Form@re
Open Journal per la formazione in rete

articoli

Assessing is not a joke. Alternative assessment practices in higher education

Valutare non è uno scherzo. Le pratiche alternative di valutazione nel contesto universitario

Margherita Di Stasio[a], Maria Ranieri[b], Isabella Bruni[c,1]

[a] *Indire, m.distasio@indire.it*
[b] *Università degli Studi di Firenze, maria.ranieri@unifi.it*
[c] *Università degli Studi di Firenze, isabella.bruni@unifi.it*

Abstract

In the context of a general renewal of teaching and learning practices in higher education, specific attention is paid to formative assessment and to the exploration of alternative assessment techniques. This paper presents a study carried out at the University of Florence involving around 200 first-year students in a course where such assessment techniques were adopted. The results showed mixed evidence. Regarding peer-assessment validity, statistical analysis gave a fair level of accordance. As for students' perception, their concerns about peer- and self-assessment emerged, especially because they understand assessment as a summative and not as a formative practice. Implications are discussed, paving the way for possible future research.

Keywords: formative assessment; peer-assessment; self-assessment; reliability; students' perceptions.

Sintesi

Nel quadro di un rinnovamento generale delle pratiche di insegnamento e apprendimento universitarie, un'attenzione particolare è rivolta alla valutazione in ottica formativa e all'esplorazione di tecniche di valutazione alternative. Questo articolo presenta uno studio condotto presso l'Università di Firenze che ha coinvolto circa 200 studenti del primo anno in un corso dove sono state utilizzate tali tecniche di valutazione. I risultati mostrano evidenze miste. L'analisi statistica ha fornito un giusto livello di accordo in merito alla validità della valutazione tra pari. Dalle percezioni degli studenti emergono invece resistenze verso la valutazione tra pari e l'autovalutazione, soprattutto perché gli studenti interpretano la valutazione come pratica sommativa e non come azione formativa. Le implicazioni di questi risultati vengono discusse nell'ottica di guidare possibili ricerche future.

Parole chiave: valutazione formativa; valutazione tra pari; autovalutazione; affidabilità; percezioni degli studenti.

[1] Although this paper has been jointly conceived, Isabella Bruni wrote sections 1, 2, 3.1 and 3.2; Margherita Di Stasio developed sections 3.3 and 4; Maria Ranieri elaborated sections 5 and 6.

1. Introduction

Nowadays, there is an increasing consensus among scholars and policy makers that support should be given to the higher education system to renew and reshape teaching and evaluation practices (McAleese et al., 2013). Scientific and technological progress as well as globalisation and labour market changes have determined a new scenario requiring a rethinking of the very role of universities. In particular, educational institutions are facing the needs of increased and differentiated users, as well as knowledge that is constantly changing and demanding new professional skills. In this context, the quality and innovation of teaching and learning processes have become a central research topic: a large corpus of studies have analysed and questioned more traditional approaches, paving the way for rethinking education in a student-centred perspective.

This paper focuses on alternative assessment techniques such as peer-assessment and self-assessment (Boud & Dochy, 2010; Gozuyesil & Tanriseven, 2017) in order to both understand how students perceive changes in assessment practices and to what extent these techniques may be considered valid. In the first part of the paper, the main findings of international research on peer- and self-assessment are described, while the second section presents a study carried out in 2017-2018 at the University of Florence, with the aim of investigating students' reactions to peer-assessment as well as peer-assessment potential in higher education.

