

Chapter 2

Is University Students' Self-Assessment Accurate?

Alicia Bolívar-Cruz, Domingo Verano-Tacoronte,
and Sara M. González-Betancor

Abstract The paper's main objective is to evaluate the self-assessment accuracy of university students. Specifically, the study analyzes the self-assessment of oral communication skills. It was carried out in a Firm Labor Organization course included in the Labor Relations and Human Resources Degree. The literature on self-assessment in Higher Education does not provide clear evidence about its accuracy, as a number of methodological problems have been detected. To reduce them, we have taken a number of precautions. Thus, a rubric was designed, and students were trained to use it. Several teachers and peers were introduced as referents, and a segmented analysis was conducted based on gender and the students' level of competence from the teachers' point of view. Results show that self-assessment accuracy is low. Moreover, regardless of the degree, men's self-assessments are higher than women's. Moreover, the findings suggest that the scoring rubric improves self-assessment accuracy when the speaker has good oral communication skills, but not when these skills are poor. These results lead us to propose the development of correction factors that can be adapted to any situation, thus allowing self-assessment to be used for summative purposes.

2.1 Introduction

In any profession that follows the principles of autonomy and self-regulation, people have to be aware of the need to assess their own work and engage in continuous learning throughout their careers (Boud, 1989; Regehr, Hodges, Tiberius, & Lofchy, 1996).

A. Bolívar-Cruz (✉) • D. Verano-Tacoronte (✉)
Economía y Dirección de Empresas, University of Las Palmas de Gran Canaria,
4th Saulo Torón, Las Palmas de Gran Canaria, Islas Canarias, 35017 Spain
e-mail: alicia.bolivar@ulpgc.es; domingo.verano@ulpgc.es

S.M. González-Betancor
Métodos Cuantitativos en Economía y Gestión, University of Las Palmas de Gran Canaria,
4th Saulo Torón, Las Palmas de Gran Canaria, Islas Canarias 35017, Spain
e-mail: sara.gonzalez@ulpgc.es

Thus, Boud (1989) suggests that one of the responsibilities of educators is to teach students to operate as professionals would. In other words, they should be capable of giving and receiving feedback and assessing their own work and that of others, which in turn would increase their professional competence. This argument supports the growing interest shown by university teachers in stimulating student participation in the learning process.

However, the literature does not provide conclusive results about people's ability to assess themselves (e.g., Campbell, Mothersbaugh, Brammer, & Taylor, 2001; De Grez, Valcke, & Roozen, 2012; Langan et al., 2008; Patri, 2002; Ward, Gruppen, & Regehr, 2002). In addition, as Stefani (1994) and Ward et al. (2002) point out, studies on self-assessment have been far from rigorous, basically due to the different criteria used by the assessors (students and teachers).

An accurate self-assessment would contribute to developing the student's critical view of his/her own work. Therefore, the present study aims to analyze the self-assessment accuracy of university students, compared to other assessors of their performance, such as peers or teachers. The study evaluates the oral communication competence, which is one of the most difficult to measure (Bolívar-Cruz et al., 2013). The methodology followed is designed to resolve several of the problems shown in the literature on self-assessment accuracy, making it possible to use this type of assessment for both formative and summative purposes.

After this introduction, the following section presents the theoretical basis for student self-assessment and its accuracy, and it addresses questions related to self-assessment accuracy. After that, we present the methodological design that guided the investigation. Next, the results of the empirical study are analyzed. Finally, we present a summary of the main results and the conclusions reached, as well as future lines of research.

2.2 The Student's Self-Assessment and Its Accuracy

This paper addresses the study of university students' self-assessment accuracy. After defining self-assessment, its main advantages will be presented, and the reasons for its low accuracy level will be analyzed.

According to Boud and Falchikov (1989), self-assessment refers to the student's commitment to judging his/her own learning, especially the achievements and results obtained.

Incorporating the student into his/her own assessment offers several advantages (Campbell et al., 2001; Dochy, Segers, & Sluijsmans, 1999; Falchikov, 2005; Gessa-Perera, 2011; Marín-García, 2009; Regehr et al., 1996; Topping, 2003), with the following being especially noteworthy: (a) it contributes to developing valuable skills for the job market, such as having a critical view of their own work; (b) it increases students' involvement in their learning; and (c) it frees the teacher to spend time on tasks with greater educational value. Specifically, in developing the ability to make oral presentations, De Grez et al. (2012) indicate that self-assessment produces improvements in grades, in perceived learning, in confidence about making

better presentations, and in the development of assessment skills. These benefits justify the use of self-assessment (Boud, 1989; Boud & Falchikov, 1989; Taras, 2010), even when it is not as accurate as it could be.

