

Application of peer review in a university course: are students good reviewers?

Maria Renza Guelfi^a, Andreas R. Formiconi^b, Marta Vannucci^c,
Lorenzo Tofani^a, Jonida Shtylla^a, Marco Masoni^{a1}

^aUniversity of Florence, Dept. of Experimental and Clinical Medicine – Florence (Italy)

^bUniversity of Florence, Dept. of Statistics, Informatics and applications – Florence (Italy)

^cCareggi University Hospital, Dept. of Maternity and Child Health – Florence (Italy)

(submitted: 9/10/2020; accepted: 2/11/2020; published: 15/10/2021)

Abstract

Peer review can be used as a teaching methodology to improve students' learning and critical thinking. However, teachers have many concerns about the reliability and validity of students' grading.

The paper describes the application of peer review as a teaching strategy to the large course of Biomedical Informatics in the School of Medicine at the University of Florence. The aim of the study was twofold: (I) assessing the validity of students' reviews, calculating the correlation between students' assigned score and instructor's assigned score; (II) assessing the validity of student's self-evaluation, calculating the correlation between student's assigned score and teacher's assigned score. To this aim a statistical analysis was performed.

The results showed a moderate concordance between the marks assigned by peers and those assigned by the instructor. Nevertheless, the comparison between the teacher median and the peer-review median shows a minimal difference that has almost no effect on changing the final grade. Instead, there was poor concordance between the marks attributed by the instructor and those relating to the student's self-evaluation. Even if further studies are needed, the promising results can begin to dispel teachers' concerns about students' grading skills that prevent the application of peer review. On such basis, the use of peer review systems can streamline the application of peer review in classes with a high number of students reducing the workload on the teacher.

KEYWORDS: Peer assessment; Peer review, Self assessment, Higher Education, Medicine, Peer Assessment Validity

DOI

<https://doi.org/10.20368/1971-8829/1135380>

CITE AS

Guelfi, M.R., Formiconi, A.R., Vannucci, M., Tofani, L., Shtylla, J., & Masoni, M. (2021). Application of peer review in a university course: are students good reviewers? *Journal of e-Learning and Knowledge Society*, 17(2), 1-8.
<https://doi.org/10.20368/1971-8829/1135380>

1. Introduction

Peer assessment is an educational strategy that requires learners to evaluate their peers based on criteria provided by the teacher (Topping, 1998). Peer assessment can be applied in different ways. Among the

different options available, it is frequently used the double blind peer review process, the established procedure that provides quality control in the production and progress of scientific knowledge.

In the educational context the peer review process implies that students evaluate and make judgment on the works of peers producing feedback reviews. At the same time students receive feedback reviews on their own work (Nicol et al., 2014).

As defined before, peer review represent a solution to act on learning through a double feedback process in which the students' works represents the basis on which reflections and judgements are made. When producing feedback, a self-evaluation process is made by students, that compares works of peers with their own. At the same time works of peers are compared to each others

¹ corresponding author - email: m.masoni@med.unifi.it

to produce feedback reviews that must take into account the criteria given by the teacher (*ibid.*). The self-evaluation process is considered very useful to facilitate the development of critical thinking which in turn will be central in lifelong learning and in the exercise of the professional activity (Geithner & Pollastro, 2016).

When receiving feedback reviews students are alerted about errors or gaps as well as misinterpretations contained in their own work. The suggestions of peers stimulate a critical thinking that leads to a reanalysis and revision of the initial work. If there is enough time in the course a self-review phase is highly recommended, as it allows students to immediately apply new ideas and different perspectives as well as to improve their writing skills (*ibid.*).

Many researchers argue that peer-review stimulates the development of student's evaluation skills that are usually ignored in traditional education. Cited benefits are (Mulder et al., 2012; Pelaez, 2002; Timmerman & Strickland, 2009):

- exposure to different perspectives in the analysis of a topic;
- development of critical thinking and problem solving skills;
- better attitude toward science;
- greater students' responsibility for their own learning.

For the rest of the paper it is useful to define the concepts of reliability and validity of peer assessment that are often misreported in the literature. Reliability is a variable that can be calculated by the consistency of marks given by peers. Validity is a variable that can be measured by the convergence between the students assigned marks and the teacher assigned mark (Bouzidi & Jaillet, 2009).

