

Analysis of validity of peer assessment: a study in Higher Education

Análisis de validez en la evaluación entre iguales: un estudio en educación superior

Pascual-Gómez, Isabel; Lorenzo-Llamas, Elena María & Monge-López, Carlos

Universidad de Alcalá de Henares

Resumen

No son muy numerosas las investigaciones que se han realizado en España sobre la participación de los estudiantes en los procesos de *evaluación entre iguales* en el marco de la Educación Superior. El objetivo de este trabajo es conocer con mayor precisión un proceso de evaluación participativa en el marco de un proyecto de aprendizaje cooperativo. Se recogieron datos de una muestra de 119 estudiantes matriculados en una asignatura del Grado de Magisterio en la Universidad de Alcalá. El procedimiento de recogida de datos fue online. Se diseñaron actividades de evaluación entre iguales y autoevaluación grupal. Los datos se compararon con las calificaciones emitidas por el profesor. Los resultados del análisis de contraste de medias arrojan una diferencia significativa de 1-2 puntos entre ambas puntuaciones a favor del alumnado. A pesar de esta sobrevaloración, el análisis de contenido del instrumento de autoevaluación concluye la utilidad de esta investigación: este tipo de procesos mejora *capacidad formativa de los procesos de evaluación* tanto para el discente como para el docente.

Fecha de recepción
21 de Julio de 2014

Fecha de aprobación
18 de Enero de 2015

Fecha de publicación
3 de Abril de 2015

Palabras clave:

Aprendizaje cooperativo; autoevaluación; carpeta de aprendizaje; Educación Superior; evaluación entre iguales; evaluación para el aprendizaje; heteroevaluación; perspectiva émica; rúbrica

Abstract

In Spain only a limited number of studies have been done regarding students evaluating their peers in higher education. The objective of this article is to analyse the way in which students that have participated in a cooperative learning project evaluated their peers. Data were collected online from a sample of 119 bachelor students that were registered in a teacher training course at the University of Alcalá. Activities were designed for both peer evaluation and self-evaluation in group work. These student evaluations were compared with ratings given by the teacher. A contrast analysis showed a statistically significant difference of 1-2 points between the mean scores, in favour of the students. Despite these overestimated values, an analysis of the content from the self-evaluation tool showed its usefulness for improving the formative assessment process for both students and teachers.

Reception Date
2014 July 21

Approval Date
2015 January 18

Publication Date:
2015 April 3

Keywords:

Cooperative learning; self-assessment; portfolio assessment; Higher Education; peer-assessment; assessment for learning; hetero-assessment; emic perspective; rubric

One of the principal tasks facing teachers is to plan and perform assessments, whether of their students, their programmes or their own teaching praxis. Proper assessment ensures the continual enhancement of the quality of the

teaching-learning processes. That is why in Higher Education attempts are made to guarantee suitable training in assessment-related competences as part of the programme

Corresponding author / Autor de contacto

Pascual-Gómez, Isabel (isabel.pascualg@uah.es). Departamento de Ciencias de la Educación de la Universidad de Alcalá de Henares. C/ Madrid, 1 19001 Guadalajara (España)

of studies pursued by students of the bachelor's degree in Infant Education.

Outside Spain, many researchers have concerned themselves with the question of how to check the validity and reliability of this kind of assessment. Among others, Montero-Fleta (2006) cites "Freeman's (1995) experiments with group work assessment, Orsmond, Merry and Reiling's (1996) research into correction criteria in joint assessment, and Kwan y Leung's (1996) and MacAlpine's (1999) research into differences between the points awarded by teacher and students" (p. 32).

Also worth mentioning is Bretones' (2008) research on meta-analytical review. Many of these studies have confirmed the reliability of this type of assessment and of the high level of agreement between students' and teachers' marks, giving positive correlations. But assessment discrepancy has also been found. Pond, Ul-Haq and Wade (1995) claim that there are occasional cases of over-evaluation due to friendships, lack of differentiation within the group, domineering members of the group who inflate their evaluations and others who fail to participate and benefit from the work of the rest.

Under the conviction that assessment should be participative and cooperative and that virtual assessment environments facilitate this kind of interaction, this study is an attempt to explore the strengths and weaknesses of an assessment experience in the framework of cooperative learning carried out from a virtual learning platform. As evaluations from different types of assessment will be compared and contrasted by means of convergent validity analysis, it is first of all necessary to clarify for the reader some of the terminology used in relation to the assessment process as carried out in this research. The basic theoretical concepts used in this article refer chiefly to two terms related to assessment modes which differ according to who does the assessing. Rodríguez and Ibarra (2011) distinguish two types: *hetero-assessment* and *self-assessment*. In a subsequent publication

(Rodríguez, Ibarra and García, 2013) the matter is stated as follows: "the prefix 'hetero' or 'self', as used in connection with assessment, is determined by the nature of the relationship that obtains between the evaluator and the object of evaluation (learning, teaching, etc.)" (p. 200).