In spite of these advantages, the incorporation of self-assessment into educational practice is limited (Boud & Falchikov, 1989), especially for summative grading purposes (Stefani, 1994). One of the reasons for this low implementation is the lack of accuracy shown by students when taking on the role of assessors of their own work.

The literature analyzing self-assessment is scant. Moreover, the results are not solid and even show contradictions. Thus, studies like those by Dochy et al. (1999) and Al-Fallay (2004) present favorable empirical results about the use of self-assessment, while others, such as those by Campbell et al. (2001), De Grez et al. (2012), Langan et al. (2008), Patri (2002), Regehr et al. (1996) and Ward et al. (2002), find results that are contrary to its use.

Methodological and psychological problems are also cited to justify the lack of consensus about self-assessment and its accuracy in Higher Education. The first problem is that the activities assessed have quite different characteristics (e.g., essays, group work, oral presentations, practical laboratory sessions), and so it is not surprising that the results do not coincide (Marín-García, 2009). Specifically in the case of oral presentations, few studies were found, which allows us to assume that it is important to perform an in-depth examination of this activity (De Grez et al., 2012; Lew, Alwis, & Schmidt, 2010; Marín-García, 2009).

Regarding the methodological problems, Ward et al. (2002) indicate that the accuracy of self-assessment has been verified through correlations analysis between students' scores and scores from an external source (teachers or peers). This approach presents various problems: (a) teachers' assessments are usually the standard for comparing students' assessments, which is, at the least, questionable (Falchikov & Boud, 1989; Topping, 2009), especially in the area of communication skills, where it is more difficult to find valid comparison patterns among expert raters (De Grez et al., 2012).

Likewise, the assumption is made that people who grade themselves act as a coherent group, which is also questionable because it would mean that all students use the same criteria and the same assessment scale (Hanrahan & Isaacs, 2001; Ward et al., 2002).

Another methodological problem related to the correlations approach is the consideration that all students act in the same way when they have the opportunity to assess their own performance. A few outliers can make the correlation index much lower than what would be expected; therefore, group heterogeneity would have to be taken into account (Ward et al., 2002).

Another aspect to consider is the influence of differences between assessors. Among these differences, an aspect that has received considerable attention is the assessor's gender and its influence on assessment quality (Archer, 1992; Falchikov & Magin, 1997). Although there are gender-based differences in self-assessment (Beyer, 1990), which seem to be related to women's lower perception of self-efficacy and confidence in their own performance (Pallier, 2003), the studies carried out in the educational setting do not provide definitive results on this topic (Boud & Falchikov, 1989). If this question is analyzed in relation to oral presentation skills, the results are not conclusive either. For example, Langan et al. (2005, 2008) detect

a significant effect of the assessor's gender, while Sellnow and Treinen (2004) do not. Meanwhile, De Grez et al. (2012) do not find a significant relationship between gender and self- and teacher assessment comparisons, although they did find one between peer and self-assessment.

Finally, the repercussions of self-assessment for the student can reduce its accuracy. Thus, Tejeiro et al. (2012) indicate that when the self-assessment can affect the grade, students' and teachers' scores do not correlate. This basically occurs for two reasons: (a) students' desire to raise their grades (Lew et al., 2010); and (b) the added pressure of assessing themselves, as Taras (2010) pointed out when indicating that poor students worry more than good ones when assessing themselves.

Based on these problems, it seems important to take a series of steps to improve self-assessment accuracy. Thus, more reliable and valid standards should be used to compare self-assessments, such as the introduction of peer assessment or forming committees of teachers as a control mechanism (Ward et al., 2002). In this sense, various studies (Campbell et al., 2001; De Grez et al., 2012; Langan et al., 2008) conclude that peer assessment is more precise than self-assessment. It is curious to observe that these studies support students' capacity for correct assessment when they have to judge the performance of others, but they cast doubt on students' ability or intention to apply this same level of rigor to their own performance.

