

Summative self-assessment in higher education: implications of its counting towards the final mark

**Ricardo A. Tejeiro¹, Jorge L. Gómez-Vallecillo¹,
Antonio F. Romero¹, Manuel Pelegrina², Agustín Wallace²,
Enrique Emberley³**

¹ Departamento de Psicología, University of Cádiz, Cádiz

² Dpto. de Psicobiología y Metodología de las Ciencias del Comportamiento,
University of Málaga, Málaga

³ Centro de Profesorado del Campo de Gibraltar, La Línea de la Concepción,
Cádiz

Spain

Correspondence: Ricardo A. Tejeiro. Departamento de Psicología, Facultad de Ciencias de la Educación, Universidad de Cádiz, Campus Río San Pedro, 11519 Puerto Real, Cádiz, Spain. E-mail: ricardo.tejeiro@uca.es

© Education & Psychology I+D+i and Editorial EOS (Spain)

Abstract

Introduction. Our study aims at assessing the validity of summative criteria-referenced self-assessment in higher education, and in particular, if that validity varies when the professor counts self-assessment toward the final mark.

Method. One hundred and twenty-two first year students from two groups in Teacher Education at the Universidad de Cadiz (Spain) participated in the study, each group following a different course in the same six-month period. The relevant difference between the two courses was that, in one of them, self-assessment was included among the assessment methods and counted for 5 per cent of the final mark. The professor was the only marker in the other course. Once the courses finished, participants completed a self-assessment questionnaire that included competences, learning results and contents, and were asked to give themselves a mark. Self-assessment data were compared with the marks given by the professor, and the 10 students with the highest discrepancies were interviewed.

Results. In both groups, the students' self-assessments were higher than the marks given by the professor, with significantly higher differences in the students with poorer results and in the group in which self-assessment counted towards the final mark. In this group, no relationship was found between the professor's and the students' assessments.

Discussion. When self-assessment does not count towards the final mark, students' and professor's assessments tend to be highly similar; when self-assessment counts towards the final mark, over and underestimations increase dramatically. The main reasons that we found for this are the desire to obtain the highest possible grades and the stress associated with self-assessment. Implications of implementing self-assessment in higher education are discussed.

Keywords: Education; Learning; Self-assessment; University; Educational Psychology.

Received: 03/03/11

Initial acceptance: 05/16/12

Final acceptance: 07/18/12

La autoevaluación sumativa en la enseñanza superior: implicaciones de su inclusión en la nota final

Resumen

Introducción. Se presenta un estudio que tiene por objetivo comprobar la validez de la autoevaluación sumativa criterial en la enseñanza superior y, de forma particular, si en dicha validez influye el que la autoevaluación sea considerada para elaborar la nota final.

Método. Participaron en la investigación 122 alumnos de primer curso del grado de Educación Infantil de la Universidad de Cádiz (España) pertenecientes a dos grupos diferentes, cada uno de los cuales cursaba, en el mismo periodo semestral, una asignatura distinta. La diferencia relevante entre ambos grupos era que en una de las asignaturas se incluía la autoevaluación entre los procedimientos de evaluación y se le asignaba un peso del 5% sobre la nota final; en la otra asignatura, toda la nota era asignada por el profesor. Al concluir el período lectivo, los participantes completaron un cuestionario de autoevaluación de competencias, resultados de aprendizaje y contenidos, y se asignaron una calificación. Sus respuestas fueron comparadas con la nota asignada por el profesor, y se entrevistó a los 10 alumnos cuyas auto-calificaciones mostraron mayor divergencia.

Resultados. En ambos grupos, el alumnado tendía a sobreestimar sus resultados respecto a las calificaciones asignadas por el profesor, incrementándose la diferencia de forma significativa en los alumnos con peor calificación y en el grupo cuya autocalificación influía en la nota final. En este grupo, de hecho, la calificación por el profesor y la autocalificación no guardaron ninguna relación.

Discusión. Cuando la autoevaluación no influye en la nota, el alumnado se evalúa de una manera muy similar a como lo hace el profesor, pero cuando la autoevaluación influye en la nota, la discrepancia se incrementa, notablemente, tanto por exceso como por defecto. Los principales motivos que se encuentran para ello son el deseo de obtener la máxima calificación posible y la presión añadida de tener que calificarse a uno mismo. Se analizan las implicaciones de cara a la aplicación de la autoevaluación en la enseñanza superior.

Palabras Clave: Educación; Aprendizaje; Autoevaluación; Universidad; Psicología de la Educación.

Recibido: 03/03/11

Aceptación inicial: 16/05/12

Aceptación final: 18/07/12

Introduction

Self-assessment constitutes one of the most interesting topics in the literature on assessment in higher education. There are a variety of reasons for this, including the desire to foster greater democratization of the relationships between students and professors as well as the consolidation of autonomy and self-regulation as the basis for learning and professional competence (Boud, 1989; Dochy, Segers & Sluijsmans, 1999; Sluijsmans, Moerkerke & Dochy, 1998).