In this article, *hetero-assessment* will refer to assessment in which teacher and students play quite distinct roles, with the teacher as the agent who assesses the students and their learning. In contrast, student assessment of the teacher's role would, according to López (2009), be not so much hetero-assessment as a process of joint assessment aimed at improving the teacher's training and teaching.

Self-assessment may be defined as the assessment people conduct of themselves, or of their own learning process. As Falchikov (1986 and 2005) puts it, in *self-assessment* students judge their own performance and/or output in line with their own assessment criteria. Falchikov and Boud (1989) note the benefits of this type of assessment in so far as it involves students in the decision-making process regarding their own learning. Roberts (2006) believes this is a valuable mode of assessment since students "reflect critically, record the process and may suggest their own learning levels" (p. 3). More thoroughgoing are Rodríguez et al. (2013) who define self-assessment as the "process through which students analyse and evaluate their actions and/or their productions" (p. 202). In education, it is customary to refer to the self-assessment of the main agents in the teaching-learning process, namely, teacher and student; nevertheless, it is most commonly used in relation to the student. Yet Panadero and Alonso-Tapia (2013) have qualified this in their suggestion that self-assessment enriches students learning since it is "a process performed by students to self-regulate their learning [...] and a teaching strategy" (p. 173). In our study we apply the term to a group process rather than an individual one, a process in which outcomes were reached on the basis of the negotiated consensus of

individuals, that is to say, of the student members of a cooperative learning group.

The students' participation in this study's assessment processes is evaluated on the basis of previous research which emphasise both its benefits and its needs, as we saw earlier in this section. In line with this positive approach, our study covers *peer assessments*, which is taken to be the maximum expression of that kind of student participation. According to Falchikov (1995), "peer assessment is the process by means of which groups of individuals give marks to their companions" (p.175); such is the assessment carried out in our study. As a result, and as Topping (2009) explains it, the students consider and specify the level, value and quality of a product (the subjects practical assignments) and the performance of other students of the same status (companions in a cooperative learning group).

At the same time, in our study all the assessment types indicated so far had to contribute to achieving a single academic goal, namely, learning on the part of the students. This is where a theoretical-practical framework also evaluated in the course of our research comes in, namely, *assessment for learning*. According to López (2009), by "assessment for learning" is meant a clear orientation towards enhancing students' learning rather than any mere control and marking of learning. More broadly, some experts like Boud and Falchikov (2006) speak of processes developing both self-regulating capacities in the learning process itself and life-long learning. Boud and Falchikov's (1989) definition has been refined by Brown and Glasner (2003), who distinguish between the fostering of self-regulation and the control of the learning process itself, which they term formative assessment. Such a distinction is irrelevant to the goals and limitations of our study, and we have accordingly only worked with assessment for learning as defined by López (2009) and Alonso-Tapia and Hernansaiz (2013), for whom it is a pedagogical tool which has an influence on the way students set about learning (Alonso-Tapia

and Hernansaiz, 2013) and may foster the development of long-term learning competences (López, 2009).

To this end *the features of assessment for learning* which have shaped our study need to be specified. Our point of departure are those errors which, according to Sanmartí (2007), teachers make and which hamper students' learning: the ways in which the environment is perceived, reasoning processes, forms of communication, ways of feeling (beliefs, values and emotions) and the ideas that are transmitted. Thus, the features of genuine assessment for learning may be summarised as: the design of activities which enable the origin of difficulties encountered in the learning process to be identified; consideration of the processes involved in understanding ideas; furnishing students with the right assistance for them to be able to overcome learning difficulties. As Alonso-Tapia and Hernansaiz (2013), "[...] the primordial function of assessment is to identify the origin of students' difficulties concerning the competence at stake, with the view to offering them the specific assistance they require in order to make progress" (p. 8).

Method

The chief goal of this study is to obtain information about the development of a cooperative learning process and about the assessment processes which take place in the framework of a university-level bachelor's degree in Infant Education. The specific objectives were:

- To ascertain the convergent validity by means of calculating the agreement level between the assessments carried out by the actors involved in the teaching-learning process, namely, the teacher (hetero-assessment) and the students (self- and peer-assessment).
- To interpret the data after checking against an emic evaluation (Cochran-Smith and Lytle, 2002).

Procedure

Our study’s research plan conforms a descriptive analysis by means of self-observation with the aid of mixed data collection and analysis techniques. The data collection techniques employed were individual records of third-party behaviour (a model rubric designed for data collection in peer-assessment and an evaluation scale for hetero-assessment) and individual records of intra-group behaviour (a questionnaire with open questions designed for group self-assessment).