2. Literature review

In the renewed perspective of active learning, evaluation becomes a central element with a shift from a summative to a formative approach, which entails that assessment should be embedded in teaching activities to foster learning processes (Brown, 2005; Notti, 2017; Sambell, McDowell, & Montgomery, 2013). Alternative assessment techniques and constant feedback could accompany learning and become a lever for improvement, reflection and self-regulation, while students develop their own assessment capacity (Grion, Serbati, Tino, & Nicol, 2018; Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006). Differently from standardised assessment techniques, non-traditional approaches are based on realistic contexts and performance-oriented, and teach students to evaluate themselves and their colleagues, taking into consideration different learning styles. In fact, an active involvement of students in the evaluation process seems to correspond to an equally active role in the management of learning processes (Boud, Cohen, & Sampson, 2001; Falchinov, 1995; Nicol, Thomson, & Breslin, 2014; Pereira, Flores, & Niklasson, 2016). Furthermore, these evaluation practices can fulfil a variety of needs both at the management level (as in the case of large-size classes) and on the didactic and training one (locus of control in learning processes) (Felisatti & Giampaolo, 2014).

In non-traditional approaches, students play an active role as assessors and have the possibility of working effectively in the process, through an autonomous construction of meaning, which is a key competence for lifelong learning (Boud et al., 2001; Nicol et al., 2014; Sambel & McDowell, 1997). In particular, we can define self-assessment as the involvement of learners in making judgements about their own learning (Boud & Falchikov, 1989): it is mainly a formative practice, aimed at fostering reflection on one's own learning processes and results (Boud, 1995). On the other hand, peer-assessment is defined as "an arrangement for learners to consider and specify the level, value or quality of a product or performance of other equal-status learners" (Topping, 2009, p. 20). It

happens generally with the use of instruments or checklists which have been previously designed by the teacher or discussed and constructed by the group itself. Peer-assessment may be formative or summative, and can take different forms, depending on the way in which assessment is made (Boud et al., 2001; Liu & Carless, 2006; Nicol et al., 2014; Topping, 2009). In particular, we can distinguish between i) peer review with students providing qualitative feedback to their colleagues, ii) peer rating, when students rate their peers using a given set of performance or personal characteristics scale and iii) peer grading with students assigning grades for summative assessment.

Assessment practices are widely investigated, specifically with reference to issues such as accuracy, validity and reliability (Cho, Schunn, & Wilson, 2006; Falchinov & Boud, 1989; Falchikov & Goldfinch, 2000). However, researchers have also pointed out that peer-assessment validity and reliability need to be further explored: larger scale studies are still scarce, while a common metric for experiments has not yet been identified (Bouzidi & Jaillet, 2009; Cho et al., 2006).

As far as self-assessment is concerned, the tendency to overrating or underrating is common with a value that is inversely proportional to the learning outcomes: weaker students tend to overrate themselves, while better prepared students tend to underrate themselves (Boud & Falchinov, 1989). Nevertheless, self-assessment accuracy improves over time, especially when teachers give feedback on students' self-assessment. As for peer-assessment, reliability can be measured by the similarity between the marks given by peers, while validity is a variable that can be measured by the similarity between the marks attributed by peers and those given by teachers. Generally speaking, peer-assessment requires students to be fair and accurate with the judgements they make regarding their peers: some studies showed that reliability could be affected by friendship or by a propensity to uniformity. Many studies compared peers' and teachers' assessment, with acceptable results (Bouzidi & Jaillet, 2009; Calabrese, 2018; Cecchinato & Foschi, 2018; Cho et al., 2006). Falchinov and Goldfinch's meta-analysis (2000) focused on validity and tested the impact of different context variables identified in primary studies such as subject area, how the assessment is carried out and the nature of the criteria used, the number of peers and faculty involved. The following variables were identified as influential:

- nature of the assessment task: marking several individual dimensions is less valid than giving a global judgement based on criteria. Furthermore, assessment provided by peers is closer to the teacher's when students evaluate academic products rather than professional ones;
- status of criteria: the agreement between students and teachers is higher when students have been involved in selection of criteria;
- level of course: peer-assessment in advanced level courses is more valid than peer-assessment in introductory ones.

Regarding the number of peers, the meta-analysis (Falchinov & Goldfinch, 2000) concluded that there is no evidence to support the idea that multiple ratings are more valid than single ones. As for the subject area, no clear difference in validity was found.