Another way to improve accuracy is to employ assessment formats that are easy to use, reliable and with high content validity. Thus, the use of rubrics is a form of assessment that makes it possible to rate the quality of students' contributions and performance levels in different areas, specifying, before doing the activity, the factors or variables that will be analyzed and the requirements for each (Andrade & Du, 2005; García-Ros, 2011; Jonsson & Svingby, 2007). Rubrics, therefore, can reduce assessment subjectivity and produce greater agreement among the scores. Likewise, it is necessary to provide adequate student training in the use of these rubrics and facilitate opportunities for self-assessment (Marín-García, 2009) throughout the degree programs.

The final precaution is related to the existence of differences in the raters' assessment behavior. This problem can be addressed by segmenting the group of students according to their performance on the activity, for example, following the teacher's criteria, in order to observe the phenomena of self-indulgence in the worst students and self-demanding behaviors in the best.

2.3 Methodological Design

As mentioned above, the purpose of this study is to rate the self-assessment accuracy of university students. Thus, it aims to verify whether the self-assessment of oral communication skills, in a summative assessment context, is sufficiently accurate compared to other sources of assessment, once a series of methodological

Table 2.1 Reliability of the rubric

	Cronbach's alpha
Teachers	0.8296
Peers	0.8704
Self-assessment	0.7965

precautions have been incorporated into the process. Thus, we propose three specific objectives:

- Find out whether it is possible to obtain a high level of self-assessment accuracy through the use of rubrics.
- Verify whether self-assessment accuracy is related to the speaker's gender.
- Analyze whether there are differentiated patterns of behavior when students are segmented according to their teachers' ratings of their presentations.

The study was carried out in Firm Labor Organization, an obligatory course taught in the Degree of Labor Relations and Human Resources (hereinafter, LRHR), which is worth six credits. The participants in the study were 92 students who assessed their classmates and themselves while performing a test consisting of making an oral presentation in teams of two people. In addition, each of the presentations was assessed by two teachers, the one responsible for the subject and another unrelated to it, both with considerable experience in assessing oral presentations.

In order to unify the assessment criteria, the students were given a rubric elaborated by teachers with experience in rating oral presentations in the university context. This rubric consisted of ten assessment criteria that included the main dimensions of the skills analyzed. In turn, each criterion was rated on a three-level scale (1—deficient, 2—acceptable, 3—excellent), and a detailed description was provided of the necessary requisites for each level. Before the presentations were made, all of the assessors had access to the rubric, they were carefully told about its functioning, and any doubts were clarified. In order to increase the students' degree of involvement, the grade received on the oral presentation was linked to the final grade in the course (i.e., summative assessment).

After collecting the assessments of the presentations, the rubric's reliability was rated through inter-rater agreement (García-Ros, 2011), considering three groups of raters: teachers, self-assessment and peers. This consistency was measured by applying Cronbach's alpha to the rubric, obtaining the results presented in Table 2.1. As the table shows, there is good internal consistency in the three groups; therefore, the rubric is considered reliable (Cortina, 1993).

In order to fulfill the proposed objectives, the global score for each speaker was obtained from the sum of the scores given by the assessors on each criterion on the rubric. Thus, variables are generated for the global score given by the teachers, the global score given by the peers and the self-assessment score. When there is more than one assessor (teachers and peers), the average of the score awarded by each rater is used, that is, the teachers' mean and the peers' mean. To facilitate comparison, the averages calculated were rounded to the first integer, given that the self-assessment can only produce integers.

The first specific objective of this study was to determine whether the assessments can be considered accurate. Therefore, the level of agreement among the scores of the three assessors, teachers, peers and self-assessment, was analyzed graphically. Moreover, we presented the main descriptive statistics, along with their statistical significance, measured through tests for equality of means.

The second objective was to analyze the self-assessment accuracy with regard to gender. To do so, the sample was segmented based on the gender of the speaker in the cases where the difference between the self-assessment and the teacher's assessment was statistically significant. Histograms of frequency were used, as well as the basic descriptive statistics, to later try to identify a linear relationship between the grades through the simple linear correlation coefficient. If this relationship was not detected, the possible independence among the variables was analyzed by applying the Spearman coefficient.

Finally, the identification of behavioral patterns by groups of students (third specific objective) was carried out through graphic analysis, segmenting the sample based on the scores the students received from their teachers.