The procedure was carried out in line with the criteria proposed by Ibarra, Rodríguez and Gómez (2012) for peer-assessment, and in the following stages:

- a. A preparatory phase in which we discussed the assessment criteria regarding participation in the group and the assessment tools themselves.
- b. The formation of cooperative learning groups.
- c. The performance of two group assignments working on contents related to the subject.
- d. The assignment of roles and allocation of tasks within the cooperative learning groups.
- e. The assessment of the process. This was performed by the students

who shared information about how they carried out the assignment and the involvement of the other participants (group self-assessment). The outcome was evaluated by the teacher (using the evaluation scale) and by the students. The eLearning platform “Aula Virtual” was used to ensure the anonymity of each assessment as a means to controlling the honesty of the responses given.

- a) Our study’s self-observation process was systematised by means of repeated data gathered during each of the practical assignments carried out in pursuance of subject contents and during monitoring of the assessment process (hetero-assessment, self-assessment and peer-assessment) with the instruments designed for the purpose.

Participants

119 first-year students taking the subject “Pedagogical Diagnosis and Observation Techniques in the Infant Classroom” as part of the bachelor’s degree in Infant Education of the University of Alcalá took part in the study, together with their teacher. The composition of the large groups (class/groups A, B and C) and subgroups (a total of 29 cooperative learning groups) are shown in Table 1.

Table 1.- Distribution of students in large groups and subgroups.

	N	Number of subgroups	Students means per subgroup
Group A	41	10	4.1
Group B	42	9	5.0
Group C	36	10	4.6

Instruments

The assessment and data collection tools used in this research were: a rubric for the peer assessment procedures; a tailored questionnaire for the purposes of the negotiated group self-assessment procedures; a tailored evaluation scale for the hetero-assessment. At the start of the term these three

tools were made available on the online campus “Aula Virtual”, which uses the *Blackboard Learning System* platform. The tools were discussed in face-to-face classes so that from the beginning of the academic year the students might have precise, detailed knowledge about their assessment and there

might be fluent communication throughout the process.

Peer assessment tools

The rubric evaluated the work group according to a series of categories with an influence on cooperative work. It was devised in line with Heathfield's *Group Assessment Scheme* (2003). It consisted of the following indicators:

- Regular attendance at group meetings.
- Contributing topic-related ideas.
- Research material.
- Analysis and preparation of topic.
- Support and motivation of the group members.
- Effective contribution to final outcome.

The rubric combine don a single quantitative and/or qualitative scale the criteria enabling the students' learning level to be evaluated and the progressive levels of performance for each task. Three evaluation categories (0-5-10) were proposed for describing the performance levels.

Self-assessment tools

The best approach for accomplishing the group self-assessment was thought to be a tailored questionnaire composed of open questions which would allow the students to reflect and justify from an emic perspective (Cochran-Smith and Lytle, 2002) their processes of work and learning within the group. The questionnaire had a first block of open questions for identifying the roles each member had assumed within the group and a second block for identifying the strong points and difficulties encountered in the course of the assessment activities, as well as the agreements reached and decisions taken by the group. After going through this process of reflection, the members of each cooperative work group had to reach consensus on the final mark (in base 10) for the whole set of activities performed as a group in the learning

folder about their perception of cooperative learning level attained.

Hetero-assessment tools

With a view to triangulating the sources of information, a tailored assessment scale was designed setting out the indicators the teacher would bear in mind when assessing the group assignments. This scale was made available on the eLearning platform at the start of the term so that the students could familiarize themselves with the indicators and bear them in mind when carrying out the group activities. The indicators were:

- Number of assignments handed in.
- Organization, development and clarity of the practical work.
- Reflections on what was gained from the practical work and its connection to theoretical issues.
- Bibliographical references or type of arguments used.
- Formal aspects: presentation.

These indicators were associated with four performance levels (very good/good/satisfactory/redo) on a grade 4 Likert-type scale.

Data analysis

These data selection and analysis techniques were employed:

- *Quantitative data analysis* of statistics with the aid of SPSS Statistics v.19.0 software. The analysis of related samples included descriptives, correlational analysis, related means comparison (T test) and effect size (Cohen's r and d).
- *Qualitative data analysis* of contents with the aid of NVivo v.10.0 software. The analysis was adjusted to the emerging categories related to the study's theoretical framework regarding learning assessment (Alonso-Tapia and Hernansaiz, 2013; Carless et al., 2006; Gallego, Sánchez and Cubero, 2011; Otero, Yuste and Alzás, 2011; Rodríguez and Ibarra, 2011;

Sanmartí, 2007). To avoid any subjectively arbitrary categorisation, the analysis was performed by two researchers who worked independently before pooling their results. Thus, two large categories, or category families, were established: *outstanding positive points* and *difficulties encountered*.