Several authors have suggested that self-assessment makes it easier for the student to monitor, direct, and regulate actions toward goals of information acquisition, expanding expertise, and self-improvement (Paris & Cunningham, 1996; Paris & Paris, 2001), therefore constituting a fundamental element for effective learning both during and after the university years (Black & William 1998; Taras 2001). Nevertheless, despite this apparent consensus, there is a vast array of definitions and a wide variety of forms in which self-assessment is conceived and implemented.

Firstly, English-speaking authors usually distinguish among *self-assessment*, *self-evaluation* and *self-grading*. *Self-assessment* refers to the process of formative assessment that allows the students to reflect on the quality of their own work and learning, to judge the degree in which it reflects the established criteria and to identify their strengths and weaknesses as a basis for improvement (Andrade & Boulay, 2003; Goodrich, 1996; Gregory *et al.*, 2000; Hanrahan & Isaacs, 2001; Paris & Paris, 2001; Tan, 2008). Usually, *self-assessment* does not imply the assignment of marks by the student, which is a common feature of *self-evaluation*, defined by Andrade and Du (2007) as the process that involves the students in making summative judgments about their work, leading to the assignment of a mark (*self-grading*). In the literature in Spanish, *auto-evaluación* has been by far the most extensively used term, which leads to confusions when its meaning is not sufficiently defined.

Aside from terminology, noteworthy differences exist in the way in which self-assessment is conceived. Kelvin Tan interviewed 16 professors from diverse disciplines and found 5 different conceptions, depending on what the students are to judge in each case (Tan, 2008): (a) their own self-assessment behavior; (b) their knowledge of self-assessment practices; (c) the criteria in the study program; (d) their improvements within the study program (the

students' judgments are only considered to benefit their improvement within the program); and (e) their general self-assessment skills, not limited to the study program.

One example of what might be considered as the “f” conception –in the line of Tan (2008) but not cited by him– will be presented in our study, this is, involving the students in the judgment of their improvements within the study program while directly counting their judgments towards the final grades. There are two possibilities, depending on whether the students' judgment is the only source of marks or just a part of them; we will deal with the second possibility, which is included in the teaching program of one of our subjects. This article thus analyzes a case of summative criteria-referenced self-assessment. The expression “criteria-referenced” refers to the existence of a previously specified educational objective, in our case included in the teaching program of the subject (Popham, 1983). It is summative because it is applied at the end of the study program for the purpose of justification and accountability (Scriven, 1967).

Both pros and cons are found in the literature regarding this type of self-assessment. The first include complementing the professors' marks, the promotion of a more adjusted self-concept, promoting in the students an attitude more open to external judgments and to the critical self-analysis of their own performance, and reducing the perceived social distance between professor and student. Limitations include the lack of objectiveness, the tendency to consider oneself as a good student or professional, the difficulty to quantify the assessment, the possibility that assessment might become a form of self-justification, and the existence of individual differences in self-assessment skills (Barber, 1997).

Our study is partly based on the comparison between the marks given by the professor and by the student, which in fact constitutes the most common procedure in the studies on self-assessment (Boud & Falchikov, 1989; Brew, 1999; Falchikov, 2005; Falchikov & Boud, 1989; Gruppen et al., 1997; Hafner & Hafner, 2003). The validity of the marks given by the students has been usually measured by the degree of similarity with the marks given by the professors (Tan, 2004). Also, assessment criteria use to be exclusively defined by the professor or the academic institution (Falchikov & Boud, 1989; García & Floyd, 1999; Hanrahan & Isaacs, 2001; Longhurst & Norton, 1997), disregarding the students' tendency to include their own criteria, such as time spent or effort, as revealed by some studies (Sullivan & Hall, 1997;

Taras, 2001). Some scholars thus defend the convenience of involving both professors and students in the setting of criteria (Dochy & McDowell, 1997; Stallings & Tascione, 1996).

Two differences with regard to previous studies were included in our work. Firstly, considering that from the early 1990s both the professional and the academic profiles tend to be defined in terms of competences (Lucas, 2007), we widened self-assessment to include the students' self-grading as well as their judgments about their perceived acquisition of competences, learning results and contents. Secondly, we compared self-assessment in two different groups; in only one of them, self-grading counted towards the final marks. Lew et al. (2010) suggested that accuracy in self-assessment might be associated with the way in which the students perceive the assessment process, although in their work they focused on the perception of the degree in which self-assessment influences learning. In our study, it seems reasonable to expect that those students whose judgments count towards their final marks will be more motivated for self-assessment and therefore will perform it more accurately. We are conscious that there is at least an alternative hypothesis: that these students will be more tempted to over score their achievements, and thus their judgments might be less related to those of their professor.