Other studies investigated students' perceptions of peer-assessment (Hanrahan & Isaacs, 2001; Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006; Planas Lladó et al., 2014). In their analysis of students' perceptions of peer- and self-assessment, Hanrahan and Isaacs (2001) identified benefits and drawbacks organised in eight general dimensions: difficulty, improved understanding of marking, discomfort, productive, problems with implementation, read others work, developed empathy, and motivation. Similar findings

also emerged in a study by Lindblom-Ylänne, Pihlajamäki, and Kotkas (2006), where students expressed positive feelings about peer-assessment together with the difficulty of being critical of their peers. Finally, in their research on peer-assessment in higher education, Planas Lladó and colleagues (2014) found students underlining the responsibility coming with peer evaluation processes and their distrust in the abilities of their peers to peer-assess.

3. Research Methods

3.1. Context and sample

In this paper we report the results of a study exploring the use of alternative assessment techniques within the course of New Technologies and Education at the University of Florence in 2017-18. The course included a two-month activity on fake news to make students able to critically analyse and discuss online news. The activity named Teaching and learning about fake news, was structured in four phases:

1. introduction to definition and typologies of fake news including examples (EAVI [European Association for Viewers Interests], 2017);
2. elaboration of an individual essay focused on an analysis of some fake news;
3. essay review based on self-evaluation and classroom discussion on fake news cases in order to clarify doubts and share interpretations;
4. delivery of the revised essay and online peer-grading.

The activities were delivered in a blended mode through the use of Moodle as a teaching and learning platform: out of 216 students enrolled in the course, 172 delivered their essays while 167 participated in the peer-assessment exercise. The large majority of students were female (159) in their first-year of academic career, which means aged between 19 and 20 years.

With regard to the self-assessment task, students were provided with a checklist aimed at guiding them in the review of their work. The checklist consisted of a battery of questions such as: “Did you recognise the type of fake news, arguing your choice? Did you use reliable information sources, quoting them in the text? Were you able to correctly identify the author(s) and describe his/her intentions? Did you describe the message and the type of reactions it seeks to elicit in the audience? Did you explain the level of impact that these fake news may have?”

Coming to the peer-assessment phase, each student received two anonymised essays and evaluated them through the use of a rubric including four criteria (critical capacity, completeness, clarity, formal correctness) and four performance levels (Figure 1). The rubric was shared with the students during a lecture, although it was not discussed in depth, due to the high number of students and the limited time. The peer-grading activity had only a formative purpose. This is the reason why we adopted an evaluation tool based on individual criteria rather than on a global judgement: the purpose was to bring students to better reflect on each individual aspect of peers’ works through appropriate guidance. The main purpose being formative, the evaluation given by peers did not affect the final grade. In the meantime, essays were also evaluated by the teacher in collaboration with two teaching assistants. At the end of this phase, students received two grades, one for their

essays, and another for the quality of their assessment as peers: the closer the students' assessment was to that of teachers, the better the evaluation was assessed.

Criteria	Performance levels			
	Excellent (3 points)	Good (2 points)	Sufficient (1 point)	Poor (0 point)
Critical capacity	The topic is analysed in a timely manner, with in-depth analyses, grounded and personal reflections.	The analysis is fitting with the task and presents some interesting personal insights.	The topic is analysed in a correct way but with no personal insight on any aspects.	The essay is limited to reporting facts, sources and information without giving a summary or analysis of any kind.
Comple-teness	The subject is treated in all its aspects, with a wealth of perspectives, plurality of sources, relevant and documented examples.	The discussion is ample and touches on a variety of aspects and themes, but they are not clearly documented.	The subject is dealt with in a complete manner, but it has not been thoroughly examined.	The subject is treated in a superficial and hasty way with obvious gaps and incorrect information.
Clarity	The essay is clearly structured, with an explicit guiding thread; shows internal coherence and consistence with the topic.	The essay is fluent, but the internal structure is not explained or consistent.	The essay, while not showing inconsistencies or contradictions, does not offer a linear treatment and is difficult to read.	The essay shows internal inconsistencies between the various topics and the specific theme.
Formal correctness	The essay is clearly and correctly written, without syntactical, grammatical or typographical mistakes.	The essay is written in a simple and correct way.	The essay presents some unclear sentences, some mistakes and typos.	The essay is formally incorrect with many mistakes related to morphology, syntax and grammar.