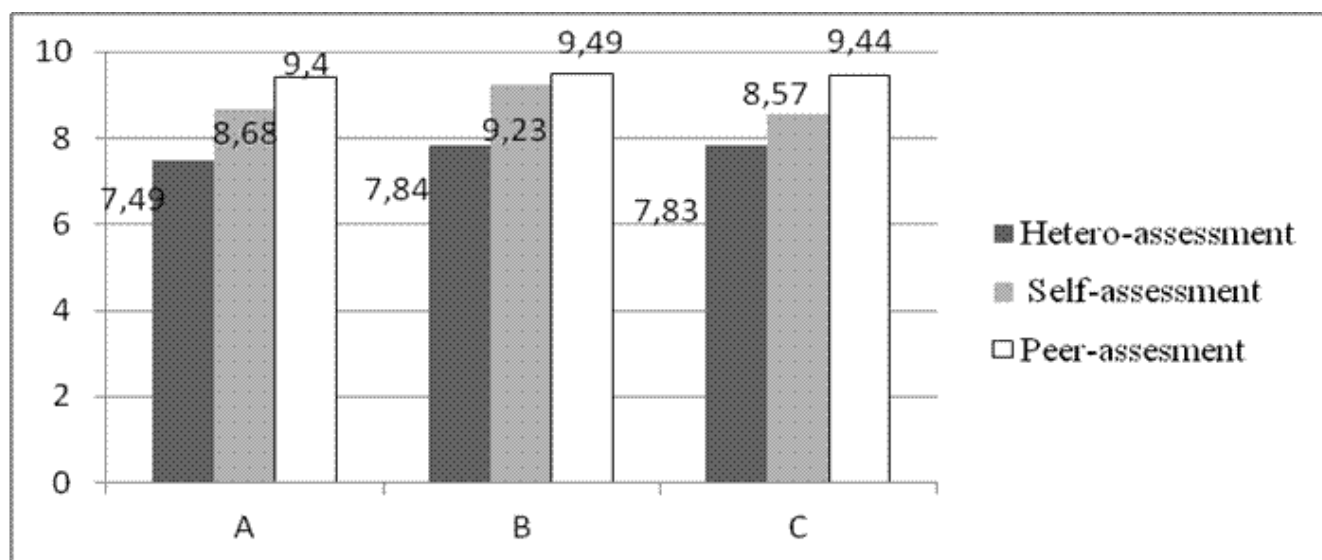
Reactivity was controlled by means of the students' and teacher's self-records, with no interference from the researchers.

Results

Level of agreement between different assessment types: convergent validity

Graph 1 shows how the average points students awarded to their companions were high (between 9.40 and 9.49 points out of a total of 10), with no statistically significant differences between groups ($F=0.59, p=.85$). The points deriving from the negotiated self-assessments were also high (between 8.57 and 9.23 points); in this case there were statistically significant differences as far as group B was concerned in so far as its points were higher than those of the other groups ($F=4.39, p=.00$). The teacher awarded lower marks (by between 1.5 and 2 points) with greater dispersion; in this case the differences between groups were not statistically significant ($F=.99, p=.37$).

Graph 1.- Average marks in three assessment types by sample group



As for the evaluation of the cooperative learning indicators, students evaluated most highly *support and motivation of group processes* (9.1 points); *materials contributed at meetings and sessions received* the lowest evaluation (8.4 points).

Table 2 shows the degree of association between the points students gave to their

companions for one assignment (peer assessment), the points awarded by the teacher (hetero-evaluation) and the points negotiated by the members of the group (self-assessment). In all combinations the correlations were weak and largely insignificant.

Table 2.- Correlations of related samples.

	N	r	p
Peer assessment and hetero-assessment	119	.201	.028
Peer assessment and self-assessment	119	.152	.099
Hetero-assessment and self-assessment	119	.022	.812

La Table 3 shows the comparison of means for these same points. This time statistically

significant differences were found for the three cases.

Table 3.- Means comparison (T-test for related means) for the different assessments and calculation of Cohen's statistic.

	Mean	Standard deviation	Typical error	t	gl	Sig.	d
Peer assessment and hetero-assessment	1.72	1.34	.12	14.006	118	.00	1.64
Peer assessment and self-assessment	.60	1.19	.10	5.527	118	.00	0.97
Hetero-assessment and self-assessment	-1.12	1.61	.14	-7.567	118	.00	0.66

According to Morales (2008), a statistically significant difference is not necessarily a large or an important one. Accordingly, we decided to calculate the *effect size*. The aim of this calculation is to determine whether statistically significant effects are important in the research's field of application. In the present case, we first calculated Cohen's d statistic before transforming it into a correlation coefficient. We obtained a mid-range correlation coefficient for the difference between peer assessment and hetero-assessment ($r=.64$), a lower one for the difference between peer assessment and self-assessment ($r=0.43$) and a more moderate one for the difference between self-assessment and hetero-assessment ($r=.31$). In the light of these results, students may be said to be less stringent evaluators and also somewhat inconsistent.

