Form@re
Open Journal per la formazione in rete

articoli

Peer assessment practices in an online context: does the group size matter?

Pratiche di valutazione fra pari in un contesto on line: quale impatto ha la numerosità del gruppo?

Chiara Pelati[a], Valentina Grion[b], Lan Li[c], Anna Serbati[d,1]

[a] *Università degli Studi di Padova, chiarapelati.cp@gmail.com*
[b] *Università degli Studi di Padova. valentina.grion@unipd.it*
[c] *Bowling Green State University (US), lli@bgsu.edu*
[d] *Università degli Studi di Padova, anna.serbati@unipd.it*

Abstract

The present study examined how group size may impact students' participation, perceptions, and work quality in an online peer assessment activity. One hundred sixty-three college freshmen were randomly assigned into three conditions that consisted of either 3-student groups, 6-student groups, or 9-student groups. Students reviewed each other's projects within their groups. Upon receiving peer feedback, students improved their own work. The data analysis suggested that students in bigger groups more actively participated in peer review discussion than students in smaller groups. For students' perceptions, students' responses to SCS and IMI questionnaires did not show considerable differences between the three groups. For work quality, it seemed that students in bigger groups outperformed students in smaller groups.

Keywords: peer assessment; group size; feedback; students' participation; higher education.

Sintesi

Il presente lavoro considera l'impatto della numerosità del gruppo su partecipazione, percezioni e qualità del lavoro degli studenti in attività on line di valutazione fra pari. 163 matricole sono state distribuite, in modo casuale, in gruppi di lavoro da tre, da sei o da nove studenti. Essi hanno esaminato i prodotti dei pari all'interno del proprio gruppo. Dopo aver ricevuto feedback dai pari, gli studenti hanno migliorato il proprio lavoro. L'analisi dei dati ha permesso di rilevare che gli studenti di gruppi più numerosi partecipano più attivamente alla discussione di peer review, rispetto a quelli di gruppi meno numerosi. Non vi sono, invece, differenze significative di risposta ai questionari SCS e IMI. Considerando la qualità del lavoro, si è constatato che gli studenti di gruppi più numerosi ottengono prodotti migliori di quelli di gruppi meno numerosi.

Parole chiave: valutazione fra pari; numerosità del gruppo; feedback; partecipazione; università.

[1] Ascription: the present study was conducted in full collaboration by the authors. However, C. Pelati wrote §5; V. Grion wrote §3 and §4; L. Li wrote §2; A. Serbati §1; §6 was written by V. Grion and A. Serbati together.

© 2020 Author(s). Open Access article distributed under CC BY-NC-ND 4.0

1. Overwiew on peer assessment

In the last several years, the educational literature has demonstrated the importance of students playing an active role in assessment processes (Grion & Serbati, 2019; Sambell, McDowell, & Montgomery, 2013), highlighting the importance of feedback as a means to support learning (Hattie & Timperley, 2007). Feedback represents the process, in which students receive information regarding the quality of their work by comparing their work with a standard, with the aim of improving current performance (Black & Wiliam, 2006; Sadler, 1989). Feedback is very informative and offers students the opportunity to confirm, modify, and integrate their knowledge, providing useful information for planning personal development.

The recent literature considers feedback to be a dialogical interactive process, in which teachers and students are collaboratively involved. In such a process, students play an active role and are engaged in a reflective activity aimed at improving their learning process (Evans, 2013). To make feedback received meaningful, students need to go through a cognitive process including: decoding and interpreting feedback, comparing it with their own work, identifying areas of improvement and ways to develop them, and reflecting on the overall process in order to be able to transfer what they learned to other contexts. The growth of feedback literacy (Carless & Boud, 2018) needs continuous opportunities for students to try out and distinguish different notions of quality, with the ultimate aim of progressively increasing independent evaluative judgement capability.

An effective strategy to promote an active role for students in assessment and the development of feedback literacy is to create peer review scenarios, in which students are called not only to receive and use feedback, but also to construct and provide feedback on peers' work by applying a set of criteria of quality in a concrete and rigorous way (Carless & Boud, 2018; Sambell, Brown, & Race, 2019). Research (Grion, Serbati, Tino, & Nicol, 2017; Nicol, Thomson, & Breslin, 2014) demonstrates that both processes of providing and receiving feedback may produce learning benefits. In the providing phase, students spontaneously compare their own work with that of peers and with criteria previously collectively identified (Figure 1), and then transfer ideas generated in the comparison to improve their own work. In the receiving phase, students have the opportunity to hear other points of view and become more aware of the weaknesses or possible mistakes in their own work and learn how to constructively receive comments from others.

