsell, Forslund & Chiriac, 2021). Although group work is a complex teaching and learning activity to design and implement, it is also an important and valuable form of assessment widely used in higher education, with positive effects for both students and lecturers (Davis, 2009).

In the present paper, we have explored PA and SA as strategies aiming to overcome the dilemma of the students' assessment in work group settings. Our argument claimed that 1) involving students in their assessment contributes to making assessment more transparent, participative, and fairer. 2) Assessing the process carried out by students during the development of their group assignment could represent a way to engage students with the task and contribute to making students aware of the importance of collaboration and group work skills. In addition, we assumed that assessing the process not only contributes to overcoming the potential problem of assigning all the group members the same grade but also makes group dynamics fairer and more accurate. 3) PA and SA could offer an adequate context for involving students as active agents in their assessment, allow lecturers to distinguish between the performance of individuals in a group setting and, in this way, enhance the sense of fairness in assessment.

The study allows us to draw some suggestions that can assist academics in implementing group task assessment. First, our findings have shown that students believe SA and PA make the assessment process fairer in a context in which SA and PA procedures are extremely rare. The students valued being an active part of the assessment process and engaged with the assessment criteria at the beginning of the process. Overall, students find SA and PA to be a satisfactory experience and a valuable strategy that enhances their sense of fairness in assessment. Additionally, as future educators, they feel that this experience allows them to develop their evaluative judgement and to improve their assessment skills.

Second, the perception of fairness is influenced by some of the singularities of the experience design and implementation, from group-forming criteria to the group climate and trust between members. For instance, the perception of fairness is higher when the PA has been conducted individually than shared with peers. This idea leads us to suggest that it is necessary to establish time and space to conduct the SA and PA processes individually, as agreeing with peers could reduce the objectivity in student assessment.

Third, being involved in PA and SA does not generate tensions at the group level, as might have otherwise been expected. At the same time, the SA and PA strategy democratizes the assessment process, as this is not only in the teachers' hands but also shared and agreed upon within the group. Managed carefully and adapted to the group characteristics, PA can be a valuable source for group cohesion and development. Creating the context for this to succeed is in the lecturer's hands.

Fourth, involving students in the assessment process through PA and SA practices contributes to enhancing the sense of responsibility, autonomy and reflection among students, as many studies have demonstrated (Panadero et al., 2019).

Any decision in teaching on the type of assessment we choose as academics should be a thoughtful and informed one, taking into consideration not only recent studies but also class characteristics. Therefore, this requires knowing and understanding our students' needs and interests and designing their learning process considering aspects such as the group composition, size and members' level of trust and confidence.

Despite the limitation of the study to a small sample of students and teachers, the study focused on the magnitude of the fairness and equity concepts attached to the assessment processes and the responsibility we all have in designing and conducting assessment practices that allow individualization and personalization of student learning. In addition, the study we conducted serves to raise awareness of the importance of student voices in the assessment process and their responsibility as active partners in it. Finally, we have attempted to propose strategies that challenge the uniformity of grades in group settings and highlight the importance of PA and SA as a means towards making assessment a fairer and more transparent process that is more democratic and formative.

References

Agud, I., Alcalde, R., Armengol, C., Barrera-Corominas, A., Borras, A., Buñuel, F., Castro, D., Díaz-Vicario, A., Duran-Bellonch, M., González, H., Ion, G., Mercader, J., Muñoz, J.L., Orta, M., Parera, M., Pérez-Maldonado, C., Rodríguez-Gómez, D., Salat, E., Sánchez, A., & Zolle-Valero, T. (2018) (Coord..). *Pauta per a l'avaluació de la contribució individual dels membres al procés grupal en el treball en equip universitari*. Dipòsit Digital de Documents. https://ddd.uab.cat/record/196345

Andrade, H.L. (2010). Students as the Definitive Source of Formative Assessment: Academic Self-Assessment and the Self-Regulation of Learning. In: *NERA Conference Proceedings 2010,* 25 (pp. 1-18). https://opencommons.uconn.edu/nera_2010/2