Concerning the process: emic perspective

As for the first large category family, *outstanding positive points*, nine analytical categories were determined in order to classify the information that emerged from the students' group self-assessment. Only units of analysis related to the theoretical-practical *assessment for learning* indicators were evaluated. The proportion of fragments codified in the content analysis for each analysis category as a percentage of total fragments codified in this category family is shown in Table 4. In their written remarks students made most frequent reference to *task allocation* (262) and *attendance and participation* (230); they also gave considerable thought to issues related to *values* (151) and *ideas* (111).

Table 4.- Percentage of self-assessment content fragments codified for each analytical category regarding “outstanding positive points”

Analytical Categories for Outstanding positive points	%
Good internal communication	1
Management of time	3
Ideas contributed to group	13
Attendance and participation	27
Task allocation	30
Practice of subject contents	1
Resources used	5
Value of teacher’s role	2
Value of cooperative learning	18
Total	100

Some issues arose concerning the contributions of different members of the groups in relation to *internal communication* as a source of cooperative learning. The following comment is one example:

Like S1 and S2, they have a lot to say and communicate well within the group, with good ideas about the activities and great capacity to make very interesting contributions, but what they find more difficult is to talk of express themselves in public, maybe because of their age.

This dimension is closely related to *ideas contributed to group*, a source of knowledge base don personal experiences: “brings to bear ideas related to teaching because of experiences of close family”; “shares experiences based on his little brother”.

Also related to the previous category are some references to knowledge management and decision-making: “before doing the practicals each of us gave some opinions and ideas and pooled them and then we organised the work according to the task”; “on the whole we all contributed ideas and talked about them as a group until reaching a final agreement about how to perform the activities”.

Also worthy of note are some comments in the *attendance and participation* category:

The decisions were taken in view of all the contributions and the work was done equitatively, with all of us benefiting and learning subject contents and always achieving the assignments own goals.

Although corresponding only to a tiny percentage, also worth highlighting is this favourable comment from the *value of teacher’s role* category regarding assistance from the teacher:

Maybe we all had some problem or other when doing one task or another, but the teacher was always to hand to help with any doubt or difficulty in the practicals.

As for the *difficulties encountered* category family, six categories were determined for analysis regarding *assessment for learning*. The proportion of fragments codified in the content analysis for each analysis category as a percentage of total fragments codified in this category family is shown in Table 5. In this case, what most worried students were issues related to *time* (75), along with various matters related to *activities* (62) and *participation* (42).

Table 5.- Percentage of self-assessment content fragments codified for each analytical category regarding “difficulties encountered”

Analytical categories for Difficulties encountered	%
Subject contents	6
Management of time	37
Resources used	4
Participation	21
Classroom spaces	1
Practice of subject contents	31
Total	100

Among the difficulties related to performing the practical activities during the course “analytical category *practice of subject contents*), students’ comments reveal two dimensions. Most frequently reference was made to the utility of carrying out activities aimed at analysing real cases with a view to forging a direct link between theoretical knowledge and practice: “we would have liked to do some practicals in an infant school, although we know it’s impossible”; “it would have been more interesting if we had been able to do it with a real case so that we’d have been more involved”. Efforts were made to satisfy this perceived need with the aid of genuine documents or videos as the current programmes of study do not make room for any performance of practical activities by first-year students of Infant Education in schools.

The second dimension had to do with the difficulty of the tasks and the capacity of the students to complete them successfully: “at the beginning we didn’t know how to set about the job because of its conceptual dimensions and its difficulty”.

Finally, there were numerous references to matters related to organizing the work and managing the time available for performing the activities (analytical category *management of time*): “we think that with a bit more time and correction in class we would have acquired the knowledge the practical were intended to equip us with”; “we think there

were too many practical activities for such a short time, and they all piled up on top of us”.

Discussion

The goal of this study was to learn more about the development of a cooperative learning experience, more particularly, about its process of participative assessment. To this end the two groups involved (teaching staff and students) were used as sources of information.

The results obtained from the data analysis allow us to conclude by emphasising the significant differences between the marks awarded by the teacher and by the students, the tendency being for the latter to give higher points in the assessment processes than the former. Whether in the *self-assessment* or in the *peer assessment*—and particularly in the latter—little difference was observed between the various indicators evaluated in the peer assessment. More detailed analysis gave as a result that the differences encountered in the processes of *peer assessment* were of little significance.

Our findings need to be considered in the light of the study’s methodological limitations: a sample drawn from a single measurement level of a single subject, degree programme and university allows little scope for generalisations, although the sample size of participating students did make for great variability within said measurement level. In line with the opinion of Panadero, Alonso-

Tapia and Huerta (2014) and previous research (Andrade, Du and Wang, 2008; Panadero and Jonsson, 2013; Sadler and Good, 2006), well designed rubrics were shown to contribute to enhancing academic performance, to sharpen differentiation of that performance, and to emphasise the assessment of the final outcome. According to Panadero et al. (2014), this last aspect “makes students pay more attention to their marks. [...] Therefore, rubrics may make students more concerned about their marks and activate performance and/or avoidance goals instead of learning goals” (p. 166). However, that research is not conclusive. In order to compensate that goal-achievement bias and to assess whether our study yielded results related to assessment for learning, the group self-assessment questionnaire was designed around open questions, the responses to which amount to evidence in the direction of the positive learning production. It may therefore be concluded that *assessment for learning* was achieved.