Figure 1. Peer-assessment rubric.

3.2. Research Questions

In this paper, we focus on peer-assessment with the double aims of investigating its level of validity, meant as the level of agreement between students' and teachers' grades (Cho et al., 2006; Falchikov et al., 2000), and exploring students' perceptions and reactions to alternative assessment techniques.

We identify two research questions:

- RQ1. Is peer-assessment a valid technique? In other words, to what extent are students' assessments of their peers similar to the marks attributed by teachers?
- RQ2. How do students perceive alternative assessment techniques and their role in these types of evaluation process?

3.3. Data Collection and Analysis

To investigate peer-assessment accuracy and validity, two datasets were constructed: one containing the grades given by the teachers, which was considered as an expert evaluation and one including the grades given by the peers. The validity of teachers' evaluations was based on the multiple exercises that teachers made to reach consensus on the final grades of a sample of five essays, while the validity of students' assessment was measured at level

of individual criterion comparing students' grades to teachers' ones using Cohen's kappa coefficient (κ). Kappa Coefficient, a statistic that is used both in medical and social studies to measure inter-rater reliability for qualitative variables, is based on values ranging from 0, in case of agreement by chance, to 1, for perfect inter-observer agreement. These values are usually analysed through the Strength of Agreement benchmarks proposed by Landis and Koch (1977) in order to discuss Kappa Coefficients, as reported in Figure 2.

Kappa Statistic	Strength of Agreement
< 0.00	POOR
00-0.20	SLIGHT
0.21-0.40	FAIR
0.41-0.60	MODERATE
0.61-0.80	SUBSTANTIAL
0.81-1.00	ALMOST PERFECT

Figure 2. Benchmarks for the strength of agreement (Landis & Koch, 1977, p. 165).

To explore students' perception, a questionnaire was administered through the e-learning platform at the end of the activity. 126 students anonymously responded. The survey included closed questions to evaluate students' level of satisfaction, their perceptions of learning gains related to the topic and of the effectiveness of the teaching approach, including the use of self- and peer-assessment. Furthermore, students' suggestions for improvement and other comments were obtained through two open-ended questions. Quantitative variables were analysed using descriptive statistics procedures, while open answers were coded in parallel by the authors in order to identify students' perceptions.

4. Results

4.1. RQ1. Is peer-assessment a valid technique? In other words, to what extent are students' assessments of their peers similar to the marks attributed by teachers?

The statistical analysis of Cohen's kappa suggests that the concordance between teachers' grades and grades of students as peer assessors is limited: indeed, the result relating to the Strength of inter-observer agreement is only FAIR (Figure 5).

Rubric criterion	Kappa	Strength of Agreement
Critical capacity	0.31	FAIR
Completeness	0.33	FAIR
Clarity	0.33	FAIR
Formal correctness	0.30	FAIR

Figure 5. Strength of agreement for individual rubric criteria.

To go deeper into the analysis of differences between teachers' and students' grades, descriptive statistics are provided below that show how teachers and students graded the same essays per criterion.

Regarding the grade attributed to the criterion *Critical capacity* (Figure 6), both teachers and students mostly chose Good. Low was a marginal choice for both. The difference

emerges in the attribution of Sufficient, used by the teachers less than the students (12% vs 30%), and Excellent, used by the teachers more than the students (31% vs 17%).