An important factor that could amplify the benefits of peer review is to organize peer feedback scenarios in groups, where feedback processes happen in groups rather than between individuals. Serbati, Grion, and Fanti (2019) have recently found that discussion and feedback co-construction with peers increase the depth and breadth of student reflection by exposing students to diverse points of view and engaging students in intellectual loops that include active negotiation, identification of misconceptions, and peer collaboration to write appropriate feedback.

2. Group size

There is a controversy in the literature concerning the optimal size for productive and meaningful group work (Odo, Masthoff, & Beacham, 2019). Some studies report that larger group sizes correlate with better group performance, while others validate that smaller group settings are more effective (Jungjoo, 2013). For instance, Hashmi (2017) suggests that larger groups, as compared to smaller groups, generate more collective intelligence.

Yet, some researchers assess that moderate group sizes such as five (Fay, Garrod, & Carletta, 2000) or seven (Blenko, Mankins, & Rogers, 2010) may be more feasible to facilitate group work. Inconsistent results regarding group size have also been reported in the area of peer assessment. For example, Valacich, Dennis, and Nunamaker (1992) reported that students in larger nine-member peer review groups outperformed those in smaller three-member peer review groups in idea generating. Nevertheless, other studies seem to suggest that smaller peer review groups such as pairs would likely hold promise in creating and sustaining effective learning environments (e.g. Bennett, Parker, & Smigiel, 2012), and engaging students and encouraging students' participation (e.g. Hung, Chen, & Samuelson, 2016). The effectiveness of working in smaller peer assessment groups seems to be endorsed by a recent study conducted by Li, Gao, and Guo (2019). Their study investigated how social media and group size may impact students' performance and intrinsic motivation in peer assessment. Students were assigned to three groups: a three-member wiki group, a three-member social messaging group, and a six-member social messaging group. As reviewers, students rated and commented upon each other's English essays in their own respective groups. As reviewees, students received and acted upon feedback their peers provided to them. The findings of the study revealed that peer review facilitated by social messaging and wiki yielded comparable results on student essay writing. Students in the small messaging group developed better quality essays and reported higher intrinsic motivation than did students in the large messaging group.

3. Research questions

In light of the literature, the presented case study examined the effects of group size on participants' group participation, perceptions of learning activities, and work quality. Specifically, there are three research questions this study aimed to answer:

- Q1. Does group size influence student participation when working in different-sized peer assessment groups (group of 3, group of 6, and group of 9)? If so, how?
- Q2. Does group size influence student perceptions of learning activities when working in different-sized peer assessment groups (group of 3, group of 6, and group of 9)? If so, how?
- Q3. Does group size influence quality of student projects when working in different-sized peer assessment groups (group of 3, group of 6, and group of 9)? If so, how?

4. Method

4.1. Study context and participants

The study took place in an undergraduate Educational Sciences course at the University of Padua during the academic year of 2018-2019. One hundred and sixty-three freshmen (88% females and 12% males) who were enrolled in the course participated in the study. The eight-week course was designed in a blended format with face-to-face lectures alternating with online activities in a Moodle environment. This represented quite a new environment for students, because the bachelor's degree in Educational Sciences is mainly organized

with face-to-face lectures. It aimed to help students gain knowledge and transversal skills in the field of experimental pedagogy, with a focus on assessment practices.

4.2. Research design and procedures

Students were randomly assigned to three collaborative groups by using an automatic group randomisation available in the Moodle platform:

- small-size group (G1) – 57 students divided in 19 subgroups of three students;
- medium-size group (G2) – 52 students divided in 8 subgroups of six students and one subgroup of 4 students;
- large-size group (G3) – 54 students divided in 6 subgroups of nine students.

As a class activity, students participated in an online peer assessment activity in their respective groups. All students followed four steps and interacted in four different forums to perform the assigned task to build an assessment rubric aimed to assess a previously defined competence (Figure 1). Every rubric was peer assessed by two students, therefore each student received feedback from two peers. However, all participants in the group could comment or help each other during the whole process.