Notwithstanding our study’s limitations, that *joint assessment* is beneficial by no means seems unreasonable. Although such assessment was shown to be an imperfect system, the results for the indicators in the students’ emic perspective nonetheless demonstrated the benefits and utility of the process and assessment tools designed for our study. Analysis of the contents of the students’ reflections as gathered by means of the group self-assessment corroborated this assertion to the extent that our study’s efforts to let the data speak enabled student’s comments to be heard: the development of positive communication processes; enriched learning thanks to contributions from group members; development of basic social skills favouring self-regulation of group work (equitative task allocation) and producing “consensus”. In theory, all of this should foment group cohesion, but no positive evidence was collected in this regard. It was also interesting to note the positive evaluation students gave to the teacher’s availability as a back-up resource.

However, empirical evidence also emerged which questioned the good results of the process. The students asked for systems of external learning regulation in which attention from the teacher, the chance to perform practical activities in class time and a reduction in task difficulty take pride of place. Students found it difficult to regulate their own learning and looked for excuses to avoid doing it. The most suggestive result in this respect had to do with the commencement of the learning self-regulation process in the case of most participating students. This may be due in part to the teacher’s provision of spaces for individual and group reflection throughout this study’s joint assessment processes. Consequently, future research may consider improving assessment design and assessment instruments in order to redound in an assessment process geared towards learning on the part of students and teacher alike.

If the goal of university assessment is not only to evaluate cognitive competences but also to have account of procedural and attitudinal competences, students’ assessments of their own learning surely amount to a particularly important variable. In line with Ibarra et al. (2012), it was noteworthy that despite some problems, in the course of this experience the students managed to: (a) improve their learning processes and outcomes; (b) improve their interpersonal skills; and (c) experience for themselves the strengths and weaknesses of one of the teacher’s most complex professional competences, namely, assessment. For the students, assessment was transformed from a peripheral task to be performed at the end of a process to an integral part of the learning process itself in so far as students learn from reflecting on their own and their companions’ work (Gil and Padilla, 2009).

References

- Alonso-Tapia, J. & Hernansaiz Garrido, H. (2013). *Evaluar para el aprendizaje. Problemas en la práctica de la evaluación de competencias: el caso de la comprensión*

- de tablas y gráficos. Assessment for learning. Difficulties in competencies practices: the case of table and graph understanding. *Revista educativa Hekademos*, 13, Año VI, 7-18.
- Andrade, H., Du, Y. & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practices*, 27(2), 3-13.
- Boud, D. & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31 (4), 399-413. Available from http://pages.ramapo.edu/~vasishth/Learning_Outcomes/Boud+Assessment+Long-term_Learning.pdf
- Bretones Román, A. (2008). Participación del alumnado de Educación Superior en su evaluación. *Revista de Educación*, 347, pp.181-202.
- Brown, S. & Glasner, A. (2003). *Evaluar en la Universidad. Problemas y nuevos enfoques*. Madrid: Narcea.
- Carless, D., Joughin, G. & Liu, N.F. (2006). *How assessment supports learning: Learning-oriented assessment in action*. Hong Kong University-Press.
- Cochran-Smith, M. & Lytle, S.L. (2002). *Dentro / Fuera. Enseñantes que investigan*. Madrid: Akal.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self-assessments. *Assessment & Evaluation in Higher Education*, 11 (2), 144-166.
- Falchikov, N. y Boud, D. (1989). Student self-assessment in Higher Education: a meta-analysis. *Review of Educational Research*, 59 (4), pp.395-430.
- Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International*, 32, 175-187.
- Falchikov, N. (2005). *Improving Assessment Through Student Involvement. Practical solutions for aiding learning in higher and further education*. London: Routledge-Falmer.
- Gallego, B.; Sánchez, C. & Cubero, J. (2011). La pro-alimentación y la retro-alimentación prospectiva entre iguales como estrategias para el desarrollo del pensamiento crítico y el aprendizaje autónomo. En Evalfor, *Evaluar para aprender en la universidad: Experiencias innovadoras en el aprendizaje a través de la evaluación* (pp. 83-97). Madrid: Bubok Publishing.
- Gil Flores, J. & Padilla, T. (2009). La participación del alumnado universitario en la evaluación del aprendizaje. *Educación XXI*, 12(1), pp. 43-65.
- Heathfield, M. (2003). Evaluación en grupo para fomentar un aprendizaje de calidad. In S. Brown, A. Glasner (Eds.) *Evaluar en la universidad. Problemas y nuevos enfoques*, (pp.155-166). Madrid: Narcea.
- Ibarra Saiz, M.S., Rodríguez Gómez, G. & Gómez Ruíz, M.A. (2012). La evaluación entre iguales: beneficios y estrategias para su práctica en la universidad. *Revista de Educación*, 359, 206-231.
- López Pastor, V.M. (2008). Desarrollando sistemas de evaluación formativa y compartida en la docencia universitaria. Análisis de resultados de su puesta en práctica en la formación inicial del profesorado. *European Journal of Teacher Education*, 31(3), 293-311.
- López Pastor, V.M. (Coord.) (2009). *Evaluación formativa y compartida en Educación Superior. Propuestas, técnicas, instrumentos y experiencias*. Madrid: Narcea.
- Montero-Fleta, B. (2006). Investigación cualitativa y cuantitativa en la coevaluación: correlación profesor-alumno. In F. Watts & A. García-Carbonell, (Eds.),