The present work was performed in the Faculty of Education in Universidad de Cádiz (Spain), where 25 out of the 61 degree subjects in 2011/2012 included self-assessment, which represents 41% of the total (50% in Childhood Education, 42.9% in Primary Education, 40% in Physical Activity and Sports and 20% in Psychology). There are substantial variations in the way self-assessment is considered in the different teaching programs. In short: only 4 out of the 25 subjects specified tasks and techniques for self-assessment and quantified the weight of self-assessment towards the final marks. In our opinion, this reflects the lack of common criteria and the existence of certain confusion with regards to self-assessment in higher education. This was confirmed in several informal contacts with the professors of these subjects.

Objectives and hypotheses

There are still few studies about the validity of self-assessment in Spanish universities or about the possible effect of counting self-assessment towards the final marks. Also, although numerous studies have asked the students to assess their achievements both quantitatively (assignment of marks) and qualitatively (describing their strengths and weaknesses),

there is a noteworthy scarcity of studies on self-assessment of competences, these defined as in the teaching programs.

Therefore, we established four objectives: (1) to assess the correlation between the students' self-assessments and the marks given by the professor; (2) to identify possible explanations for the differences between the students' and the professors' assessments; (3) to analyze the direct self-assessment of competences, learning results and contents; and (4) to advance in the knowledge of forms to improve the students' reflections and knowledge of their own performance.

A review of the literature suggests the following hypotheses: (1) the marks self-assigned by students will be similar to those given by their professor; (2) the similarity between the students' and the professor's assessments will be higher when the former are asked to give themselves a mark between 0 and 10 (which is the usual marking procedure in most of their academic life) than when they are asked to assess their acquisition of competences, learning results and contents, because these aspects are less familiar and less concrete; (3) there will be a higher tendency to overestimate than to under estimate one's own achievements; (4) self-assessment will concur more with the professor's assessment in students with higher marks than in those with poor marks; and (5) counting self-assessment towards the final grades will be associated with a higher agreement between the self-assigned mark and the mark given by the professor. The studies that lead us to these hypotheses will be commented in the results discussion.

Method

Participants

Two groups of 60 and 62 students from the degree in Childhood Education in the Faculty of Education in Universidad de Cádiz (Spain) participated in the study. A vast majority of the participants in both groups were females (95% and 95.2% respectively); their ages varied from 18 to 32 years ($M=20.98$; $SD=3.41$ and $M=21.82$; $SD=3.70$, respectively). Although both groups were following the same studies with the same subjects, self-assessment was analyzed in the *Tutoría y Familia* (Tutorship and Family) course in one of them (further referred to as TYF) whereas the other (PSE) was analyzed in the *Psicología de la Educación* (Educational Psychology) course. Both courses are received in the first semester of the first year,

with 14 teaching weeks that include a session of 2.5 hours (with the whole group) and two sessions of 1 hour (half the group attending to each session). The same professor (first author) gave both courses, following the same teaching methods: lectures with active participation of the students, individual practices consisting of bibliographic research and group practices consisting of the preparation and exposition of contents.

The procedure for the assessment of learning results was the fundamental difference between both groups. In the PSE group, marks were exclusively given by the professor, weighing the results of the theory exam (60%), practices (30%) and participation of the student in class (10%). In the TYF group, marks were calculated with the same procedures (counting 60%, 30% and 5% on the final mark, respectively), but also with 5% of self-assessment based, according to the teaching program of the subject, in an “additional report about the student’s learning provided by the student or the group”.

Instruments

An ad hoc self-assessment questionnaire was designed in which the students were asked to indicate, on a subjective 5 points scale (from “very low” o “very high”) the degree in which they believed that they had acquired each of the competences, learning results and contents included in the program of the course. Each item was written in the same terms used in that program. The questionnaire for the TYF group included 1 competence, 4 results and 10 contents, whereas the questionnaire for the PSE group included 1 competence, 6 results and 9 contents. The students were also asked to indicate, on a scale from 0 to 10 points, the final mark that they thought they should receive. Annex 1 shows the questionnaire for the PSE group.

Procedure

The questionnaires were applied at the end of the teaching period, in the same session of the theory exam and before it. At the beginning of the session, the researcher explained briefly the purpose of the self-assessment questionnaires. Students in group PSE were told that the objective was to gain a better knowledge of the students’ perception on their own learning in order to improve future editions of the course. The same explanation was given to group TYF with the addition that, as indicated in the teaching program, the self-assigned marks would directly count for 5% of the marks. This is, assigning oneself 10 points would add 0.5 points to the mark given by the profesor (which ranged from 0 to 9.5); self-assigning

0 points would add 0 to the mark of the professor; the other self-assigned marks would weigh proportionally between 0 and 0.5 points.

Marking of theory exam, practices and participation was performed before the revision of the self-assessment forms, in order to ensure the independence between the marks given by the professor and the self-assessments.

Once the marking process was finished (including the communication of marks to the students and the review of exams when demanded), the 10 students with higher differences between self-marking and the marks given by the professor were interviewed. The sessions were video recorded. The interviewer was the professor of the course, which implies a certain risk of social desirability. Nevertheless, the existence of a previous personal relationship between the interviewer and interviewed is considered as a valuable aspect in quantitative research (Rubin & Rubin, 1995). Further, the possible negative effects of social desirability were minimized by the fact that the interviews were carried out after the conclusion of the semester, when the students knew that the professor would not give them more classes in the present course and, given the structure of the degree, probably neither in the future.