Feedback is a time consuming activity, but it is considered a crucial issue to enhance students' learning (Higher Education Funding Council for England, 2011). To provide feedback from peers is seen as an alternative to instructor's feedback and it can be useful to reduce teacher's workload, especially in courses with a high number of students. However, there are criticisms on peer-review as a teaching methodology.

One of the main issue of concern is the poor reproducibility and validity of student generated grades: in addition to having a poor knowledge of the discipline, they have rarely carried out review activities so the reliability of their evaluations could be poor (Cho et al, 2006). Others draw attention to the possible distortions resulting from friendships and deals between students (*ibid.*). In the latter case, the double-blind review allows to overcome the problem, since the student does not know the identity of the authors of the papers to be reviewed, nor of those who will review the one produced by him (Guelfi et al., 2019).

On the other hand, there are reasons to believe that peer assessment can be just as good as that of the teacher. In the first place, the latter may have reliability problems due to the high number of documents that he has to examine with the need to speed up the evaluation process. This situation does not occur in students peer review, where each one has a small number of papers to review which they can spend more time on. Secondly, a single paper is analyzed by several learners and the reliability of the set of assessments could be higher (Cho et al., 2006). Thirdly, the higher number of received feedbacks than those of the teacher can give a broader view of the topic (Topping, 1998). Finally, students' feedback reviews can be written in a more accessible way.

Applying peer review is a time consuming process, especially when teachers have classes with a high number of students. In this case it is crucial to have a peer review system, an educational tool that streamlines peer assessment implementation and reduces the teacher's workload. Peer review systems allow students to upload papers and then to distributes them randomly and anonymously, assigning each student the papers to be reviewed.

There are many peer review systems now available. Calibrated Peer Review (CPR) is a widely used program developed at UCLA that allows to hone ("calibrate") students' evaluation skills with sample assignments comparing their ratings to those assigned by the instructor (Robinson, 2005). CPR has been successfully used in hundreds of educational organizations all over the world. Moodle is a widespread Learning Management System that has a module, named Workshop, to manage peer review. Moodle workshop has been tested and it represents a reliable technology for peer review (Strang, 2015).

Higher education represents a context where large classes are common and peer review is frequently applied as a teaching methodology to enhance learning (Luckner & Purgathofer, 2015). Several studies have been published in the literature that used the peer-review process in the biomedical area.

In biology courses, B. Timmerman and D. Strickland (2009) have shown that not only graduated students but also undergraduates can be effective peer reviewers and that peer review improves reasoning skills, scientific writing and attitudes towards science. In a human physiology program, it has been demonstrated that peer review is one of the most effective learning activities that enhance students' perception of their scientific literacy and writing skills (Geithner & Pollastro, 2016). In a Doctor of Pharmacy curriculum, the inter-rater reliability of students' evaluations versus faculty evaluations was assessed through the CPR system. The results showed a fair inter-rater reliability between

scores assigned by pharmacy students and faculty members (Isaacs et al., 2020).

The assumption at the basis of this study is that the teacher is a reliable assessor. Instead, the literature reports many concerns about student's assessment skills. The focus of the study is to answer the question if students can be reliable reviewers as good as the teacher.

The article describes an experiment conducted in a large course at the School of Medicine of the University of Florence where peer review as a teaching strategy was applied. The aim of the study was twofold:

- assessing the validity of students' reviews, calculating the correlation between scores assigned by peers and those assigned by the teacher;
- assessing the validity of student's self-evaluation, calculating the correlation between student's assigned score and teacher's assigned score.

2. Materials and Methods

The study was conducted in the 2018/19 academic year in the course of Biomedical Informatics at the School of Medicine of the University of Florence. The course assigns 3 credits and takes place in the second semester of the first year of the medical degree.

It was delivered over nine weeks in blended learning mode, with about 60% of the teaching activities carried out remotely by Moodle and the Massive Online Open Course (MOOC) platform Federica of the University Federico II of Naples. The MOOC used in the course was entitled "Il Web e la ricerca di informazioni in rete" and it was developed by MRG and MM.

The face-to-face lessons were highly interactive by means of a Student Response System. These lessons were held one day a week and lasted for 4 hours. Between one face-to-face meeting and the next, a series of learning activities were carried out remotely using Moodle. The distance learning activities, which were mandatory in order to pass the final exam, were tracked and analyzed. The topics covered by the course were finalized to the production of a paper.

At the end of the course the peer-review process was applied. To the students were asked to:

- produce an individual paper on a topic chosen by the instructor;
- review and evaluate five papers produced by peers (peer review) by means of criteria provided by the teacher (rubric);
- then to evaluate its own paper applying the same rubric.

After the completion of the final work a face to face lesson discussed pros and cons of peer review and possible areas of improvement.

To carry out these activities, the Moodle Workshop module was used, an essential tool to automate and speed up process management.

The study cohort consisted of 330 students. The peer-review and the self-assessment activities were carried out by 95.10% of the students.

As showed during the face-to-face lessons, the student's work starts from a clinical scenario to be transformed into a searchable clinical question. The clinical scenario had to be an original student's idea, taking into account the novice knowledge possessed by the first year medical students. The clinical question had to be structured following the PICO model, a paradigm of Evidence Based Medicine. Then, after a PubMed subject search, the student had to choose the correct type of study to answer the clinical question, arguing it on the basis of the pyramid of evidence (Greenhalgh, 2014).

To produce the paper, the student had a maximum time of 10 days available, after which the essay had to be uploaded to the platform without putting one's name and surname and without identification marks in order to make it possible to maintain anonymity in the peer review activity. After the papers delivery phase, the Moodle Workshop module distributed them randomly and anonymously, assigning each student five documents to review (*Double blind peer review*). Each student was also asked to evaluate their own paper. The review phase was to be completed within 14 days. The double-blind peer review eliminates possible bias resulting from agreements between students.

The choice of 5 reviews to be performed by each student was highly considered. Some researchers have argued that from 5 to 7 reviews are the ideal number to increase the level of validity of the reviews themselves and they allow the authors to learn by comparing the opinions received (Cho et al, 2006).

To facilitate peer review and to guide students through the evaluation process rubrics are used. Rubrics are criteria-based marking schemes that raise the quality of assessment helping students to provide and receive standard feedbacks (Jones et al., 2017). Rubrics allow assessors to establish criteria and to define level of performance for each criterion. A rubric is usually presented as a table with criteria in the rows and rating with performance level in the columns.

In the study, the assignment has been divided in four detailed criteria: choice of keywords, PICO model, Pubmed subject search, choice of type of study. Every row of the rubric contains a criterion. The right column contains the scale range for each criterion, Unfortunately, it has not been possible to define precise

levels of performance for each rating due to the wide variety of errors that could be made by students. To overcome the problem, many examples of correspondence between errors and scores for each criterion were shown during the face-to-face lessons.

Table 1 shows the rubric used in the study with four evaluation criteria and, for each criterion, the scale range that reviewers can assign.

CRITERIA	SCALE RANGE
Correct Keywords choice	0-1
Transforming clinical question into a PICO model	0-4
Pubmed subject search	0-2
Choice of appropriate type of study related to the clinical question	0-3

Table 1 - The rubric used in the study.

To dispel students' doubts about the reliability of the reviews received from peers, each paper was evaluated by the teacher. The teacher's evaluation was essential to compare marks assigned by peers and the instructor's rating.

The final grade of the exam was calculated by adding:

- the scores acquired in the activities carried out on the e-learning platform during the course delivery (*maximum 11 points*);
- the teacher's score given to the student's paper (*maximum 10 points*);
- the score that measures the students' ability in evaluating the papers produced by peers and their own. The score was calculated by the algorithm implemented in Moodle Workshop that takes into account, for each paper reviewed, the difference between the teacher's mark and the student's mark for each criterion (*maximum 10 points*).

Each student's paper was assigned a teacher's mark expressed in tenth, a mark (in tenth) given by each of the five reviewers in addition to that assigned by the student himself. A statistical study was carried out in which teacher's evaluation, peers' evaluation and self-evaluation were compared.

Preliminarily, the main tests (Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling tests) to assess the normality of the three mark distributions were applied to the collected data: in the case of a test with a p-value < 0.05, the assumption of normality was refused. According to the results of the preliminary tests, normal distributions using Student-T test, and non-normal distributions using Wilcoxon signed rank Test were compared. The Lin's concordance correlation coefficient to test the agreement between teacher's marks and marks derived

by the peer-review process, and between teacher's marks and self-assessment marks was used. A p-value less than 0.05 is considered statistically significant.

An Excel spreadsheet was used to perform this analysis. Each student was anonymized and identified with a progressive serial number. For each student, three values were recorded: the mark assigned to the student's paper by the teacher; the average mark assigned to the students by the peers at the end of the peer-review process; the self-assessment mark, assigned by the student to his own paper. All marks were expressed in tenth (*minimum 1, maximum 10*) and represented as a continuous numerical variable; for each numerical variable, mean, median and standard deviation (SD) were reported.

3. Results

The relationship between the teacher's marks and the average marks derived by the peer-review process is represented by the scatter-plot in Figure 1.

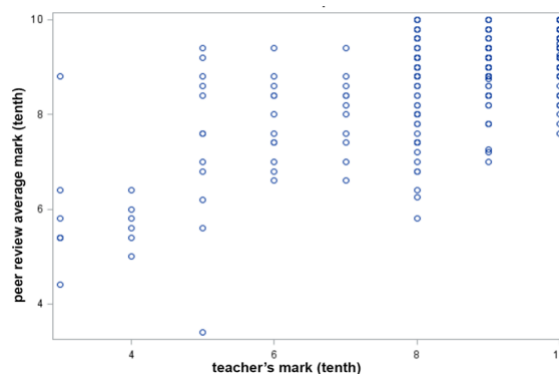


Figure 1 - Scatter-plot showing relationship between "teacher's mark", on the abscissa axis, and "peer review average mark", on the ordinate axis, for each student.

The box-plot in Figure 2 represents the distributions of the three variables "teacher's mark", "peer-review" and "self-assessment": the thick line represents medians.

Considering the first two distribution, the descriptive statistics are the following: "teacher's mark" mean 8.74 (SD 1.69), median 9 versus "peer-review" mark mean 8.89 (SD 1.12), median 9.2; the difference between the two means of the distributions is -0.15.

Table 2 shows the results of normality tests applied to the distributions "teacher's mark" and "peer-review": all the four tests (Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling) give a Statistica test resulting in a p<0.05; then, the hypothesis of normality of the distributions was not accepted, and, for the following comparison the non-parametric Wilcoxon signed rank test was used (Table 3).

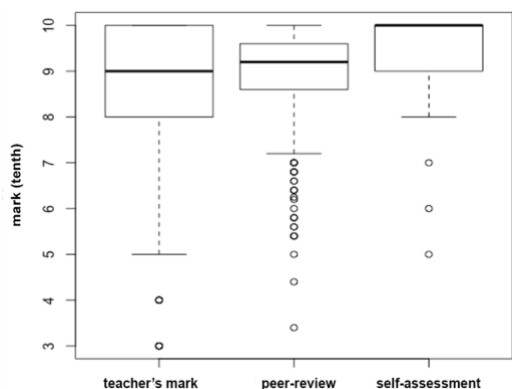


Figure 2 - Boxplot of the three distributions: “teacher’s mark”, “peer-review” mark and “self-assessment” mark.

Considering data showed in Figure 2 and Table 3, it is clear that there is a significant difference between the two distributions: the evaluation of the students tended to be higher than that attributed by the teacher. Nevertheless, as we can see comparing the medians (“teacher’s mark” median 9.0 and “peer-review” median 9.2), this difference is minimal and has no practical impact in modifying the final mark for the course.

The bar-plot in Figure 3 displays the absolute frequencies (in ordinate axis) of the differences, for each student, between the teacher’s mark and the average mark derived from the peer-review process: it is clear that more than 70% of the differences in the order of ± 1 compared to the teacher’s mark.

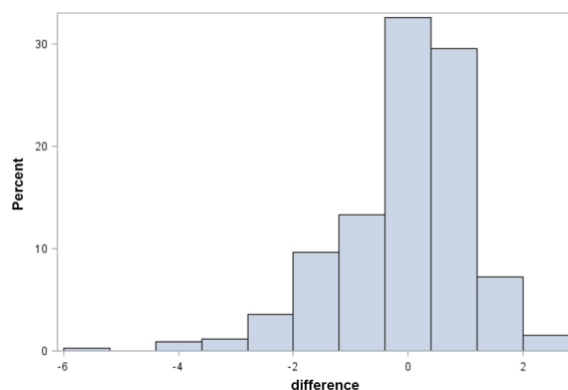


Figure 3 - Frequencies of differences between “teacher’s mark” and “peer-review” mark, for each student of the study sample.