Barfield, R.L. (2003). Students' perceptions of and satisfaction with group grades and the group experience in the college classroom. *Assessment and Evaluation in Higher Education, 28*(4), 355-370. https://doi.org/10.1080/0260293032000066191

Brown, G. T. L., & Harris, L. R. (2013). Student Self-Assessment. In J. H. McMillan (Ed.), *The SAGE Handbook of Research on Classroom Assessment* (pp. 367-393). Thousand Oaks. https://doi.org/10.4135/9781452218649.n21

Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in higher education, 5*(2), 243-255. https://doi.org/10.1080/135625100114885

Dijkstra, J., Latijnhouwers, M., Norbart, A., & Tio, R. A. (2016). Assessing the "I" in group work assessment: State of the art and recommendations for practice. *Medical Teacher, 38*(7), 675–682. https://doi.org/10.3109/0142159X.2016.1170796

Dyrud, M.A. (2001). Group projects and peer review. *Business Communications Quarterly 64*(4), 106-12.

Entwistle, N., & Waterston, S. (1988). Approaches to studying and levels of processing in university students. *British journal of educational psychology, 58*(3), 258-265. https://doi.org/10.1111/j.2044-8279.1988.tb00901.x

Flores, M.A., Veiga Simão, A.M., Barros, A. & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. *Studies in Higher Education, 40*(9), 1523-1534. https://doi.org/10.1080/03075079.2014.881348

Forsell, J., Forslund Frykedal, K., & Chiriac, E. H. (2020). Group work assessment: assessing social skills at group level. *Small Group Research, 51*(1), 87–124. https://doi.org/10.1177/1046496419878269

Forsell, J., Forslund Frykedal, K., & Chiriac, E. H. (2021). Teachers' perceived challenges in group work assessment. *Cogent Education, 8*(1), 1-16. ttps://doi.org/10.1080/2331186X.2021.1886474

Fraile, J., Pardo, R., & Panadero, E. (2018). Autoevaluación y autocalificación en el Grado en Ciencias de la Actividad Física y del Deporte: estudio censal de las guías docentes. *Profesorado. Revista de Currículum y Formación del Profesorado, 2*(3), 163-182. https://doi.org/10.30827/profesorado.v22i3.7997

Gipps, C. & Stobart, G. (2009). Fairness in assessment. In C. Wyatt-Smith & J.J. Cumming (ed.), *Educational assessment in the 21st century* (pp. 105-118). Springer.

Glebhill, M. & Smith, P. (1996). *Student perceptions of learning with reference to group work*. Occasional Paper 12, Buckingham College of Higher Education, Brunel University, London.

Hannaford, L. (2017). Motivation in group assessment: a phenomenological approach to post-graduate group assessment. *Assessment and Evaluation in Higher Education, 42*(5), 823-836. https://doi.org/10.1080/02602938.2016.1195787

Hassanien, A. (2006). Student experience of group work and group assessment in higher education. *Journal of teaching in travel & tourism, 6*(1), 17-39. https://doi.org/10.1300/J172v06n01_02

Ion, G., Díaz-Vicario, A., Barrera-Corominas, A., Mercader, C., & Suárez, C.I. (2021). Encouraging co-responsibility: a tool for students' involvement in the assessment process. In P. Baughan (ed.), *Assessment and Feedback in a Post-Pandemic Era: A Time for Learning and Inclusion* (pp. 161-168). Advance HE.