Data analysis

The statistical analysis of data was performed using the SPSS (version 15.0) program. For the statistical analysis, the 5% of self-marking was subtracted from the total marks obtained by the TYF group. Since the maximum score was then 9.5 points, all marks were multiplied by 1.053 to convert them into a 0-10 points scale.

Pearson's r correlation coefficient was utilized to know if there were differences between the self-assigned marks and the marks given by the professor. Since a part of the professor's marks was subjective (assessment of practices), two successive analysis were performed utilizing total marks and scores in the written exam.

Means in marks and self-grading in each group were compared with Student's T . The correlation between marks and discrepancies was calculated to assess the possibility of differences between the students with higher and with lower marks. Finally, in order to know if the students are more accurate in their self-assessment when they are asked to give themselves 0-10 points than when they are asked to assess their achievement of competences, learning re-

sults or contents, the scores in learning results items were added and grouped in a single weighed score; the same was performed with contents items (this was not necessary with competences because they included a single item). The correlations among scores in the 3 resulting variables, self-assessment and marks were then calculated for both groups.

Results

A positive relationship between self-grading and the marks given by the professor was found for group PSE but not for TYF, both when total marks and only scores in the written exam were considered (table 1).

Table 1. Relationship between self-grading and marks given by the professor

Group	Mark	<i>r</i>	<i>P</i>
TYF	Total	.099	.454
	Exam	.270	.838
PSE	Total	.331	.009
	Exam	.351	.005

Table 2 shows means and standard deviations of marks and self-gradings. In both groups, self-grading provided higher scores than the marks given by the professor ($M = .111$, $SD=1.133$ for PSE; $M = .280$, $SD=1.293$ for TYF), but differences were not statistically significant ($t_{(61)}=.774$, $p=.442$ for PSE; $t_{(59)}=1.676$, $p=.099$ for TYF).

Differences between self-grading and the marks given by the professor ranged from +4.1 to -2.4 points in group TYF, and +2.3 to -1.1 in group PSE. The students with higher and with lower marks differed in the degree of discrepancy, both in TYF ($r=.705$; $p=.000$) and in PSE ($r=.473$; $p=.000$). In both groups, differences increased (higher over-grading) as marks decreased.

Table 2. Differences between self-grading and marks given by the professor

Group	Marker	Mean	St.Dev.
TYF	Student	7.72	.922
	Professor	7.44	1.00
PSE	Student	6.89	1.05
	Professor	6.78	.887

In both groups, the self-assigned mark was significantly related with the reported acquisition of competences, learning results and contents (table 3). Regarding the relationship between these and the marks given by the professor, we found positive significant correlations in PSE for learning results and contents, but not for competences.

Table 3. Relationship of marks and self-grading with self-assessment of competences, learning results and contents

Factor	Group	Assessment of	<i>r</i>	<i>P</i>
Marks	TYF	Competence	.172	.189
		Learning results	.007	.960
		Contents	.066	.618
	PSE	Competence	.086	.509
		Learning results	.440	.000**
		Contents	.282	.027*
Self-grading	TYF	Competence	.377	.003**
		Learning results	.543	.000**
		Contents	.594	.000**
	PSE	Competence	.427	.001**
		Learning results	.520	.000**
		Contents	.499	.000**

Regarding the qualitative analysis of interviews, it should be noted that all 10 students with higher differences between marks and self-grading were included in group TYF. The reasons alleged for over-grading were: (a) difficulty to appropriately assess their own acquisitions; (b) desire to obtain the highest possible mark; and (c) not knowing the reason. Those who under-estimated their marks gave the following reasons: (a) difficulty to appropriately assess their own acquisitions; (b) avoidance of giving a selfish image; and (c) becoming stressed.

Discussion

Is there a relationship between the students' self-grading and the marks given by the professor? When self-assessment does not count towards the final marks, our results suggest a positive answer. This is in line with previous findings (Boud, 1986; Falchikov & Boud, 1989; Grupper et al., 1997; Hafner & Hafner, 2003; Sullivan & Hall, 1997; Taras, 2001). It is not an unanimous result, though. Simon Cassidy (2007) compared self-grading and professor's marking in a sample of first year undergraduate students and found a low although significant positive correlation ($r = .25$). He concluded that a majority of the students showed good self-

assessment skills that were not present in a quarter of his sample. Correlations between self-grading and the professor's marks in other studies with wide samples stand around $r = .20$ (Lew et al., 2010); in their discussion, the authors conclude that students tend to show a rather poor performance when judging their own learning process.

Is the relationship between self-grading and the professors' marking influenced by the way in which self-assessment is accomplished? Our data yield a negative response. In both groups, self-grading was significantly related with the reported acquisition of competences, learning results and contents. Regarding the relationship between self-assessment of these aspects and the mark given by the professor, group PSE provided significant positive correlations for learning results and contents but not for competences. These results may be due to the fact that competences are written in a more general manner, or perhaps they are influenced by the number of items for each aspect: only one for competences while 4-6 for learning results and 9-10 for contents.