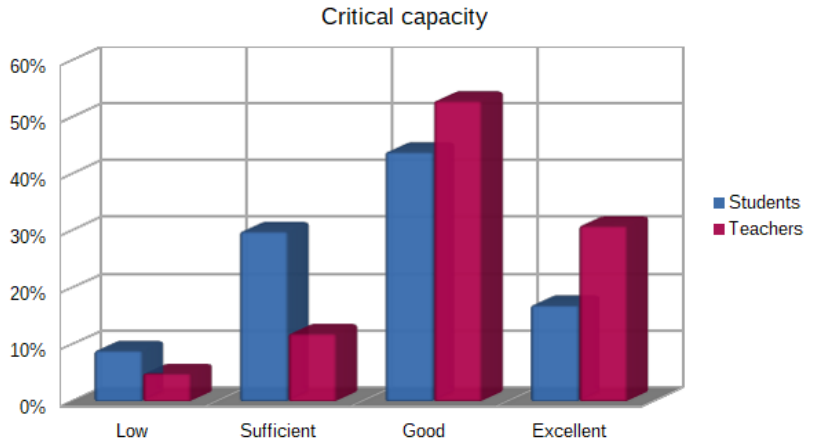

Figure 6. Frequency distribution of students and teachers grades for criteria critical capacity.

Concerning the criterion *Completeness* (Figure 7), teachers' and students' grades are mostly similar in case of Low and Excellent, while Sufficient was attributed to different extents (22% of the students vs 11% of the teachers) as well as Good (31% of the students vs 46% of the teachers).

Figure 7. Frequency distribution of students and teachers grades for criteria completeness.

Moving to *Clarity* (Figure 8) teachers and students converge on attributing positive scores, although with some distinctions: in many cases, peers go for Good (31%) and Excellent (42%) with a total of 73% of positive grades; similarly, teachers' positive grades reached 88%, but overturning the levels between Good (60%) and Excellent (28%).

Figure 8. Frequency distribution of students and teachers grades for criteria clarity.

With respect to *Formal correctness* (Figure 9), grade orientation is similar to Critical capacity: both teachers and students mainly chose Good (44% of the students, 53% of the teachers). Low was a marginal choice for both (9% of the students, 5% of the teachers) while an important difference emerges in the attribution of Sufficient with teachers choosing it less than students (18% vs 36%), and Excellent used by students less than teachers (16% vs 33%).

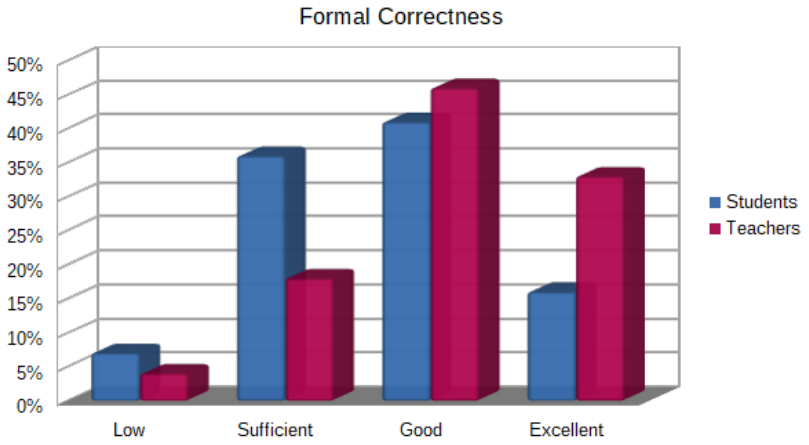

Figure 9. Frequency distribution of students and teachers grades for criteria formal correctness.

4.2. RQ2. How do students perceive peer grading?

Not surprisingly, students found the activities of self-assessment and peer grading unusual compared to their common academic experiences. Data gathered through the questionnaire allow us to identify the lights and shadows of peer-assessment in higher education from the students' perspective.

When asked to express their opinions about the proposed activities (in terms of difficulty, interest and pleasantness), students indicated that peer grading is challenging with 44% finding it difficult, while only 29% interesting and 33% pleasant (Figure 10).

Students' perceptions of alternative assessment techniques were explored in depth through specific questions whose answers are summarised in Figure 11.