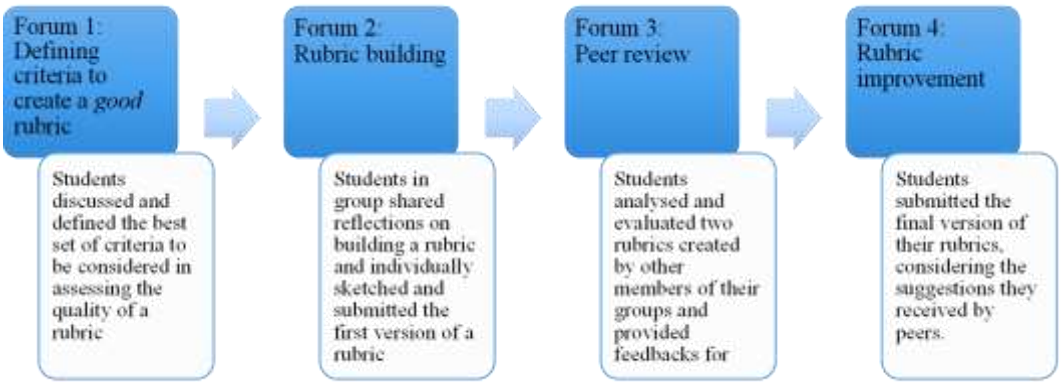

Figure 1. Four Steps Followed by Students in the Peer Assessment Activity.

The researchers adopted a simultaneous mixed-methods research model (Tashakkori & Creswell, 2007) to determine the effects of group size on student learning in groups. To answer the research questions, three main aspects were investigated: a) student participation in their respective groups, b) student perceptions, c) work quality (quality of assessment rubrics built by students). Student participation and perceptions were measured through quantitative methods on the whole sample of participants in the study (n=163). Work quality was qualitatively analyzed on a sample of 54 rubrics randomly selected from the three groups (18 rubrics from each group of G1, G2, and G3). Quantitative and qualitative data were collected and analyzed over three concurrent phases. The findings derived were then triangulated to answer the three research questions.

To analyze student participation, their posts in forums were extracted from Moodle using a tool that tracks the number of the posts for each group and each forum. Collected data were analyzed on Excel, using descriptive statistics. To collect data on student perceptions, two questionnaires were sent to students to invite students' responses after they completed the peer assessment activity. The first questionnaire contained the Sense of Community Scale (SCS). The SCS, designed by Balboni, Perrucci, Cacciamani, & Zumbo (2018),

includes 36 items on a four-point Likert scale with 1 representing "strongly disagree" and 4 representing "strongly agree". This scale aims at investigating student perceptions on three dimensions: membership (19 items), influence (6 items) and fulfilment of needs (11 items). The second questionnaire consisted of a translated and adapted version of the Intrinsic Motivation Inventory (IMI) designed by Ryan (1982). The original version of IMI has seven sub-scales and 45 items in total rated on a seven-point Likert scale, with 1 representing "not at all true" and 7 representing "very true". The researchers chose 4 subscales that were relevant to this study to be included in the survey: Interest/Enjoyment, Effort/Importance, Perceived Competence, and Relatedness. The translated Italian version of the sub-scales was validated by a panel of experts at the University of Padua and then translated back to English by a native speaker researcher.

To assess the quality of students' projects, a sample of 18 students' assessment rubrics were randomly selected from each group, totalling 54 projects. Two independent judges evaluated students' projects using criteria that rank learners based on four performance levels: Exemplary (score range: 28-30), Proficient (score range: 24-27), Developing (score range: 18-23), and Inadequate (score range: 18 or lower). The two judges discussed and negotiated to reach a consensus in case of disagreements. The assessment of students' projects included both qualitative and quantitative components. Besides assigning each project a score, the two judges also noted down the most current strengths and weaknesses of each project assessed. After the assessment procedure, the set of marks and qualitative observations were integrated to provide a global insight on the groups' performance trends.

5. Results

5.1. Student participation

Research Question 1: Does group size influence student participation when working in different-sized peer assessment groups (group of 3, group of 6, and group of 9)? If so, how?