Linn, R.L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15*(1), 1-16. https://doi.org/10.3102/01623737015001001

Linn, R.L., Baker, E., & Dunbar, S (1991). Complex, performance based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21. https://doi.org/10.2307/1176232

Maguire, S., & Edmondson, S. (2001). Student evaluations and assessment of group projects. *Journal of Geography in Higher Education, 25*(2), 233–240. https://doi.org/10.1080/03098260120067664

Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education, 45*(8). https://doi.org/10.1080/0

Nieminem, J., & Tuohilampi, H. (2020). 'Finally studying for myself' – examining student agency in summative and formative self-assessment models. *Assessment & Evaluation in Higher Education, 45*(2), 1-15. https://doi.org/10.1080/02602938.2020.1720595

Nordberg, D. (2008). Group projects: More learning? Less fair? A conundrum in assessing postgraduate business education. *Assessment & Evaluation in Higher Education, 33*(5), 481-492. https://doi.org/10.1080/02602930701698835

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. 10.1080/02602938.2019.1600186

Panadero, E., & Brown, G.T.L. (2017). Teachers' Reasons for Using Peer Assessment: Positive Experience Predicts Use. *European Journal of Psychology of Education, 32*(1), 133–156. https://doi.org/10.1007/s10212-015-0282-5

Panadero, E., Brown, G.T.L., & Courtney, M.G.R. (2014). Teachers' Reasons for Using Self-Assessment: A Survey Self-Report of Spanish Teachers. *Assessment in Education: Principles, Policy & Practice, 21*(4), 365–383. https://doi.org/10.1080/0969594X.2014.919247

Panadero, E., Fraile, J., Fernández-Ruiz, J., Castilla-Estevez, D., & Ruiz-Díaz, M.A. (2019). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment and Evaluation in Higher Education, 44*(3), 379-397. https://doi.org/10.1080/02602938.2018.1512553

Parsons, D. (2004). Justice in the classroom: Peer assessment of contributions in group projects. *Proceedings of the 17th NACCQ, Christchurch*, 145-152.

Perry, B. (2010). Exploring academic misconduct: Some insights into student behaviour. *Active Learning in Higher Education, 11*(2), 97-108. https://doi.org/10.1177/1469787410365657

Quesada, V., Rodríguez, G., & Ibarra, M.S. (2017). Planificación e Innovación de la Evaluación en Educación Superior: la Perspectiva del Profesorado. *Revista de Investigación Educativa, 35*(1), 53–70. https://doi.org/10.6018/rie.35.1.239261

Ruel, G., Bastiaans, N., & Nauta, A. (2003). Free riding and team performance in project education. *International Journal of Management Education, 3*(1), 26–38.

Sambell, K., McDowell, L., & Brown, S. (1997). 'But is it fair?': an exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23*(4), 349–371. https://doi.org/10.1016/S0191-491X(97)86215-3

Simpson, G., & Clifton, J. (2015). Assessing postgraduate student perceptions and measures of learning in a peer review feedback process. *Assessment & Evaluation in Higher Education, 41*(4), 501-514. https://doi.org/10.1080/02602938.2015.1026874

Springer, L., Stanne, M.E., & Donovan, S.S. (1999). Effects of small group learning on undergraduate Science, Mathematics, Engineering and Technology: a meta-analysis. *Review of Educational Research, 69*(1), 21-51. https://doi.org/10.3102/00346543069001021

Topping, K.J. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.2307/1170598

Van Aalst, J. (2013). Assessment in collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. K. K. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 280-296). Routledge.

Valentine, N., Durning, S., Shanahan, E. M., & Schuwirth, L. (2021). Fairness in human judgement in assessment: a hermeneutic literature review and conceptual framework. *Advances in Health Sciences Education,* 26(2), 713-738.

Winstone, N., Pitt, E., & Nash, R. (2021). Educators' perceptions of responsibility-sharing in feedback processes. *Assessment & Evaluation in Higher Education, 46*(1), 118-131. https://doi.org/10.1080/02602938.2020.1748569

Funding:

The project titled "*Enhancing the individualized evaluation of group projects through peer evaluation strategies and feedback*" was funded by the Teaching Innovation Projects initiative of the Office of Teaching Quality (OQD) at the Universitat Autònoma of Barcelona (2017).

Supplementary Material

Please find the following supplemental material visualised and available to download via Figshare in the display box below.

A-nova analysis:

Descriptive analysis:

Means and SD of each item of the survey: