

---

## Peer Assessment in Architecture Education

*Mafalda Teixeira De Sampayo,*

Lisbon University Institute, Lisbon, Portugal, mgts@iscte.pt

*David Sousa-Rodrigues,*

*Cristian Jimenez-Romero and Jeffrey Johnson*

The Open University, Milton Keynes,

United Kingdom, david.rodrigues@open.ac.uk

### Abstract

The role of peer assessment in education has become of particular interest in recent years, mainly because of its potential benefits in improving student's learning. It also presents benefits in time management by allowing teachers and tutors to use their time more efficiently to get the results of student's assessments quicker. Peer assessment is also relevant in the context of distance learning and massive open online courses (MOOCs).

The discipline of architecture is dominated by an artistic language that has its own way of being discussed and applied. The architecture project analysis and critique goes beyond the technical components and programme requirements that need to be fulfilled. Dominating the architecture language is an essential tool in the architect's toolbox. In this context peer assessment activities help prospective architects to develop skills early during their undergraduate education.

---

In this work we show how peer assessment acts as a formative activity in architecture teaching. Peer assessment leads the students to develop critical and higher order thinking processes that are fundamental for the analysis of architecture projects. The applicability of this strategy to massive open online education systems has to be considered as the heterogeneous and unsupervised environment requires confidence in the usefulness of this approach. To study this we designed a local experiment to investigate the role of peer assessment in architecture teaching.

This experiment showed that students reacted positively to the peer assessment exercise and looked forward to participating when it was announced. Previously to the assessment students felt engaged by the responsibility of marking their colleagues. Subsequently to the first iteration of the peer assessment, professors registered that students used elements of the qualitative assessment in their architecture discourse, and tried to answer the criticisms pointed to their projects by their colleagues. This led their work in directions some hadn't considered before.

The marks awarded by the students are in good agreement with the final scores awarded by the professors. Only in 11 % of the cases the average score of the peer assessment differed more than 10 % from marks given by the professors. It was also observed that the professor's marks were slightly higher than the average of the peer marking. No correlation was observed between the marks given by a student as marker and the final score given to that student by the professors.

The data produced in this experiment shows peer assessment as a feedback mechanism in the construction of a critical thought process and in the development of an architectural discourse. Also it shows that students tend to mark their colleagues with great accuracy. Both of these results are of great importance for possible application of peer assessment strategies to massive open online courses and distance education.

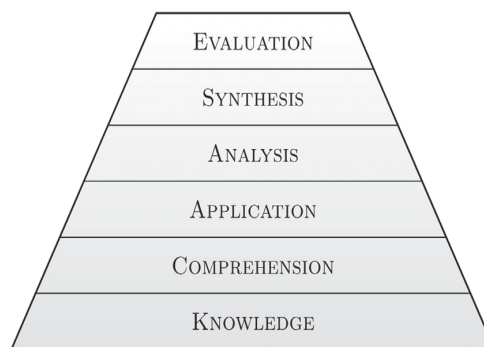
## Keywords

Education, MOOC, Architecture, Peer Assessment, Marking

---

## 1. Introduction

The role of peer assessment in education has become of particular interest in recent years, mainly because of its potential benefits in improving student's learning [1] and benefits in time management by allowing teachers and tutors to use their time more efficiently to get the results of student's assessments quicker [2]. Peer assessment has also relevant in the context of distance learning and massive open online courses (MOOCs) [3]. These education systems have scalability problems in the cost of evaluating students and new strategies are being researched to lower the cost of evaluating many students. In this context peer assessment is very important because it is scalable. For each new student we get one new marker for the system and efforts are being made to differentiate between good and poor markers [4]. Although these efforts are oriented towards objective learning subjects, peer assessment can also be applied in the context of subjective fields, like architecture, painting or music, where it has an intrinsic pedagogic value as a formative activity.



**Figure 1: Bloom's taxonomy six levels of the cognitive domain.**

The discipline of architecture is dominated by an artistic language that has its own way of being discussed and applied. The architecture project analysis and critique goes beyond the technical components and programme requirements that need to be fulfilled. Dominating the architecture language is an essential tool in the architect's toolbox. The establishment of a method of 'doing architecture' in the student's early learning years is a slow process. It is impossible to reduce the architecture practice to one dimensional aspect and therefore it is of utmost importance for students to develop a critical thinking process about the architecture design process [5]. In this context peer assessment activities can help them develop skills early in their undergraduate education.

In this work we show how peer assessment acts as a formative activity in architecture teaching. Peer assessment leads the students to develop critical and higher order thinking

---

processes that are fundamental for the analysis of architecture projects [2, 6]. The applicability of this strategy to massive open online education systems has to be considered as the heterogeneous and unsupervised environment requires confidence in the usefulness of this approach. To study this we designed a local experiment to investigate the role of peer experiment in architecture teaching.

The method used in the peer assessment experiment aims help student to develop their higher order cognitive skills as defined by Bloom's taxonomy [7, 6, 8]. Bloom identified six levels within the cognitive domain. The six levels (see figure 1) of the cognitive domain<sup>1</sup> characterise fundamental questions that educators have in relation to the learning objectives set for students. This experiment potentiates the development of the the highest order levels of *analysis*, *synthesis*, and *evaluation* in the students. Analysis refers to the ability to break down material into its component parts, the ability of the student to look at an architecture project and be able to identify the main components of the project that are relevant. Synthesis is associated with the ability to establish a discourse about the project, the ability to have an architecture vocabulary that allows the student to produce communication about the architectural object being analysed. Evaluation is concerned with the ability to judge the value of the material for a given purpose. These judgements are defined against architectural criteria and this learning outcome is the highest in the cognitive hierarchy because it contains elements of all levels (including the lower levels knowledge, comprehension, and application).

In the context of massive open online courses (MOOC) the assessment made by paid staff is prohibitively high and other ways of assessing students achievements has to be derived. Typically the assessment is made by using multiple choice questions, computer assisted marking or peer marking.

It is this final aspect of peer marking that is of particular interest, because it makes the task "evaluation" scalable because each new student that enters the system is also a marker for the system. The downfall of this approach is that not all students are equivalent markers. The same way they will not learn the subjects with the same speed, they won't be able to mark others with the same level of competence and accuracy exhibited by a professional marker. Our experiment shows that this difficulty can be solved by collective action in the form of peer assessment.

The experiment described here was setup to test the premise that any two good markers will show a marking behaviour that is consistent over time and that two bad markers will present uncorrelated marking behaviours. Our hypothesis is that based on these premisses, one can identify different quality markers.

---

<sup>1</sup> Bloom also defined an affective domain and a psychomotor domain in his classification [7]

---

## 2. Methods

In our experiment we have 45 students from two professors of the architecture programme of the Lisbon University Institute. The two classes belong to the course *Architecture IV* (4<sup>th</sup> semester in the undergraduate programme). This course is mainly practical with classes taking place in a simulated architecture office environment.

The peer assessment activities were made during two distinct time periods. The first peer assessment session was held midway the semester and the marking occurred only among classmates. The second peer assessment activity occurred at the end of semester simultaneously with the student's final examination. This time students assessed each other's work irrespective of their class. In both phases each student assessed three other randomly chosen students.

The assessment was divided in a qualitative part and a quantitative mark. Each student was asked to identify the positive aspects of the assessed project and also to identify its flaws and future improvements. They also marked the overall project with a score based on the achievements of the project against the programme of the exercise proposed for the semester.

The first assessment was done in the classroom during a 3h period during which the students were given a slip of paper to fill (depicted in figure 2). They had to fill the slip of paper with both qualitative and quantitative assessment about the project.

A marking guide was distributed (see supplementary material) where students were reminded of the education objectives that were expected at that phase of the semester (the first phase was held on the 6<sup>th</sup> week of a 12 week course). The students were informed to only consider the aspects mentioned in the marking guide and the materials presented by their colleagues for assessment and to disregard any previous experience and interactions they might have had with that project and its authors. This was very important as classes are conducted in a simulation of a architecture studio and it is normal for students to share ideas, discuss projects and solutions. In the peer assessment exercise it was asked that students disregarded any previous information about the projects they were analysing.

The second phase of the peer assessment was conducted in a similar setting with the exception that it occurred at the end of the semester during the public presentations of their work to a jury. This jury was formed by the two professors of architecture and by an external invited member (usually a professional architect). This public examination occurred over two days and the students were asked to do the peer assessment based on their colleagues presentations.

---

PROJECT ID	POSITIVE ASPECTS	MARK (0-100)
MARKER ID	ASPECTS TO IMPROVE	

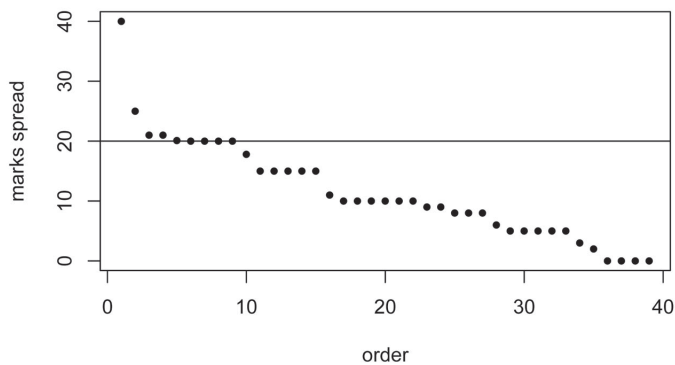
**Figure 2: Slip of paper used in the Peer Assessment**

### 3. Results

This experiment showed that students reacted positively to the peer assessment exercise and looked forward to participating when it was announced. Previously to the assessment students felt engaged by the responsibility of marking their colleagues. Subsequently to the first iteration of the peer assessment, professors registered that students used elements of the qualitative assessment in their architecture discourse, and tried to answer the criticisms pointed to their projects by their colleagues. This led their work in directions some hadn't consider before.

The quality of the peer assessment process was very high and through textual inspection of the student's answers the professors concluded that the limited space available for the qualitative aspects forced students to synthesise and develop a critical thought process. Globally the comments made by the students were very assertive, but in some cases they showed that some students still didn't possess an architecture discourse capable of communicating in architecture language. The peer assessment was very useful in identifying such cases.

The quantitative marks awarded by the students in the second peer marking period are in good agreement with the final scores awarded by the professors. Internal consistency between the marks assigned by the different students was high. In a 100 point scale the spread of the marks was low in the majority of the cases as shown in figure 3. Only in one case the marks varied 40 points while the majority showed less than 20 point of variation.



**Figure 3: Marks spread, ordered from highest spread to lowest spread (in a 100pt scale).**

The observation that the marks spread was low allowed us to compute a simple arithmetic average of the marks without incurring in great errors or needing advanced schemes for computing a score from the individual marks.

Alternatives to the simple average would include the removal of marks that diverged from the average more than a fixed threshold or considering a weighted computation based on other information like students previous reputation. The latter option would require an iterative marking process that we didn't have in this experiment while the former would not be fully useful in this scenario as the number of marks per student was low (on average each student marked another three students work) making the outlier detection difficult, if not impossible.

Only in 5 cases the average score of the peer assessment differed more than 10 % from marks given by the professors (students outside the range  $[-2,+2]$  in figure 4). This represents less than 12 % of the students. It was also observed that the professor's marks where slightly higher than the average of the peer marking. No correlation was observed between the marks given by a student as marker and the final score given to that student by the professors. This seems to imply that a good marker doesn't necessarily need to be a good student, mainly if following a marking guide.

Besides the agreement between the averaged marks and the final mark given by the jury observed in figure 4, we wanted to understand if there was a correlation between the marking activity and the learning activity of the students. Is a good student also a good marker? For this we computed the marker average error by comparing the marks they assigned to those given by the jury and computed the Pearson correlation between this average error and the mark they received as students by the Jury.

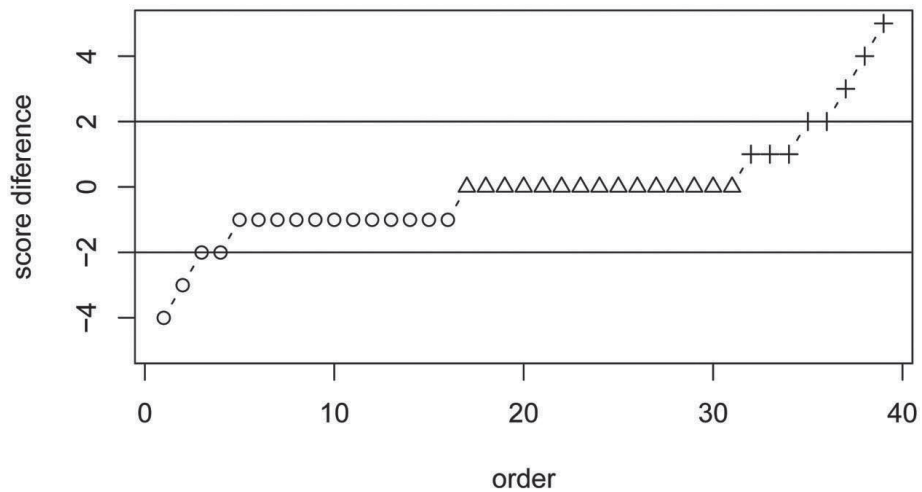


Figure 4: Difference between the average mark and the jury score (in a 20pt scale).

The results show that there is no correlation between the two activities. The Pearson correlation value is  $-0.063$ . This lack of correlation is also observed in the plot of the average marker's error against their final mark in figure 5.

This lead us to think that in this particular context, the marking activity and the application of their architectural knowledge isn't yet correlated. While the marking activity was positive correlated with the final marks awarded by the jury, with a Pearson correlation of  $+0.65$ , meaning that student were able to analyse and criticise other student's work, this ability was not directly correlated with their own work performance as if the application of the concepts they used while evaluating others was not taken fully into consideration in the output of their projects.

We found this result surprising and it highlighted the importance that the peer assessment activity has. It exposes students to aspects of architecture that they know in abstract, but that they failed to consider when are transferring that knowledge into their own work.

By moving the students from the role of the architect to the role of the critic, they become aware of aspects of the architectural process that they didn't consider in their own work. This leads to a reflection process and forces them to try to fix the deficiencies in their architectural practice in subsequent work.



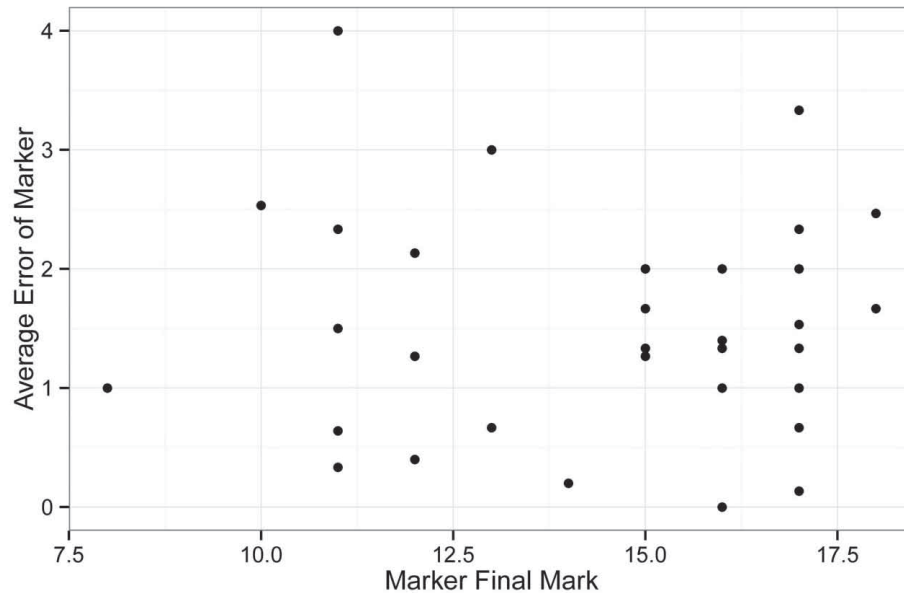


Figure 5: Average marker's error vs. marker final grade (in a 20pt scale).

## 4. Discussion

In this paper we presented an educative experiment that incorporates peer assessment in the teaching of Architecture to second year undergraduate students. The experiment of peer assessment had a formative aspect to the teaching of the students as it allowed the development of a critical thinking process about other students projects.

This activity presents itself as a very interesting way to tackle some pedagogic objectives, namely those that are usually in the top of Bloom's taxonomy and require the higher cognitive skills of *analysis*, *synthesis*, and *evaluation*. The exercise allowed students to engage in high level thinking about architecture project. It was observed in class that students incorporated aspects of the languages used during the experiment in their architectural discourse. Also, the students employed elements and suggestions from their colleagues into their respective work, meaning that the peer assessment worked clearly as a feedback mechanism for students. This has the main advantage of relieving the task of giving feedback on their progress exclusively from the professor(s) of the discipline, allowing for more feedback points during the semester.

---

These observations prove the usefulness of the peer assessment strategy as a formative tool. The data produced in this experiment shows peer assessment as a feedback mechanism in the construction of a critical thought process and in the development of an architecture discourse and language.

The quantitative aspect of this experiment shows a correlation between the marks given by a student and the final mark given by an expert panel (professors, professionals). Students tend to mark their colleagues with great rigour. This is a very important conclusion because it is of great importance for possible applications of these strategies to massive online courses and distance education.

The application of peer marking to massive online education poses many problems but the tool can be used for aspects of the formative process besides the attribution of a final mark. In this experiment we showed that the marking was in agreement with the final expert mark. This is good indicator for future experiments in larger scales. The success of this experiment can be attributed to the controlled environment and the engagement of the students. The task of doing peer assessment was considered by the students to be fun. In an online course people might not be engaged all at the same level. Their starting baseline might be more heterogeneous than in traditional courses. They might drop off and not do their assigned assessments, or just rush over the assessment to get it done. These kinds of problems need to be tackled if the next iteration of this work in order to apply peer assessment to massive online courses. In any case there are already strategies of identifying good markers and bad markers by using iterative peer assessment schema [4] and it is our objective to explore these results in the development of future education systems.

## Acknowledgments

The authors would like to thank Professor Helena Botelho of ISCTE-IUL for her contributions to this experiment.

## References

- [1] McDowell, L. and Mowl, G.; Innovative assessment – its impact on students; In Gibbs, G. (ed.) *Improving student learning through assessment and evaluation*, Oxford: The Oxford Centre for Staff Development, 131–147, 1996.
- [2] Sadler, P. and Good, E.; The Impact of Self-and Peer-Grading on Student Learning. In *Educational Assessment* 11.(1), 1–31, 2006.
- [3] Jimenez-Romero, C.; Johnson, J.; De Castro, T. 'Machine and social intelligent peer-assessment systems for assessing large student populations in massive open online education' In *12<sup>th</sup> European Conference on e-Learning ECEL-2013*, SKEMA Business School, Sophia Antipolis, France, 30–31 October 2013
- [4] Johnson, J.; Jimenez-Romero, C.; Rodrigues, D.M.S.; Bromley, J.; Willis, A.; Hypernetwork-based Peer Marking for Scalable Certificated Mass Education; In *European Conference on Complex Systems*, Lucca, Italy, 2014
- [5] Krüger, M.; Do Paraíso Perdido à Divina Comédia – Reflexões Sobre o Ensino de Arquitectura em Países Anglo-saxónicos e Latinos In *Jornal Arquitectos* 201, 36–47, 2001

---

[6] Bloom, B. S.; Mastery learning. In J. H. Block (Ed.), In *Mastery learning, theory and practice* 47–63. New York: Holt, Rinehart, and Winston, 1971.

[7] Bloom, B. S.; Taxonomy of Educational Objectives: The Classification of Educational Goals Longman group, U.K., ISBN-13: 978–0679302117, 1969.

[8] Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P., Raths, J., and Wittrock, M. C. A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives. *Anderson LW and Krathwohl DR. New York: Addison Wesley Longmann, 2001.*

## Assessment guide

(Translated from the original version in Portuguese)

### Architecture IV

As defined in the by the curriculum of Architecture IV with respect to the learning objectives of this teaching unit, architecture is defined as an experimental activity, where the project is a moment of synthesis that integrates many factors – context, idea, form, function, scale, language. The project is based on the development of concepts through drawings and models (its main interments) in the realm of the creative process.

The teaching methodologies privilege a practical approach to the learning process where students are incentivated to use varied means of representation, both as research tools and as communication support.

In this methodologies it is incentivated the argumentation ability of students about the options taken by them in the execution of their work and the ability to argue about the qualities of other students work. This aims the development and consolidation of the students critical thinking in interpreting the location and the quality of the proposed spaces.

It is proposed to the students that they do a critical evaluation (both qualitative and quantitative) of their students colleagues.

The evaluation must take into consideration the following factors:

1. The student knows the content of the exercise of this course and must ponder what is asked in this guide (including the elements that the students are expected to submit at this phase).
2. The final evaluation of the student (as marker) is weighted by his critical thinking ability.
3. The student must only assess the result of the project and not its development.
4. The student (as marker) must not communicate with the colleague author of the marked project during the assessment (can not make questions about the project).

At the end of the exercise each student will receive the feedback with his/her colleagues assessments.