2.4 Analysis and Discussion of the Results

Figure 2.1 shows an initial examination of the level of agreement among the three collectives involved in assessing the oral presentations. It can be observed that the teacher assessment coincides more with the peer assessment than with the self-assessment. In fact, as would be expected, the self-assessment is higher than the other two in most cases. Another noteworthy result is that the teachers' scores are quite similar the peers' scores, although the latter present less variation and, therefore, discriminate less, as shown in the studies by Kwan and Leung (1996), Magin and Helmore (2001) and Marín-García (2009). Thus, we can establish that the use of the rubric seems to bring the peer and teacher assessments closer to each other than to the self-assessment.

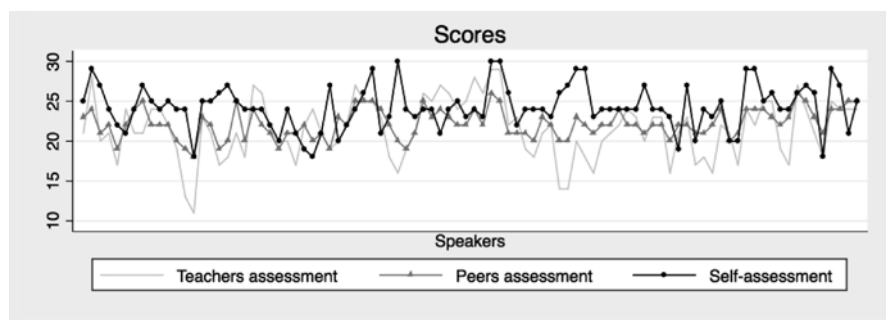


Fig. 2.1 Scores given to the speakers by type of assessor

Table 2.2 Descriptive statistics of the global score on the presentations

Number of presentations		92
Min.–max.	Teachers	11–29
	Peers	18–26
	Self-assessment	18–30
Mean	Teachers	21.49
	Peers	22.18
	Self-assessment	24.29
Standard deviation	Teachers	3.78
	Peers	1.83
	Self-assessment	3.10

To improve the accuracy of this analysis, we carried out a descriptive statistical analysis. The results can be seen in Table 2.2. The table shows that the range between the minimum and maximum scores of the teachers is superior to the range of the peers, while the range corresponding to self-assessment is always located above the minimum and maximum of the other assessors. Therefore, while peers seem to give more intermediate scores, and teachers use a broader range of scores, the students give themselves higher grades. Moreover, and examining the level of inter-rater agreement, it can be seen that there are no statistically significant differences between the assessment means of teachers and peers, while the differences between either of these two and self-assessment are statistically significant. Therefore, once again, teachers and peers score similarly, while self-assessment offers divergent values.

In order to discover the influence of some personal variables on students' assessments, in our case gender, it is necessary to segment the sample based on whether the speaker is a man or a woman (second specific objective). Figure 2.2 shows the histogram of frequencies of the global score awarded by each of the assessors, based on the gender of the speaker.

A general tendency can be observed in which the distribution is displaced toward the right as we change the assessor (teacher, peers, self-assessment), for both sexes, although this tendency seems to become stronger when self-assessment is performed by men. In general, peers give higher scores to the communicative competence of the speakers than the teachers do, regardless of the gender of the speaker. In addition, the self-assessment of this skill is higher than the peers' perception, and this difference seems to be greater in men than in women.

Although the results seem clear, it is necessary to find out whether these gender-based differences can be considered valid. Therefore, contrasts of differences in means are conducted, and the tendency of the scores by assessor group is analyzed (see Table 2.3). First, examining only the coefficients of the differences, it can be observed that the direction of the teachers' and peers' scores usually coincides. Thus, both groups give higher scores to women than to men. However, when the significance is analyzed, the peer ratings do not show significant differences based on gender. Even so, differentiating the speaker by gender is relevant, as it shows that men's self-assessment is systematically higher than women's, with the differences