There are few references to this topic in the literature. Fitzgerald, White and Gruppen (2003), for instance, found that students showed higher skills to self-assess their knowledge than to assess their clinical abilities, which are more difficult to define operatively.

Do the students tend to over-estimate rather than to under-estimate their achievements? We found such a tendency, in despite of the high coincidence between the students' and their professor's marking. These results are in line with reports from several authors (Falchikov & Boud, 1989; Lew et al., 2010; Sullivan & Hall, 1997) but not with others (Andrade & Du, 2007; Cassidy, 2007; Taras, 2001). Cassidy (2007), for instance, found 56% of under-estimation and 40% of over-estimation.

Are the students with higher marks more accurate in self-assessment? Such is the result of this and other studies (Falchikov & Boud, 1989; Lew et al., 2010; Orsmond, Merry & Reiling, 1997; Sullivan & Hall, 1997). This seems a reasonable result, because the ability to follow and judge one's own learning processes, performance, strengths and weaknesses is one of the factors that influence the students' competence (Lew et al., 2010).

Is the accuracy in self-assessment influenced by its counting towards the final grades? Yes, as revealed by the analysis of marking differences and by the fact that the correlation

with the professor's marks dramatically decreased in group TYF until $r = .10$. Nevertheless, against our expectations, we didn't find a simple tendency to over-grading oneself, but rather both over and under marking. This is, we found a general loss in assessments' reliability.

It has been suggested that the low correlation between the students' and their professors' markings may be due to the fact that the first don't have enough access to their own learning process (Lew et al., 2010). The lack of experience might also account for the low correlation: it has been reported that last years' students provide more accurate self-assessments than those in the first years (Falchikov & Boud, 1989; Gibbs, 1995). Nevertheless, both factors are common to both groups in our study: they may therefore have reduced the correlations in both, but they don't explain the differences between them.

The results of the interviews may provide an answer. In fact, they reveal two main reasons for the increase of differences when the students know that their self-grading will count towards the final marks:

(a) The simple desire to increase their marks. This result relates to those of Maguire, Evans and Dyas (2001), who found that a sample of first year undergraduates faced self-assessment with a "strategy" directed to obtaining good results with the minimum effort. Also, Lew and Schmidt (2006) reported that many students think that they can use their self-assessments to affect the way their professor assesses their achievements. As Andrade and Du (2007) concluded, if a teacher asks his or her students to assess themselves, and counts those assessments towards the final mark, then the most awoken and qualification-motivated students will assign themselves an A.

(b) The added stress when one assesses him or herself. Taras (2001) pointed out that the requirement of self-assessment generates different reactions, the worse students being more worried than the best ones. Many complain that they don't have neither the experience nor the knowledge, and express their concern about the possibility that their self-assessment will not coincide with that of their professor. Hanrahan and Isaacs (2001) remarked the doubts of the students regarding the reference criteria, and Andrade and Du (2007) found that initial reactions to self-assessment included uneasiness and the feeling of being unable to perform it. Frequently, the students assessed themselves according to the perceived expectations of their professor.

Implications for the use of self-assessment

The fact that differences increase when self-assessment counts towards the final marks does not invalidate the first of our results: when that is not the case, students are able to assess themselves at least as reliably as their professor. This is a relevant result for the implementation of formative self-assessment activities in the class-room, with the double purpose of facilitating the students' tracking of their own progresses and training their self-assessment and self-control skills (Falchikov & Boud, 1989). Self-assessment is therefore feasible when it is utilized as a tool for the students to critically review their own work (Andrade & Du, 2007). Furthermore, students consider that this type of self-assessment provides benefits including an improvement in the understanding of the assessment process, the development of critical thinking and higher empathy with the professors, as well as and an increase in the motivation for improvement (Hanrahan & Isaacs, 2001).

According to Kraayenoord and Paris (1997), one of the purposes of the true assessment –which includes self-assessment– is to encourage the students to a more active involvement in the following and revision of their own progresses. In this regard, undergraduate university students, at least in the first year, keep a solid mental schema in which the role of the professor is eminently productive (of contents, proposals, individual challenges and marks) whereas the student is basically receptive. The regular practice of formative self-assessment may contribute to the replacement of this passive conception of the role of the student by a more active schema.

On the other hand, if it is true that, as suggested by numerous authors, the training in the ability to judge one's own progresses constitutes a fundamental skill for the effective learning and the future professional development (Boud, 1986; Dearing, 1997; Falchikov, 1997; Stefani, 1998; Tan, 2008), it is still more relevant in the case of those who, as our participants, follow studies with the purpose of becoming teachers. If the assessment of academic achievements will be a fundamental part of their future job, won't they benefit particularly from the regular practice of self-assessment?