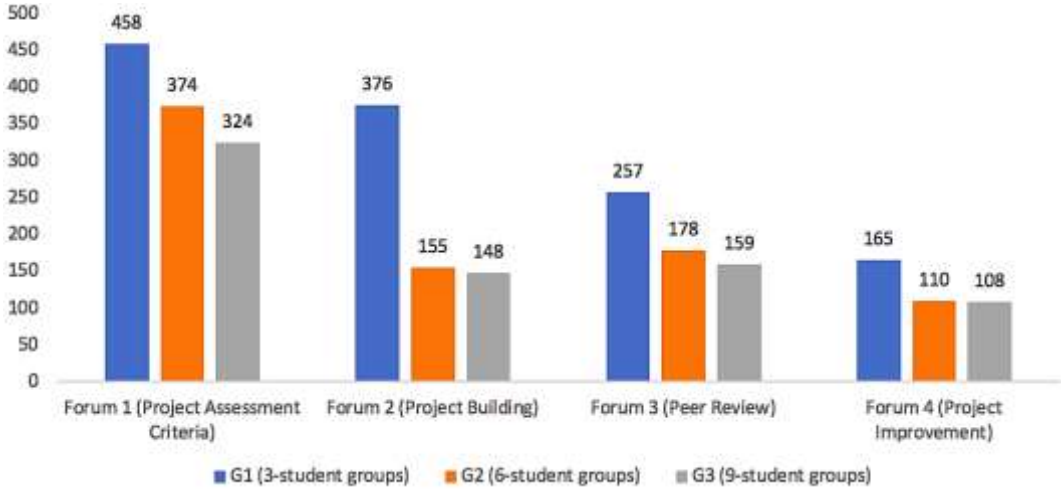

Figure 2. Total Number of Posts in Each Group in Each Forum.

G1 showed the highest performance in participation, totalling 1.256 posts across four discussion forums. G2 reported a total of 817 post contributions and G3 showed the least

active participation with 739 posts. Interestingly, each group reached a peak of activity when discussing the best set of criteria to be considered in assessing their projects (Forum 1). The participation records of each group hit a low when students were asked to improve their projects by considering the suggestions received from peers (Forum 4).

Observing the distribution of students' participation records for each group (Figure 2), we found some interesting trends. First, in the forum dedicated to peer review (Forum 3), we observed an increase of interactivity in G2 and G3, as compared with the constant upward trend of G1. Second, the forum dedicated to the rubric building (Forum 2) reported the largest gap between the number of posts of G1 and those of Group 2 and Group 3.

Considering statistics on individual participation, we calculated the average of posts per student in each group. On the overall activities, G1 students had a mean of 22 posts; G2 students 16 posts; while G3 students had an average of 14 posts per student. It seems that the average number of posts decreased as group size increased.

5.2. Student Perception

Research Question 2: Does group size influence student perceptions of learning activities when working in different-sized peer assessment groups (group of 3, group of 6, and group of 9)? If so, how?

The results of the two questionnaires (SCS and IMI) investigating students' perceptions are presented below. The analysis of students' responses to SCS did not show considerable differences between the three groups (Figure 3).