- La evaluación compartida: investigación multidisciplinar*, (pp. 31-46). Universidad Politécnica de Valencia. Available from <http://www.upv.es/gie/Publicaciones.html>
- Morales Vallejo, P. (2008). *Estadística Aplicada a las Ciencias Sociales*. Madrid: Universidad Pontificia de Comillas.
- Muñiz, J. & Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación*, 5, 13-25.
- Navaridas Nalda, F. (2002). La evaluación del aprendizaje y su influencia en el comportamiento estratégico del estudiante universitario. *Contextos Educativos*, 5, 141-156.
- Otero, M., Yuste, R. & Alzás, T. (2011). Coevaluación en la plataforma virtual Moodle: una propuesta práctica para la mejora del proceso de enseñanza en Educación Superior. In Evalfor, *Evaluar para aprender en la universidad: Experiencias innovadoras en el aprendizaje a través de la evaluación*, (pp. 62-71). Madrid: Bubok Publishing.
- Panadero, E. & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. DOI: [10.1016/j.edurev.2013.01.002](https://doi.org/10.1016/j.edurev.2013.01.002)
- Panadero, E. & Alonso-Tapia, J. (2013). Autoevaluación: Connotaciones teóricas y prácticas. Cuándo ocurre, cómo se adquiere y qué hacer para potenciarla en nuestro alumnado. *Electronic Journal of Research in Educational Psychology*, 11 (2), 551-576. Disponible en: <http://www.redalyc.org/articulo.oa?id=293128257012>
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on construct validity of peer assessment, perceived fairness and comfort, and performance. *Studies in Educational Evaluation*, 39(4), 195-203. DOI: [10.1016/j.stueduc.2013.10.005](https://doi.org/10.1016/j.stueduc.2013.10.005)
- Panadero, E., Alonso-Tapia, J. & Huertas, J.A. (2014). Rubrics vs self-assessment scripts: effects on first year university students' self-regulation and performance / Rúbricas y guiones de autoevaluación: efectos sobre la autorregulación y el rendimiento de estudiantes universitarios de primer año. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 37(1), 149-183. DOI: [10.1080/02103702.2014.88165555](https://doi.org/10.1080/02103702.2014.88165555)
- Pond, K., Ul-Haq, R. & Wade, W. (1995). Peer review: a precursor to peer assessment. *Innovation in Education and Training International*, 32, 314-323.
- Roberts, T. S. (2006). *Self, Peer and Group Assessment in E-Learning*. Hershey, PA: Information Science Publishing.
- Rodríguez Gómez, G. & Ibarra Saiz, M.S. (Eds.) (2011). *e-Evaluación orientada al e-Aprendizaje estratégico en Educación Superior*. Madrid: Narcea.
- Rodríguez Gómez, G., Ibarra Sáiz, M.S. & García Jiménez, E. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de Investigación en Educación*, 11 (2), 198-210. Recuperado de <http://dialnet.unirioja.es/servlet/articulo?codigo=4734976>
- Sadler, P.M. & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), pp. 1-31.
- Sanmartí, N. (2007). *Evaluar para aprender: 10 ideas clave*. Barcelona: Graó.
- Topping, K.J. (2009). Peer assessment. *Theory Into Practice*, 48, 20-27.