Some suggest a negative answer to this question. The results in Lew et al. (2010) seem to indicate that self-assessment is not learned through practice, which was also the result of

Eva et al. (2004). Others report just the opposite: self-regulation in general can be trained with explicit instruction, reflexive meta-cognitive discussions and modeling (Álvarez, 2009; Eissa, 2007), and Dochy, Segers and Sluijsmans (1999) found that the accuracy in self-assessment in a sample of students improved with practice over time.

If these last statements are true, this is, if self-assessment can be trained, then two new questions spring up. First, when should these practices be implemented? Taras (2001), citing also Boud (1986) and Wood, Marshall and Hrymak (1988), suggests that the optimum moment is during the first or second year of undergraduate studies, when students are more receptive and a greater accumulative value can be gained. Second, will self-assessment skills generalize to other contexts? Andrade and Du (2007) indicate that, while generalization is possible in some instances, it is not the most common outcome.

Should then self-assessment be just formative, or will also the summative self-assessment yield benefits? One of the main arguments in favor of self-assessment is the modification of the unilateral power that professors keep along the assessment process (Boud, 1995; Butcher & Stefani, 1995; McMahon, 1999; Rainsbury & Hodges, 1998; Somervell, 1993; Stefani, 1998). Authors as Maddalena Taras (2001) assert that such modification of power relationships cannot be achieved unless students are allowed to implement the summative self-assessment: if students don't have access to the process of summative assessment, their involvement in the power basis of higher education will be merely peripheral. Nevertheless, Tan (2004, 2008) argues that the participation of students in summative self-assessment only increases their power when the result of such self-assessment has priority over the professor's assessment. In this line, Race (1991) indicates that summative self-assessment can hardly be participative or redistribute power if the students know that the professor will not accept their self-assigned marks unless he or she considers them satisfactory.

Is it advisable, therefore, to implement the summative self-assessment and to make it count, total or partially, towards the final marks? Our results do not support this option. That was precisely the paradigm of group TYF, in which we found that, differently from group PSE (whose self-assessments did not count towards the marks), self-grading yielded results which differed significantly from the professor's marks. In this regard, Andrade and Du (2007) stated that counting self-assessment towards the final marks may divert the students'

attention from the quality of their work and how to improve it, compromising their honesty and separating them from their learning objective.

Not counting summative self-assessment towards the marks –therefore not redistributing power in terms of Tan (2004, 2008) and Race (1991)– does not mean that it is useless in the teaching-learning process. We consider that group PSE, whose self-assessment didn't count towards the marks, benefited from making explicit what they considered a fair mark and from the fact that this largely coincided with the mark latter assigned by the professor. This may have contributed to their acceptance, as revealed by a decrease in claims after the exam, but it is a subjective post hoc perception that we did not include in our study.

The main limitation of our findings probably comes from the fact that the professor's marks (and not necessarily the real learning of the student) were used as comparison criterion, which is somehow compromised due the existing variability among professors regarding the reliability of their assessments. Van Daalen (1999) suggested that reliability and validity can be improved in these instances by calculating the correlation between self-assessment and the judgment of several experts. This procedure (known as double-marking or multiple-marking), implies its own problems (Ecclestone & Swann, 1999), but presents the advantage of being more accurate and fair, also facilitating the coordination of criteria among professors (Tan, 2004). Practical reasons did not allow us to use multiple-marking in our research, which might limit the validity of the results.

As Sullivan and Hall (1997) suggested regarding their own results, we think that a relevant contribution of the present study is the confirmation that, although the students in general are able to adequately assess their acquisitions, in a noteworthy number of cases there is a high discrepancy with respect to the professor's assessment. The answers to the interviews also show that the students don't always understand the reasons for those differences, which may negatively affect their motivation, self-concept, attitude towards the teachers, etc. If any assessment is prone to be questioned in some degree (Boud, 1995), in higher education it must also be taken into account that the students see themselves as adults and that's the treatment they need to perceive. Assessment is an area in which they feel particularly vulnerable and, although they know that it's their work and not themselves what is being judged, the truth is that they feel emotionally involved with it and with the way it is assessed (Taras,

2001). Any difference between the professor's mark and the mark that the student perceives as fair must be clarified, and in this regard post-marking tutorships may be fundamental.

Finally, the vast majority of females in our study made it impossible to check the possible existence of sex differences in the students' summative criteria-referenced self-assessment. Regarding this issue there are confronted results in the literature, both in favor of the existence of such differences (Andrade & Boulay, 2003; Goodrich, 1996) as against it (Andrade & Du, 2007; Dweck *et al.*, 1978; Roberts & Nolen-Hoeksema, 1989), so it appears as a topic to be addressed in future research.