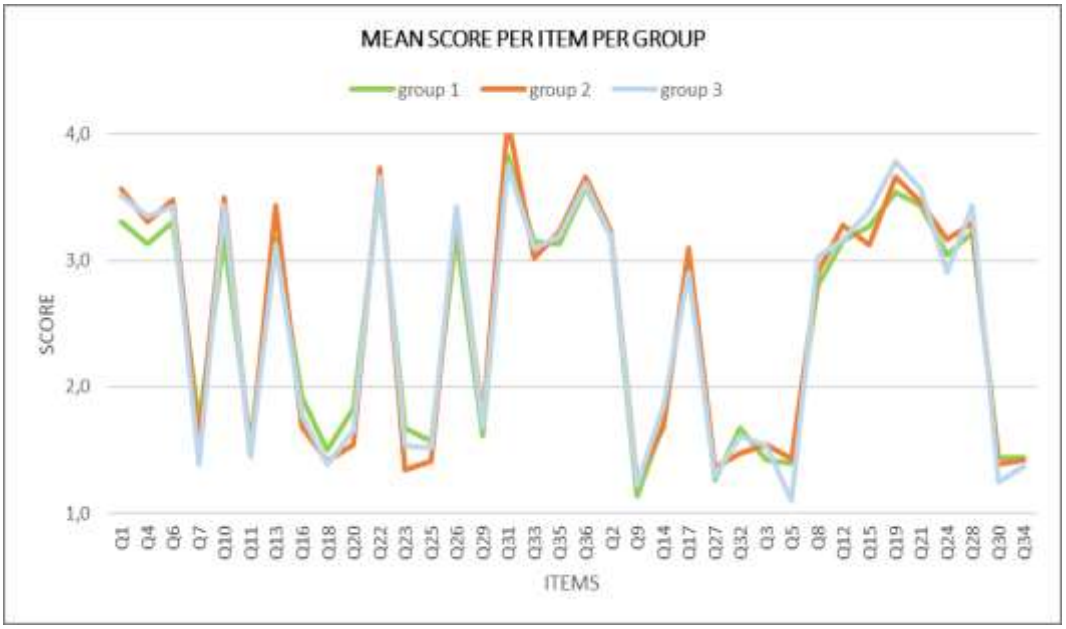

Figure 3. Mean Scores of SCS Statements per Group.

The three groups reported similar ratings for the three subscales of SCS. Figure 4 shows the means for the Membership, Influence, and Fulfillment of Needs Subscales across three groups. All means hovered around 3 on a 4-point Likert Scale with 4 representing the most positive perceptions. These data suggested that students had generally positive perceptions and feelings towards their sense of belonging to the group, their influence on activities and the fulfilment of learning needs.

	Group	*M*	*SD*
Membership Subscale	G1	2.9	0.7
	G2	3.1	0.7
	G3	3.0	0.7
Influence Subscale	G1	2.7	0.5
	G2	2.8	0.6
	G3	2.7	0.6
Fulfilment of Needs Subscale	G1	3.0	0.6
	G2	3.0	0.6
	G3	3.1	0.6

Note. M=Mean; G1: three-student Group; G2: six-student Group; G3: nine-student Group.

Figure 4. Means for G1, G2, and G3 Groups on the Membership, Influence, and Fulfillment of Needs (SCS Subscales).

Also referring to the second questionnaire IMI, the results of analysis also followed a very similar pattern for the three groups (Figure 5).

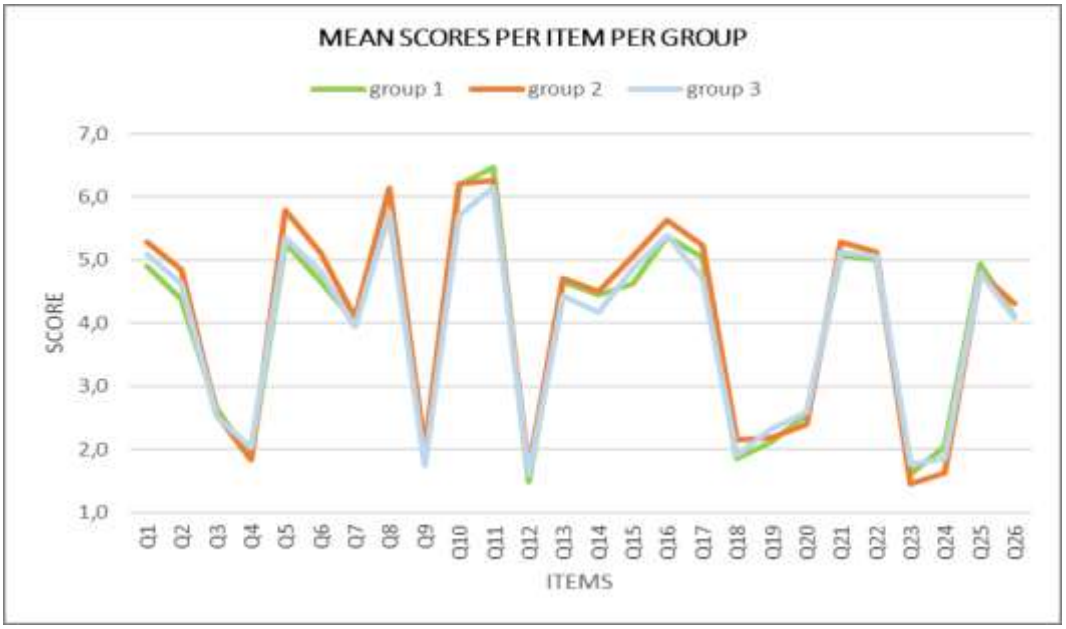

Figure 5. Mean Scores of IMI Statements per Group.

In the same way, the mean scores related to the four IMI subscales (Interest/Enjoyment, Effort/Importance, Perceived Competence, and Relatedness) did not reveal much difference between the three groups (Figure 6). The general high rating of items showed the overall positive feelings of students towards their satisfaction, motivation, engagement, learning achievements and interaction with peers. It was noticeable that the sub-scale "Effort/Importance" was consistently rated high (close to 6) across three groups on a seven-point Likert Scale.