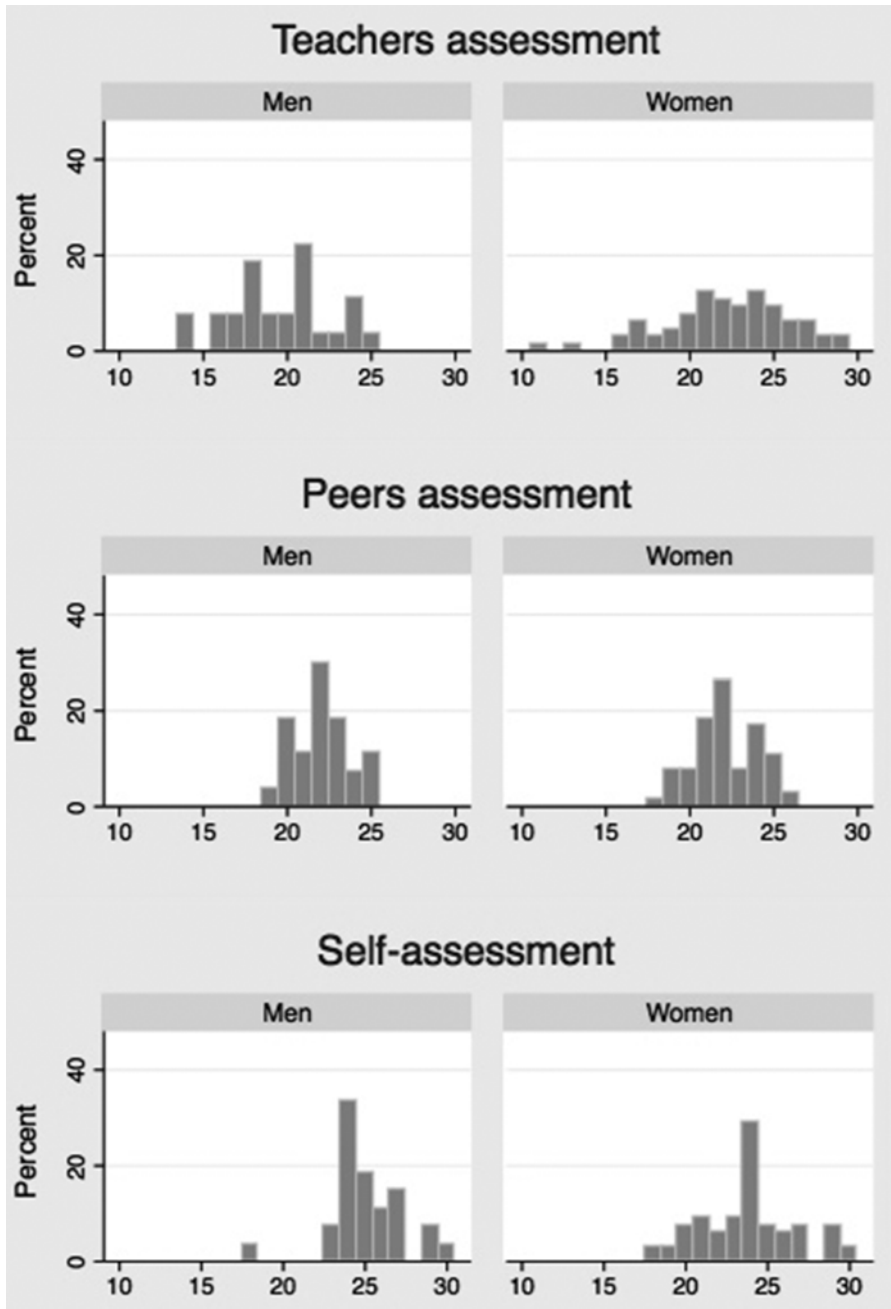


Fig. 2.2 Histogram of frequencies for the global score based on the gender of the speaker and the type of assessor

Table 2.3 Descriptive statistics of the global scores on the presentations and the speaker's gender

		Men	Women
Number of presentations		27	65
Min.--max.	Teachers	14–25	11–29
	Peers	19–25	18–26
	Self-assessment	18–30	18–30
Mean	Teachers	19.46***	22.25***
	Peers	22.07	22.23
	Self-assessment	25.55***	23.84***
Standard Deviation	Teachers	3.13	3.74
	Peers	1.66	1.91
	Self-assessment	2.61	3.16

***Differences in mean by speaker's gender statistically significant at 0.01 level

Table 2.4 Linear correlation between assessment sources by gender

	Men			Women		
	Teachers	Peers	Self-	Teachers	Peers	Self-
Teachers	1			1		
Peers	0.7184***	1		0.7781***	1	
Self-assessment	0.0305	0.0949	1	0.4306***	0.4710***	1

*** Linear correlation statistically significant at 0.01 level

being significant. Thus, self-assessment shows opposite results to the opinions of teachers and peers.