Figure 4 presents the Lin’s concordance coefficient, expressed as mean and 95% confidence interval; this analysis has been conducted considering every single mark as a discrete numerical variable.

As proposed by McBride G. B. (2005), the Strength-of-Agreement between two discrete numerical variables can be evaluate as follows:

The aim of this study is to evaluate, for each paper, the concordance between:

1. Almost perfect: Lin’s concordance correlation coefficient > 0.90 ;
2. Substantial: $0.8 < \text{Lin’s concordance correlation coefficient} \leq 0.9$
3. Moderate: $0.65 \leq \text{Lin’s concordance correlation coefficient} \leq 0.8$
4. Poor: Lin’s concordance correlation coefficient < 0.65

Test	Statistica	P-value
Shapiro-Wilk	W 0.930093	< 0.0001
Kolmogorov-Smirnov	D 0.169882	< 0.0100
Cramer-von Mises	W-Sq 1.410353	< 0.0050
Anderson-Darling	A-Sq 7.172708	< 0.0050

Table 2 - The normality tests for the two distributions: “teacher’s mark” and “peer-review” mark.

Test	Statistica	P-value
Student’s T	T -25.5886	< 0.0001
Sign	M -148.5	< 0.0001
Signed Rank	S -27115.5	< 0.0001

Table 3 - Wilcoxon Signed Rank test for comparing the two distributions, “teacher’s mark” and “peer-review” mark; the first row shows the result of Student T-test for the same distributions.

Considering our lower one-side 95% confidence limit (0.6571), we can conclude that there is a moderate concordance between the teacher’s mark and the marks derived by the peer-review process.

The same analyses were carried out to assess the concordance between the teacher’s mark and the student’s self-assessment mark. In this case, the differences are wider: “teacher’s mark” mean 8.74 (median 9) versus student’s “self-assessment” mean 9.62 (median 10). As showed in Figure 5 and 6, not only the difference between “teacher’s marks” and “self-assessment marks” is wider than that presented in the previous analysis, but also this difference is greater than 2 in about 30% of cases.

<u>Lin's concordance coefficient</u>	<u>LCL 95%</u>	<u>UCL 95%</u>
0.6594	0.6571	0.6618

Figure 4 – Lin’s Concordance Correlation Coefficient for the two distributions (“teacher’s mark” and “peer review” mark). LCL 95%: lower one-side 95% confidence limit; UCL 95%: upper one-side confidence 95% confidence limit.

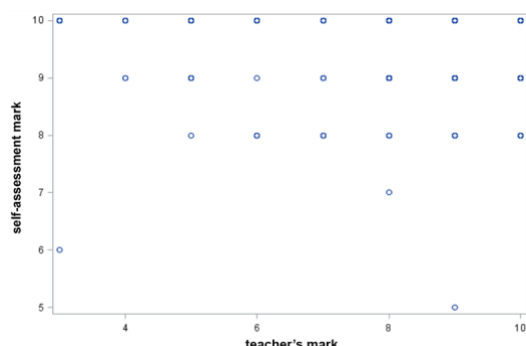


Figure 5 - Relationship between “teacher’s mark” and “self-assessment” mark distribution. it should be noted that low marks of the teacher correspond to high marks of self-assessment.

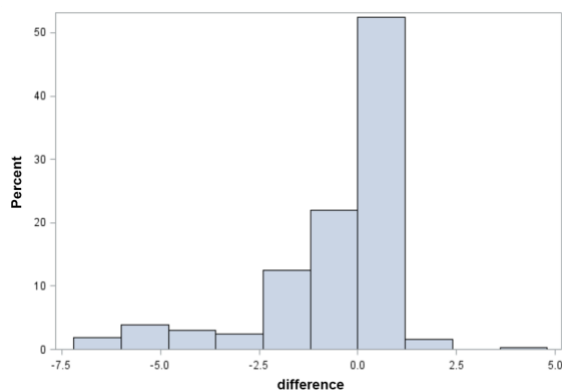


Figure 6 - Barplot showing the frequencies of the differences between the “teacher’s mark” and the “self-assessment” mark for each student.

In the case of the student’s self-assessment, a lower one-side 95% confidence limit of 0.085 indicates only a poor concordance (Figure 7).