Indicate the activities that were for you most..	pleasant	interesting	difficult
Lesson on fake news	54%	55%	2%
Fake news examples	39%	41%	2%
Paper writing	50%	47%	22%
Self-evaluation	11%	12%	25%
Paper review	18%	12%	20%
Peer-assessment	33%	29%	44%

Figure 10. Students' satisfaction with course activities (N=126).

More specifically, in terms of learning, students found the self-assessment task useful to understand their mistakes and how to improve their essays. Concerning peer-assessment, their perceptions were more neutral: students did not feel they had been more active nor that they had learned more. Nevertheless, engaging with alternative evaluation practices seems to have enabled students to take a step forward on metacognition and increased awareness of cognitive processes. Indeed, almost all the students claimed that:

- experiencing these approaches brought them to reflect on the evaluation meaning: on average, almost 90% agreed (including 59% agree and 29% strongly agree);
- experiencing these approaches brought them to reflect on the evaluation impact on their professional development (including 60% agree and 31% strongly agree).

Looking at the questions about peer-grading as a stand-alone activity, a more complex picture emerges. On one hand, more than half of the students perceived the benefits of peer grading for learning: 47% of the students agreed and 12% strongly agreed that peer grading allowed them to improve their learning results. This quantitative data is consistent with some qualitative data emerging from the open-ended questions where students defined peer grading as *interesting and useful* or *useful and constructive*. Furthermore, 34% of students were neutral on this issue. On the other hand, faced with the statement "Peer grading activity made me feel like an added value within the course", 44% of students neither agreed nor disagreed.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Self-assessment was useful to understand my mistakes	0	4.5%	14.5%	**61%**	**20%**
Self-assessment was useful to understand how to improve my outcomes	0	3%	14%	**63%**	**20%**
Peer grading activity made me feel like an added value within the course	2%	15%	**44%**	**28%**	11%
Peer grading activity allowed me to improve learning	1%	6%	**34%**	**47%**	12%
Experiencing these approaches brought me to reflect on the evaluation meaning	0%	3%	9%	**59%**	**29%**
Experiencing these approaches brought me to reflect on the evaluation impact on professional development.	0%	3%	6%	**60%**	**31%**

Figure 11. Students' perception of the impact of peer-assessment and peer grade on learning outcomes (N = 126).

Further qualitative comments help identify the reasons why students sometimes felt uncomfortable with peer grading. In some cases, this perception of discomfort seems to be linked to the awareness of not being prepared, suggesting an emerging educational need. For example, a student in an open-ended question claimed: "Personally, I found peer evaluation very challenging since it gives me such a big role of assessing my peers without having yet developed appropriate evaluation skills or having been prepared to take on such a kind of responsibility". Another student observed: "Peer evaluation came out very challenging, although it is useful to realise how difficult it is to provide objective evaluations". In others cases, although limited, students underlined difficulties in understanding the rubric and its use. One student claimed that: "the criteria levels were very similar; and I found both my colleagues' works halfway between one level and another". Another student found criteria "too narrow for someone facing this task for the first time".

5. Discussion

Overall, our study suggests that students as assessors are more severe than their teachers. Their grades were, indeed, generally lower than those given by the teachers, thus not aligning with teachers' evaluations. This result raises doubts about the validity of students' grading, leading us to reflect on the reasons for this disalignment. First of all, this could be explained by the uncertainty that students felt in the understanding of the criteria included in the rubric. Actually, criteria were not directly negotiated with students during the lesson, but just shown and shared to limit misunderstanding. However, the evaluation process is more than avoiding misunderstanding: it has to do with analysis and interpretation, that is two cognitive activities that need recursive and tuning processes of meaning construction. Further studies are necessary to go deeper in the understanding of the impact of this factor and future research should be developed to clarify how negotiation processes of evaluation criteria could be carried out in the context of large-size classes.