Based on the data in Table 2.3, a certain correlation between the peer and teacher assessments can be intuited, but not between these scores and the self-assessment. To quantify each relationship, the correlation is analyzed, and the results appear in Table 2.4.

The table shows a high linear correlation between peer and teacher ratings (71 % for men and 78 % for women). Moreover, the linear correlation between self-assessment and the other assessors is significant in the case of women, although the coefficient is low in comparison with those already mentioned (43 % correlation with teachers and 47 % with peers).

As a high correlation was not detected between self-assessment and peer and teacher ratings, even though they all saw the same presentation and used the same rubric, we considered the possibility that the relationship might not be linear. Therefore, Spearman ranges were calculated, but without finding any change in the results. These results led us to conclude that the use of the rubric seems to have brought the scores of teachers and peers closer to each other, but there was less convergence between their opinions and the self-assessments, especially when the oral presentations were made by male speakers.

As the third specific objective, we proposed that self-assessment would behave in a differentiated way in students assessed by teachers as having better or worse skills.

The analysis was performed by dividing the speakers based on gender, given that the teacher ratings had been statistically significant. The grouping of the students in one collective or the other (with better/worse level of competence) was determined through the construction of confidence intervals for the set of individuals and the gender of the student. Thus, the students who were outside the interval, constructed as a mean score plus/minus a standard deviation for their reference group, would be the best/worst. Figure 2.3 presents the score given by each of the three rating sources to each of the speakers in the collectives identified in this way.

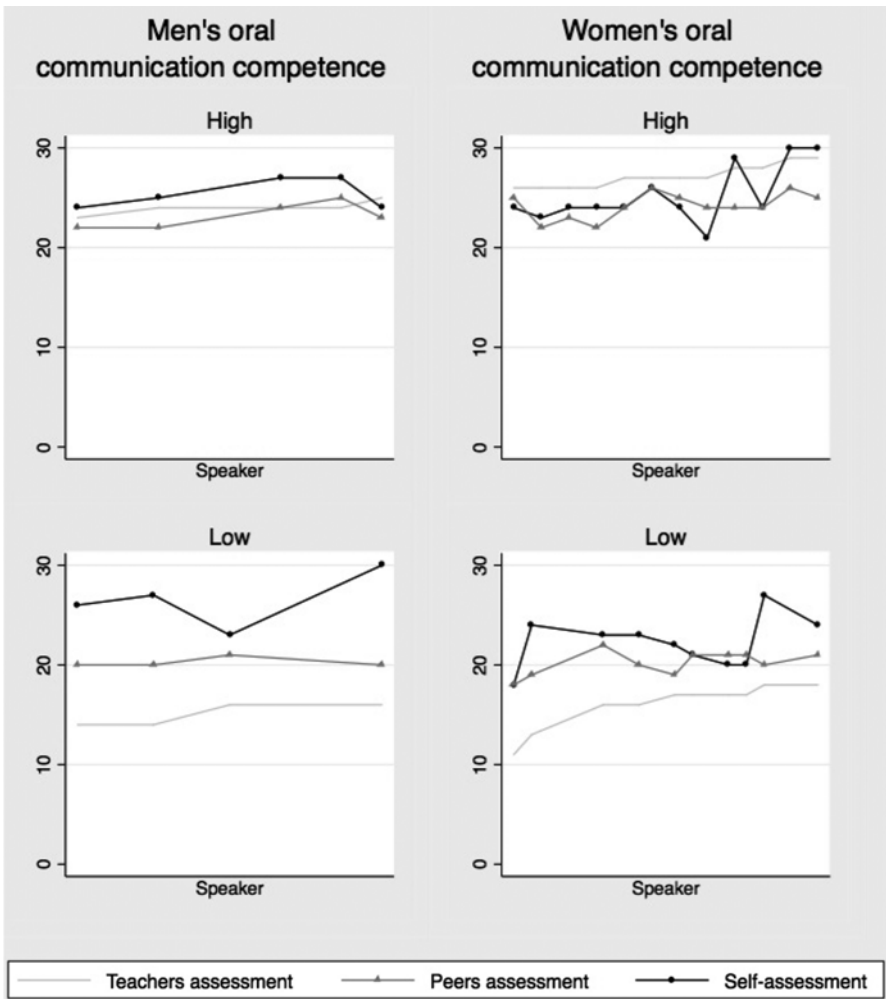


Fig. 2.3 Score given by the assessors to the speakers with high/low oral communication competence and gender

On the presentations by the students with low communication competence, without distinguishing the gender, there is greater disparity in the assessments of the three sources than on the presentations by the students with high competence, as the latter show a consensus among the scores. Therefore, the rubric seems to unify criteria when the speaker shows a high level of oral communication skills, but not when the speaker lacks or has low levels of these skills.