<u>Lin's concordance coefficient</u>	<u>LCL 95%</u>	<u>UCL 95%</u>
0.087	0.085	0.088

Figure 7 – Lin’s Concordance Correlation Coefficient for the two distributions (teacher’s mark and “self-assessment” mark).

4. Discussion

The assumption behind the study is that the teacher is a reliable assessor, while there are concerns about the student’s skills both in peer grading and self-evaluation.

The first aim of the study is to determine whether students can be considered valid evaluators as good as the teacher. To address this issue the validity of students’ reviews has been assessed, calculating the correlation between peers’ scores and teacher’s score. The study demonstrated a moderate concordance between the marks assigned by peers and the marks assigned by the teacher (Figure 4). Nevertheless, as seen in Figure 2, the comparison between the teacher median (9.0) and the peer-review median (9.2) shows a minimal difference that has almost no effect on changing the final grade. This issue reinforces the results of the study.

In the literature it is reported that from four to six reviewers is an adequate number to obtain a result of agreement between the marks assigned by the teacher and those of the students (Cho et al, 2006). In the study, choosing five reviewers for each paper contributed to the satisfactory results.

In the literature there are few articles about the use of peer review in health profession education. As cited in the introduction, peer review approach was used in a Doctor of Pharmacy curriculum to assess the inter-rater reliability of students’ evaluations versus faculty evaluations. The results showed a fair inter-rater reliability between scores assigned by pharmacy students and faculty members (Isaacs et al., 2020). The experiment was conducted with the CPR system.

Another study used the same educational tool in third-year medical students. It examined the effectiveness of CPR to teach and assess students’ patient note-writing skills using three longitudinal activities. (McCarty et al., 2005). The results showed a progressive improvement of the student’s CPR activity. The best

alignment between students' scores and faculty scores emerged in the last activity.

For the application of peer review this study used Moodle's Workshop, that do not have a calibration phase of student's grading skills. It is possible that different results would have been obtained with the CPR system.

The second aim of the study was to assess the validity of student's self-evaluation, calculating the correlation between student's score and teacher's score. The results of the study showed poor concordance between the marks assigned by the teacher and those relating to the students' self-evaluation (Figure 7). This is confirmed by Figure 5, where it can be seen that students' marks are higher than teacher marks.

Probably the students do not have enough content knowledge and metacognitive skills to carry out evaluation of their own paper. Another possible explanation of these data is a cognitive process called self-prophecy bias, where students tend to overrate their own performance (Strang, 2015).

From these results it emerges that self-evaluation is less valid than peers evaluation. These data are confirmed in another study (Cho et al, 2006). As consequence, the Authors will exclude self-evaluation from the calculation of the final grade in future studies.

The adoption of new teaching strategy can have benefits that go beyond the single course involving other faculty members. The promising results of the study promoted the implementation of the same teaching methodology in the first year course of Biology at the School of Medicine at the University of Florence.

4.1 Limits of the study

The study focused on the validity of the peer-review process. As consequence, the effectiveness of a teaching strategy based on peer-review on student learning was not considered.

The peer review cycle involves a self-review phase where students update and resubmit the assignment. This phase is highly recommended, as it allows students to immediately apply new ideas and different perspectives (Geithner & Pollastro, 2016). Unfortunately, the time constraints of the course prevented the application of this phase.

Many researchers suggest that giving feedbacks with marks and comments facilitate the evaluation of the teacher and give better learning results than using only marks (Nicol et al., 2004). In the study the high number of students prevented the use of comments due to the high teacher workload.

The reliability of the rubric and the clarity of teacher's presentation on illustrating its use are variables that

greatly affect the validity of the peer review process (Bouzidi & Jaillet, 2009).

At the end of the course during a plenary discussion some students suggested a change to the rubric. They asserted that it was difficult to extrapolate the correctness of the choice of keywords from the construction of the PICO model with a consequent uncertainty in assigning the correct grade. The teachers accepted this suggestion by simplifying the rubric for the following academic year in such a way as to divide the analysis of the work on the basis of three criteria: PICO model, Pubmed subject search, choice of type of study. Consequently, a change in the scale range of the criteria was done.

At the end, no gender difference was considered which could have provided further information.