	Group	*M*	*SD*
Interest/Enjoyment **Subscale**	G1	4.7	1.4
	G2	5.0	1.3
	G3	4.8	1.4
Effort/Importance **Subscale**	G1	5.8	1.1
	G2	5.8	1.3
	G3	5.7	1.3
Perceived Competence **Subscale**	G1	4.9	1.0
	G2	5.0	1.3
	G3	4.8	1.4
Relatedness Subscale	G1	4.9	1.4
	G2	5.0	1.3
	G3	4.8	1.6

Note. M=Mean; G1: three-student Group; G2: six-student Group; G3: nine-student Group.

Figure 6. Means for G1, G2, and G3 Groups on the Interest/Enjoyment, Effort/Importance, Perceived Competence, and Relatedness (IMI Subscales).

5.3. Work quality

The analysis of student project scores (Figure 7) indicated that G1 with 3 students in each group obtained the lowest mean score (23.6 out of 30), corresponding to the performance level "Developing". G3 with 9 students in each group got the highest mean score (27.4 out of 30), corresponding to the "Proficient" level and not far from the "Exemplary" level. The quality of students' projects appeared to correlate with the group size: project quality increased as group size got bigger.

Group	M	SD
G1	23.6	3.5
G2	26.1	2.9
G3	27.4	1.9

Note. M=Mean; SD = Standard Deviation; G1: three-student Group; G2: six-student Group; G3: nine-student Group

Figure 7. Student Project Mean Scores for G1, G2, and G3 Groups.

In G1, we observed both the variation and the dispersion of data. Project scores were spread out over a wide range from 17 to 28.5 with occurrences for almost each value. The distribution of scores in G2 was much more consistent. Projects were mostly rated between 24 and 28.5, with only two cases of higher distances from the mean. Then, marks appeared a little scattered but mostly distributed on a good level of performance, which suggested that most students performed uniformly but a few members received scores considerably far from the average. For G3, the project scores tended to be close to the mean of the data set. Only one student went out to the range of 26 to 29, which indicates the general high-performance level of the group.

6. Discussion and Conclusion

The present study examined how group size may impact students' participation, perceptions, and work quality in a peer assessment activity. The first research questions asked if and how group size may influence student participation when working in different-sized peer assessment groups. Previous studies suggested that larger group size may decrease group member's opportunity to participate in collaborative learning environments, which may negatively impact members' contributions to the group (Jungjoo, 2013). Furthermore, other studies suggest that moderately smaller group sizes such as five (Fay et al., 2000) or seven (Blenko et al., 2010) may be effective in facilitating productive group work. Additionally, Hung, Chen, and Samuelson (2016) reported that small peer review groups were more effective in engaging students and encouraging student participation. Our study seems to endorse these findings by showing that student participation in groups inversely corresponds to the group size. Students in G1 reported the highest level of participation as judged by the average posts by each student. Furthermore, observing the distribution of students' participation for each group size (Figure 1), we identified an interesting pattern that seems to suggest that group size may influence students' participation depending on the task. There might be multiple explanations for this interesting discovery. One possibility could be, when students in small groups (e.g. G1) were asked to discuss their projects, they had to share their thoughts more in depth and be actively engaged as they really counted on each other in completing the task. While in larger groups students could benefit from a richer set of ideas shared by other group members without the need for active participation. We call for future research to further explore this issue and determine how group size and the nature of tasks may be intertwined to impact student participation in groups. Given that we measured students' participation by extracting the number of posts using a Moodle tool, future studies could consider also the quality of posts in forums, in order to validate our results.

The second research questions asked if and how group size impacted student perceptions. Students' responses to both SCS and IMI questionnaires did not show much difference across three groups. All groups reported similar scores for each item on the questionnaires. Overall, it suggested the positivity of students' perceptions of the activities, which confirmed previous reported benefits of collaborative group work in a peer review process.

Focusing on the effects of group size on work quality, literature has not yet clearly identified the predictors of work quality in team collaboration and also presented conflicting results. Some researchers reported that learners work more effectively when working in smaller teams (Bennett et al. 2012; Li et al., 2019), while others suggested that students in large teams outperform those in smaller ones (Valacich et al., 1992). In terms of work quality, our findings seemed to reaffirm the value of working in bigger groups. The findings of our study revealed that smaller groups had varied levels of performance on their projects. Larger groups obtained better scores, as students in these groups built consistently higher quality projects with richer and more thoughtful content. However, if we consider students' participation, we could notice that smaller groups perform better than larger groups.