References

- Álvarez, I.M. (2009). Evaluar para contribuir a la autorregulación del aprendizaje. *Electronic Journal of Research in Educational Psychology*, 7(3), 1007-1030.
- Andrade, H. & Boulay, B. (2003). Gender and the role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research*, 97(1), 21–34.
- Andrade, H. & Du, Y. (2007). Student responses to criteria referenced self-assessment. *Assessment and Evaluation in Higher Education*, 32(2), 159-181.
- Barber, L.W. (1997). Autoevaluación. In J. Millman & L. Darling-Hammond, *Manual para la evaluación del profesorado*. Madrid: Editorial La Muralla.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Boud, D. (1989). The role of self assessment in student grading. *Assessment & Evaluation in Higher Education*, 14(1), 20–30.
- Boud, D. & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529–549.
- Boud, D.J. (1986). *Implementing Student Self-Assessment*. Sydney: Higher Education Research and Development Society of Australia.
- Brew, A. (1999). Towards autonomous assessment: using self-assessment and peer assessment. In S. Brown & A. Glasner (Eds), *Assessment matters in higher education (pp. 159-171)*. Buckingham: SRHE & Open University Press.
- Butcher, A.C. & Stefani, L.J. (1995). Analysis of peer, self- and staff-assessment in group Project work. *Assessment in Education*, 2(2), 165–186.
- Cassidy, S. (2007). Assessing ‘inexperienced’ students’ ability to self-assess: Exploring links with learning style and academic personal control. *Assessment and Evaluation in Higher Education*, 32(3), 313–330.
- Dearing, R. (1997). *Higher Education in the Learning Society*. London: HMSO.
- Dochy, F. & McDowell, L. (1997) Introduction: assessment as a tool for learning. *Studies in Educational Evaluation*, 23(4), 279–298.
- Dochy, F., Segers, M. & Sluijsmans, D.M.A. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Educational Evaluation*, 24(3), 331–50.
- Dweck, C., Davidson, W., Nelson, S. & Enna, B. (1978). Sex differences in learned helplessness: 2. Contingencies of evaluative feedback in the classroom and 3. An experimental analysis. *Developmental Psychology*, 14(3), 268–276.

- Ecclestone, K. & Swann, J. (1999). Litigation and learning tensions in improving university lecturers' assessment practice, *Assessment in Education: Principles, Policy and Practice*, 6(3), 377–389.
- Eissa, M.A. (2007). La eficacia de un programa basado en el desarrollo de estrategias de autorregulación de la escritura en estudiantes de educación secundaria con dificultades de escritura. *Electronic Journal of Research in Educational Psychology*, 7(1), 5-24.
- Eva, K.W., Cunnington, J.P.W., Reiter, H.I., Keane, D.R. & Norman, G.R. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in Health Sciences Education*, 9(3), 211–224.
- Falchikov, N. (1997, septiembre). Why do lecturers involve students in assessment? Paper presented at the 2nd *Northumbria Assessment Conference, Encouraging Partnership in Assessing Learning*. University of Northumbria at Newcastle.
- Falchikov, N. (2005). *Improving assessment through student involvement*. Abingdon, Oxon: RoutledgeFalmer.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical Solutions for aiding learning in higher and further education*. London: Routledge.
- Falchikov, N. & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis. *Review of Educational Research*, 59(4), 395–430.
- Fitzgerald, J.T., White, C.B. & Gruppen, L.D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education*, 37(7), 645–649.
- Garcia, J.A. & Floyd, C.E. (1999). Using single system design for student self-assessment: a method for enhancing practice and integrating curriculum. *Journal of Social Work Education*, 35(3), 451–461.
- Gibbs, G. (1995). *Assessing student centred courses*. Oxford: Oxford Centre for Staff Development.
- Goodrich, H. (1996). *Student self-assessment: at the intersection of metacognition and authentic assessment*. PhD Thesis, Harvard University.
- Gregory, K., Cameron, C. & Davies, A. (2000). *Self-assessment and goal-setting*. Merville: Connection Publishing.
- Gruppen, L., Garcia, J., Grum, C., Fitzgerald, J., White, C., Dicken, L., Sisson, J. & Zweifler, A. (1997). Medical students' self-assessment accuracy in communication skills, *Academic Medicine*, 72(10), S57–59.