5. Conclusions

Peer review seems a viable teaching methodology to improve students' learning and critical thinking. The use of peer review systems also makes it possible to apply this teaching strategy to classes with a high number of students, thus reducing the teaching load on the instructor redirecting part of the work to the students.

From the perspective of the teacher concerns exist about the reliability and validity of peer grading assignment. The results of the study have demonstrated a moderate concordance between the marks assigned by peers and the marks assigned by teacher. On such basis, concerns have not to prevent the application of peer review in higher education, at least with appropriate scaffolding.

Given that the concordance is moderate, further studies are needed to evaluate the application of peer review as a teaching methodology in the medical area. However, the current work demonstrate that peer review seems a promising approach to be used in the field of medical education as innovative teaching strategy.

Acknowledgements

The Authors thank Dr. Fabio Picciafuochi of the Department of Anesthesiology, Neuroanesthesia and Intensive Care, Careggi University Hospital (Florence) for the helpful translation tips.

References

- Bouzidi L., Jaillet A. (2009), Can Online Peer Assessment be Trusted?, *Educational Technology & Society*, 12 (4), 257-268.
- Cho K., Schunn C. D., Wilson R.W. (2006), Validity and Reliability of Scaffolded Peer Assessment of Writing From Instructor and Student Perspectives. *Journal of Educational Psychology*, 98, 891-901.
- Geithner C.A., Pollastro A.N. (2016), Doing peer review and receiving feedback: impact on scientific literacy and writing skills. *Adv Physiol Educ*, 40, 38-46.
- Greenhalgh T., eds (2014), *How to read a paper. The basics of Evidence Based Medicine*. Wiley Blackwell V edition.
- Guelfi M.R., Masoni M., Shtylla J., Formiconi A.R. (2019), Peer assessment nell'insegnamento di Informatica del Corso di Laurea in Medicina e Chirurgia dell'Università di Firenze. Firenze, Firenze University Press.
DOI: 10.36253/978-88-6453-890-7
- Higher Education Funding Council for England (2011), *The National Student Survey: Findings and Trends 2006–2010*. Bristol: Higher Education Funding Council for England.
- Isaacs A. N., Miller M. L., Hu T., Johnson B., Weber Z. A. (2020), Inter-Rater Reliability of Web-Based Calibrated Peer Review within a Pharmacy Curriculum. *American journal of pharmaceutical education*, 84(4), 7583.
- Jones L., Allen B., Dunn P., Brooker L. (2017), Demystifying the rubric: a five-step pedagogy to improve student understanding and utilization of marking criteria. *Higher Education Research & Development*, 36:1, 129-142.
DOI:10.1080/07294360.2016.1177000
- Luckner N., Purgathofer P. (2015), Exploring the use of peer review in large university courses. *IxD&A*, 25: 21-38.
- McBride G.B. (2005), A Proposal for Strength-of-Agreement Criteria for Lin's Concordance Correlation Coefficient. NIWA (National Institute of Water & Atmospheric Research) Client Report: HAM2005-062.
- McCarty T., Parkes M. V., Anderson T. T., Mines J., Skipper B. J., Grebosky J. (2005), Improved patient notes from medical students during web-based teaching using faculty-calibrated peer review and self-assessment. *Acad Med*, 80(10 Suppl), S67-70.
DOI: 10.1097/00001888-200510001-00019
- Mulder R., Pearce J., Baik C., Payne C. (2012), Guide to student peer review. URL: <http://peerreview.cis.unimelb.edu.au/wp-content/uploads/2012/06/Academic-guide-FINAL.pdf> (ver 25/08/2021).
- 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.
- Pelaez N.J. (2002), Problem-Based Writing with peer-review improves academic performance in Physiology. *Adv Physiol Educ*, 26, 174-184.
- Robinson R. (2001), Calibrated Peer Review™ an Application to Increase Student Reading & Writing Skills. *The American Biology Teacher*, 63(7), 474-480.
- Strang, K.D. (2015), Effectiveness of peer assessment in a professionalism course using an online workshop. *Journal of Information Technology Education: Innovations in Practice*, 14: 1-16.
- Timmerman B., Strickland D. (2009), Faculty should consider peer review as a means of improving students' scientific reasoning skills. *J Sc Acad Sci* 7: 1.
- Topping, K. (1998), Peer assessment between students in colleges and universities. *Rev Educ Res* 68, 249–276.