Another explanation could be linked to students' perceptions about peer-assessment. To this purpose, our study found controversial findings. On one hand, students recognised peer assessment as an interesting activity, improving reflection processes and increasing their awareness about the role and the implications of evaluation for learning. This is consistent with several studies in the literature which have shown how peer assessment impacts on metacognition and on the development of high order cognitive skills (Grion et al., 2018; Hattie & Timperley, 2007; Nicol & MacFarlane-Dick, 2006). Furthermore, being on the side of the evaluator allowed them to understand the complexity associated with the act of assessing. This is particularly relevant, since these students are future educators and are likely to have to deal with evaluation in their future professional lives. From this perspective, the experience had implications for learning not only in terms of cognitive processes (better understanding of concepts) but also in terms of professional development. Indeed, through putting themselves into the shoes of an assessor, they concretely lived the experience of evaluation from a totally different point of view: no longer the one of tutees but that of educators.

At the same time, students perceived how challenging the evaluation process is entailing not only revising or commenting but also taking the responsibility for grading. As we have seen, students can be involved in the assessment processes at different stages: from peer reviewing to peer grading (Boud et al., 2001; Liu & Carless, 2006; Nicol et al., 2014; Topping, 2009). In our study we found that while students felt they gained benefits from

peer reviewing, they found themselves inadequate to rate and grade their peers. The exercise required them to go out from their comfort zone and take a role they are not used to playing – at least within the formal setting of higher education. The life of the classroom is regulated by implicit norms determined by repeated behaviours which generate expectations and roles: these are called educational routines. The very basic routine is that of question and answer, where the teacher generally asks and the students answer. The asymmetry between the role of the students and that of the teacher mostly reflects the relation of power underpinning the school mechanism of questioning and answering. Evaluation processes are deeply rooted in these mechanisms and, although to different degrees of awareness, students do know them. Therefore, when asked to assess, they had to break the implicit agreement and overturn the traditional rules of the institutional grammar of school. In this sense, we can say that, when asked to grade their peers, they had to leave their comfort zone, that is their traditional roles and routines, and posit themselves in a new perspective. This feeling of inadequacy or discomfort is consistent with other studies on students' perceptions on peer-assessment (Hanrahan & Isaacs, 2001; Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006; Planas Lladó et al., 2013): one can conclude that a general feeling of distrust towards their own assessing capacities and those of their peers negatively impacted on their evaluation, with implications for accuracy and validity. An important consequence is that for peer assessment to be more valid, a cultural change in the way how teachers' and students' roles are socially conceived is needed. Evaluation has to do not only with measures but also with the cultural meanings that are associated with it. Transforming this culture both in the perspective of the students and the teachers, is crucial to increase the meaningfulness of assessing, rating and grading.

6. Conclusion

The theme of evaluation in the academic context is gaining momentum in the current debate on the quality of university teaching (Grion et al., 2018). Scholars agreed on the importance of self- and peer assessment in higher education (Nicol et al., 2014; Sambell et al., 2013), since they allow one to tackle different challenges: from management issues (large-size classes) to the pedagogical ones (locus of control in learning processes) (Felisatti & Giampaolo, 2014). Nevertheless, the results presented in this study highlight how the active involvement of students in the evaluation processes shows lights and shadows. In fact, while self-assessment was widely appreciated by students in terms of effectiveness on learning processes, peer assessment put them in difficulty when they were asked to leave their own comfort zone to perform an activity that they perceived as an overloaded task of responsibility. Although these results cannot be generalised due to the limits of the sample and those of procedures adopted, including the lack of data about self-assessment, they suggest that the adoption of innovative evaluation practices such as peer assessment requires a cultural change that affects not only the teachers, but also the students. The latter not only expect the teacher to evaluate them according to formal traditional roles, but they also find it challenging being assessors or evaluators. From this point of view, the institutional grammar which more or less implicitly informs educational practices and expectations seems to hinder the assumption of different postures. If we wish to encourage the entry of new evaluation practices, this culture should evolve towards new visions of the roles and relationships in the academic context.

Reference list

Boud, D. (1995). *Enhancing Learning through Self-assessment*. London, Philadelphia, PA: Kogan Page.

Boud, D., & Falchinov, N. (1989). Quantitative studies of self-assessment in higher education: a critical analysis of findings. *Higher Education, 18*, 529–549.