- Hafner, J. & Hafner, P. (2003) Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509–1528.
- Hanrahan, S.J. & Isaacs, G. (2001). Assessing self- and peer-assessment: the students' views, *Higher Educational Research and Development*, 20(1), 53–70.
- Kraayenoord, C.E.V. & Paris, S.G. (1997). Australian students' self-appraisal of their work samples and academic progress. *Elementary School Journal*, 97(5), 523–537.
- Lew, M.D.N. & Schmidt, H.G. (2006). *Reflection upon learning between theory and practice: A focus-group study of tutors' and students' perceptions*. Rotterdam: Erasmus University Rotterdam.
- Lew, M.D.N., Alwis, W.A.M. & Schmidt, H.G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment and Evaluation in Higher Education*, 35(2), 135–156.
- Longhurst, N. & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation*, 23(4), 319–330.
- Lucas, S. (2007). Desarrollo de competencias desde la enseñanza universitaria. Armonización con la Enseñanza Secundaria y el mercado de trabajo, desde la Psicología Social de la Educación. *Electronic Journal of Research in Educational Psychology*, 5(1), 1696-2095.
- Maguire, S., Evans, S.E. & Dyas, L. (2001). Approaches to learning: A study of firstyear geography undergraduates. *Journal of Geography in Higher Education*, 25(1), 95–107.
- McMahon, T. (1999). Using negotiation in summative assessment to encourage critical thinking. *Teaching in Higher Education*, 4(4), 549–554.
- Orsmond, P., Merry, S. & Reiling, K. (1997). Students and tutors perceptions of a good essay. *Research in Education*, 58, 81–84.
- Paris, S.G. & Cunningham, A.E. (1996). Children becoming students. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 117–147). New York: Simon & Schuster Macmillan.
- Paris, S.G. & Paris, A.H. (2001). Classroom applications of research on self-regulated learning, *Educational Psychologist*, 36(2), 89–101.
- Popham, W.J. (1983). *Problemas y técnicas de evaluación educativa*. Madrid: Anaya.
- Race, P. (1991). *Learning through assessing*. Paper presented at the Standing Conference on Educational Development. Newcastle, UK.

- Rainsbury, E. & Hodges, D. (1998). Academic, employer and student collaborative assessment in a work-based cooperative education course, *Assessment and Evaluation in Higher Education*, 23(3), 313–325.
- Roberts, T. & Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback, *Sex Roles*, 21(11–12), 725–746.
- Rubin, H. & Rubin, I. (1995). *Qualitative interviewing: the art of hearing data*. Thousand Oaks, CA: Sage.
- Scriven, M.S. (1967). *The methodology of evaluation, en Perspectives of curriculum evaluation (AERA Monograph Series on Curriculum Evaluation, n° 1)*. Chicago: Rand McNally.
- Sluijsmans, D.M.A., Moerkerke, G. & Dochy, F.J.R.C. (1998). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, 1, 293–319.
- Somervell, H. (1993). Issues in assessment, enterprise and higher education: the case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education*, 18(3), 221–233.
- Stallings, V. & Tascione, C. (1996) Student self-assessment and self-evaluation, *Mathematics Teacher*, 89(7), 548–555.
- Stefani, L.J. (1998). Assessment in partnership with learners, *Assessment and Evaluation in Higher Education*, 23(4), 339–350.
- Sullivan, K. & Hall, C. (1997). Introducing students to self-assessment. *Assessment and Evaluation in Higher Education*, 22(3), 289–305.
- Tan, K.H.K. (2004). Does student self-assessment empower or discipline students? *Assessment and Evaluation in Higher Education*, 29(6), 651–662.
- Tan, K.H.K. (2008). Qualitatively different ways of experiencing student self-assessment. *Higher Education Research and Development*, 27(1), 15–29.
- Taras, M. (2001). The Use of Tutor Feedback and Student Self-assessment in Summative Assessment Tasks: towards transparency for students and for tutors. *Assessment and Evaluation in Higher Education*, 26(6), 605–614.
- Van Daalen, M. (1999). Test usefulness in alternative assessment. *Dialog on Language Instruction*, (1–2), 1–26.
- Wood, D.R., Marshall, R.R. & Hrymak, A.N. (1988). Self-assessment in the context of the McMaster Problem Solving Programme. *Assessment and Evaluation in Higher Education*, 13(2), 107–127.

Annex 1. Self-assessment questionnaire (group PSE).

**UNIVERSITY OF CÁDIZ; DEGREE IN CHILDHOOD EDUCATION
EDUCATIONAL PSYCHOLOGY**

ASSESSMENT REPORT

NAME _____ DATE _____

- 1. Indicate with an X the degree in which you think you have acquired the following competence:**

	Very low	Low	Medium	High	Very high
To understand the educational and learning processes in the 0-6 years period, in the family, social and school contexts					

- 2. Indicate with an X the degree in which you think you have reached each of the following learning results:**

	Very low	Low	Medium	High	Very high
To understand the family as a learning context as well as the need of collaboration with the school					
To understand the importance of perception processes in early attention					
To understand and establish relationships between learning, basic psychological processes and the educational process in Childhood Education					
To elaborate action proposals based on such knowledge as well as on the differences in abilities and learning rhythms					
To identify and to know the different applications of Psychology within the school					
To be able to educatively ap-					

ply in the class-room the re- sources and instruments pro- vided by the Psychology					
--	--	--	--	--	--

3. Indicate with an X the degree in which you think you have learnt the following contents:

	Very low	Low	Medium	High	Very high
Educational Psychology with- in the context of psychological sciences					
Psychological models of the learning-teaching process ap- plied to Childhood Education and Early Attention					
Basic psychological processes in Childhood Education					
Construction of knowledge in the class-room					
Interactions in the class-room and their social, affective and learning implications in Childhood Education					
The family as learning context					
The scientific method and its applications					
Strategies for an efficient learning					
Family values and beliefs in education					

4. Round off the mark that, in your opinion, you should receive in this subject:

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

[This page intentionally left blank]