ANNEX 1

Instrumentos utilizados

Peer assessment rubric

	Student1	Student2	Student 3
Regularly attends group meetings			
Contributes ideas for the assignment			
Brings research material, analyses and prepares the topic			
Contributes to the group's cooperative processes			
Supports and motivates group members			
Contributes to final result and presentation			
Observations			

10	5	0
Attendance at group meetings		
Attends all face-to-face and virtual meetings, remaining to the end, working actively and attentively all the time, and ready to be flexible over times.	Missed some meetings, lack of interest in group work.	Missed most meetings. Often or always arrives late, inattentive or wastes most of the time.
Contribution of ideas for the assignment		
Thinks about the topic before the meeting, brings ideas about how to work, acts on the suggestions of others and is ready to challenge ideas rather than to remain silent.	Makes no prior preparations. Contributes one or two ideas. Neither opposes nor discusses ideas.	Makes no prior preparations. Contributes no ideas. Tends to reject the ideas of others rather than challenge them.
Research material, analysis and topic preparation		
Does what he/she is told, brings materials, shares out research equitatively, helps to analyse and evaluate material.	Does not always do what he/she is told, does not bring materials or share out tasks equitatively but usually gets the better of the sharing out.	Does no research. Does not do what he/she undertook to do. Does not manage his/her workload. Does not get involved in assignments and lets everybody else provide all the material.
Contribution to group's cooperative processes		
Puts aside personal differences, wishes to review progress and avoid conflict, adopts various roles as necessary, keeps the group on the right track, is flexible but topic-centred.	Puts aside personal differences. Always adopts the same role (leader, joker, etc.) regardless of circumstances. Does not take the lead in reviewing and helping the work along.	Does not take the imitative but waits to be told what to do. Always adopts the same role (leader, joker, etc.) regardless of the circumstances, creates conflict and is unwilling to review the group's progress.

10	5	0
Support and motivation of the group's members		
Is a good listener, encourages participation, adapts to a cooperative learning environment, is sensitive to issues affecting group members, supports group members with special needs.	Only cares about finishing the task. Listens to the rest but ignores their opinions when working; does not motivate or support companions' participation.	Only cares about finishing the task, speaks while others are talking and ignores their opinions. Is insensitive to individual needs and makes no contribution to the learning process.
Effective contribution to final outcome: presentation		
Willing to try out new things. Does not spoil tasks, makes quality contributions, uses own initiative, is convincing and produces high-quality work/presentation.	Gets involved in some tasks, making mediocre contributions. Does not use own initiative but produces work of satisfactory quality, doing what others tell him/her to do.	Unwilling to get involved in any task, shuns responsibility and is unconvincing. Makes a poor and limited contribution.

(Adapted from Heathfield, 2003, p. 162)

Self-assessment questionnaire

The aim of this questionnaire is to find out how the group work experience evolved. It has four blocks which ask you to reflect on:

- a) The common goals achieved by the group. It mentions difficulties encountered in the course of the work and how solutions were found.
- b) Agreements made and decision taken, above all, decisions taken as a group.
- c) The individual contributions of each group member.
- d) A negotiated group mark (base 10).

Participants	Responsibilities
Block 1: Difficulties encountered	
Block 2: Agreements and decisions	
Block 3: Contributions	
Group member	Contribution
Block 4: Group mark (negotiated)	
Other observations	

Teacher's evaluation scale

Activities delivered							
Level 1 Very good		Level 2 Good		Level 3 Satisfactory		Level 4 Redo	
Organisation, development and clarity of activities							
Level 1 Very good		Level 2 Good		Level 3 Satisfactory		Level 4 Redo	
Reflections on what was gained from practice and its connection to theory							
Level 1 Very good		Level 2 Good		Level 3 Satisfactory		Level 4 Redo	
Bibliographical consultations							
Level 1 Very good		Level 2 Good		Level 3 Satisfactory		Level 4 Redo	
Formal aspects							
Level 1 Very good		Level 2 Good		Level 3 Satisfactory		Level 4 Redo	

Authors

To know more / Saber más

Pascual-Gómez, Isabel (isabel.pascualg@uah.es).

Holds a doctorate in Philosophy and Educational Sciences from the Complutense University of Madrid. At present she is a non-tenured lecturer in the University of Alcalá's Department of Educational Sciences, where she teaches Research and Diagnostic Methods in Education. Her chief lines of research include university teaching assessment and educational psychology tools validation. Her postal address is: Aulario María de Guzmán. C/ San Cirilo s/n. 28804. Alcalá de Henares (Spain).



Lorenzo-Llamas, Elena María (elenam.lorenzo@uah.es).

Doctorate in Educational Psychology from the University of Alcalá. At present she is an associate lecturer in the University of Alcalá's Department of Educational Sciences, where she teaches Research and Diagnostic Methods in Education. Her chief lines of research include high intellectual capacities, educational research methods, catering for diversity and inclusion, and higher education assessment methods. Her postal address is: Aulario María de Guzmán. C/ San Cirilo s/n. 28804. Alcalá de Henares (Spain).



Monge-López, Carlos (carlos.monge@edu.uah.es).

Is a doctoral student of Educational Planning and Innovation. On the research staff of the University of Alcalá's Department of Educational Sciences (area: Didactics and School Organization), he holds a pre-doctoral FPU-UAH Doctoral Training contract. His postal address is Facultad de Educación. C/ Madrid 1. 19001. Guadalajara (Spain).





Revista ELectrónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).