Boud, D., Cohen, R., & Sampson, J. (2001). *Peer learning in higher education: Learning from and with each other*. London: Routledge.

Boud, D., & Dochy, F. (2010). *Assessment 2020. Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted?. *Educational Technology & Society, 12*(4), 257–268.

Brown, S. (2005). Assessment for learning. *Learning and teaching in higher education*, *1*, 81–89.

Calabrese, M. (2018). Affidabilità, criticità, valore pedagogico della valutazione tra pari in modalità sommativa con l'attività workshop di Moodle. *Atti del Moodle Moot Italia 2018*.

Cecchinato, G., & Foschi, L. C. (2018). Coinvolgere gli studenti nell'insegnamento: analisi di un percorso di innovazione didattica all'Università. *Form@re - Open Journal per la formazione in rete, 18*(1), 97–110.

Cho, K., Schunn, C., & Wilson, R. (2006). Validity and Reliability of Scaffolded Peer Assessment of Writing From Instructor and Student Perspectives. *Journal of Educational Psychology, 98*(4), 891–901.

EAVI. European Association for Viewers Interests (2017). *Beyond Fake News. 10 Types of misleading news*. https://www.rcmediafreedom.eu/Multimedia/Infographics/Beyond-Fake-News-10-Types-of-Misleading-News (ver. 10.12.2019).

Falchinov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International, 32*, 175–187.

Falchinov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: a meta-analysis. *Review of Educational Research, 59*(4), 395–430.

Falchinov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Felisatti, E., & Giampaolo, M. (2014). Personalizzare l'apprendimento nel contesto universitario. *Formazione & Insegnamento, 12*(4), 311–327.

Gozuyesil, E., & Tanriseven, I. (2017). A Meta-Analysis of the effectiveness of alternative assessment techniques. *Eurasian Journal of Educational Research, 70*, 37–56.

Grion, V., Serbati, A., Tino, C., & Nicol, D. (2018). Ripensare la teoria della valutazione e dell'apprendimento all'università: un modello per implementare pratiche di peer review. *Italian journal of Educational Research, 19*, 209–226.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: the student's views. *Higher Education Research & Development*, *20*(1), 53–69.

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

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

Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-peer-and teacher assessment of student essays. *Active learning in Higher Education*, *7*(1), 51–62.

Liu, N., & Carless, D. (2006). Peer feedback: the learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290.

McAleese, M., Bladh, A., Berger, V., Bode, C., Muelhfeit, J., Petrin, T., … & Tsoukalis, L. (2013). *Report to the European Commission on 'Improving the quality of teaching and learning in Europe's higher education institutions'*. Brussels: Publications Office of the European Union.

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.

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218.

Notti, A. M. (2017). *La funzione educativa della valutazione. Teorie e pratiche della valutazione educativa*. Lecce: Pensa Multimedia

Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: a review of research in assessment and evaluation in higher education. *Assessment & Evaluation in Higher Education, 41*(7), 1008–1032.

Planas Lladó, A., Feliu Soley, L., Fraguell Sansbelló, R. M., Arbat Pujolras, G., Pujol Planella, J., Roura-Pascual, … & Montoro Moreno, L. (2014). Student perceptions of peer assessment: an interdisciplinary study. *Assessment & Evaluation in Higher Education*, *39*(5), 592–610.

Sambell, K., & McDowell, L. (1997). The value of self- and peer assessment to the developing lifelong learner. In C. Rust (Ed.), *Improving Student Learning-improving students as learners* (pp. 56-66). Oxford: Oxford Centre for Staff and Learning Development.

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

Topping, K. (2009). Peer assessment. *Theory into Practice, 48*(1), 20–27.

Tonelli, D., Grion, V., & Serbati, A. (2018). L'efficace interazione fra valutazione e tecnologie: evidenze da una rassegna sistematica della letteratura. *Italian Journal of Educational Technology*, *26*(3), 6–23.