In light of the results obtained from triangulating the three concurrent phases of analysis, it seems reasonable to assume that in a peer review process, group size may have an effect on learner participation and work quality, whereas it does not influence students' perceptions. Our findings suggested that group size was directly proportional with work quality and indirectly proportional with learners' participation. Groups of three members (G1) showed a high participation rate during the overall activities, but their projects were

rated the lowest quality. For the groups consisting of nine students (G3), we observed the opposite trend: students showed the lowest participation rate and the highest quality of projects. Groups of six students (G2) reported an intermediate level of participation and work quality, with a tendency to be closer to the level of the nine-student groups. If we assume that group size encourages the amount of shared knowledge and skills available during the learning process, we could infer that group size may have some effects on the components of a given learning context. Based on this assumption, the interpretation of our findings also suggested that the kind and the nature of the task may influence the group size effect. Actually, in a context where the assignment includes peer review as a crucial instrument to achieve the learning goals, larger groups could benefit from a larger amount of feedback provided from or to peers, which may promote better student performance.

Reference list

Balboni, G., Perrucci, V., Cacciamani, S., & Zumbo, B. D. (2018). Development of a scale of sense-of community in university online courses. *Distance Education, 39*(3), 317–333.

Bennett, P. N., Parker, S., & Smigiel, H. (2012). Paired peer review of university classroom teaching in a school of nursing and midwifery. *Nurse Education Today*, *32*(6), 665–668.

Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and Learning* (pp. 9-26). Los Angeles, CA: Sage.

Blenko, M. W., Mankins, M. C., & Rogers, P. (2010). *Decide & deliver: 5 steps to breakthrough performance in your organization.* Boston: Harvard Business Press.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–12.

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, *11*(6), 481–486.

Grion, V., & Serbati, A. (2019). *Valutazione sostenibile e feedback nei contesti universitari. Prospettive emergenti, ricerche e pratiche*. Lecce: Pensa Multimedia.

Grion, V., Serbati, A., Tino, C, & Nicol, D. (2017). Ripensare la teoria della valutazione e dell'apprendimento all'università: un modello per implementare pratiche di peer review. *Giornale Italiano della Ricerca Educativa*, *19*, 209–226.

Hashmi, N. (2017). *The more the merrier? Understanding the effect of group size on collective intelligence*. Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology, Massachessetts, USA. https://dspace.mit.edu/handle/1721.1/113957 (ver. 23.03.2020).

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hung, Y. J., Chen, S. C., & Samuelson, B. L. (2016). Elementary EFL students' practice of peer assessment of oral classroom activities. *TESOL Journal*, *7*(2), 449–468.

Jungjoo, K. (2013). Influence of group size on students' participation in online discussion forum. *Computer & Education*, *62*, 123–129.

Li, L., Gao, F., & Guo, S. (2019). The effects of social messaging on students' learning and intrinsic motivation in peer assessment. *Journal of Computer Assisted Learning*. 1–10. https://doi.org/10.1111/jcal.12409 (ver. 23.03.2020).

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122.

Odo, C., Masthoff, J., & Beacham, N. (2019). Group formation for collaborative learning. In S. Isotani et al. (Eds.), *Artificial intelligence in education* (pp. 206-212). AIED 2019. Lecture Notes in Computer Science (vol. 11626,). Cham: Springer.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*(3), 450–461.

Sadler, D. R. (1989). Formative assessment: Revisiting the territory. *Assessment in Education*, *5*(1), 77–84.

Sambell, K., Brown, S., & Race, P. (2019). Assessment as a locus for engagement: priorities and practicalities. *Giornale Italiano della Ricerca Educativa*, *12*, 45–61.

Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. London: Routledge.

Serbati, A., Grion, V., & Fanti, M. (2019). Caratteristiche del peer feedback e giudizio valutativo in un corso universitario blended. *Giornale Italiano della Ricerca Educativa, 12,* 115–137.

Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods Research, 1*(1), 3–7.

Valacich, J., Dennis, A. R., & Nunamaker, J. F. (1992). Group size and anonymity effects on computer-mediated idea generation. *Small Group Research*, *23*(1), 49–73.