In any case, teacher and peer behaviors seem to follow the same pattern, as the results show that teachers are stricter than peers when assessing presentations by students with low communication skills, while on presentations by students with high communication competence, teachers are more benevolent than peers. Likewise, there is a clear pattern in the self-assessment of the presentations made by students with low competence, as their self-assessment is systematically higher than the ratings by the other two assessment sources. This difference is even more pronounced in men than in women. However, men with high communication competence continue to self-assess their presentations with higher scores than the other assessors, while the women tend to underrate themselves.

2.5 Conclusions and Future Research

The development of the self-assessment capacity is attracting a lot of attention in the academic world, given the importance of the student's involvement in the learning process, not only to improve his/her academic results, but also because this skill contributes to the student's professional development. However, the research carried out on self-assessment accuracy does not provide conclusive results, and it presents a lack of methodological rigor (Stefani, 1994; Ward et al., 2002). The present study has analyzed self-assessment accuracy in the university, after taking a series of methodological precautions recommended in the literature. Thus, university students' ability to rate their oral communication competence was measured, using as peer and teacher assessments as referents. This study proposed three specific objectives: (1) to find out whether the use of a rubric makes it possible to obtain a high level of agreement among the different types of assessors; (2) to verify whether self-assessment accuracy is related to the speaker's gender; and (3) to examine the existence of different types of behavior in students' self-assessment, segmenting them according to the best or worst grades given by the teachers.

A series of conclusions can be drawn from the analysis of the results. On the one hand, although the use of a rubric allows teachers and peers to assess in a similar way, the same effect does not occur when self-assessment is incorporated. This result has been found in the previous literature (e.g. Kwan & Leung, 1996). Therefore, the conclusion can be drawn that the use of the rubric provides a high level of accuracy in the case of peers and teachers, but not in the case of self-assessment. Various arguments can explain this result. First, the effect of self-assessment on the grade can influence its outcome, producing higher scores than other assessment sources and reducing the rubric's effect. Moreover, the lack of a

self-assessment habit, not involving the students in identifying the criteria, and the absence of teacher-student negotiation about the criteria to be assessed could explain the results. Finally, the differences between teacher and self-assessments may be due to the teachers' greater experience in judging oral presentations (De Grez et al., 2012). However, it should be emphasized that the students assessed their classmates with sufficient accuracy when they acted as peers. Therefore, we can say that students can be good assessors of others, but, at least according to our data, they are not good at assessing themselves.

Regarding differences among students, the results show that self-assessment accuracy is related to the assessor's sex. Although the teacher and peer ratings were oriented in the same direction (both collectives think women present better communication skills), the self-assessment behavior is not as homogeneous. In general, men give themselves higher scores than women do. Furthermore, there is no significant relationship between men's self-assessment and teacher and peer assessment, while there is in the case of women, although the levels reached are much lower than those found between teachers and peers. This interesting result requires a study that focuses more on determining the causes for this behavior by the male speakers, who systematically rate themselves higher than the other two collectives do.

On the other hand, and given our baseline idea that not all students are going to behave in the same way when assessing themselves, we were able to show the existence of various types of behavior when dividing the sample according to the teachers' scores. The results indicate that the rubric unifies the ratings when speakers with high oral communication skills are assessed, but not in the case of low ones. In students with low communication skills, the self-assessment is systematically higher than the ratings by peers and teachers. This difference is even more pronounced in men than in women. When analyzing students with a high communication level, the results are noteworthy: the men give themselves higher scores than those awarded by peers and teachers, while, with the same references, the women tend to underrate themselves.

These results lead us to consider the possibility of proposing a correction factor. This correction factor seems necessary for using self-assessment within the summative assessment process. Thus, the differences between the self-assessments of the best and worst rated students could be reduced, as well as those stemming from the speaker's gender and not justified by the quality of the work.

In spite of the findings, incorporating self-assessment and, above all, peer assessment, provides positive opportunities (Boud, 2007; De Grez et al., 2012; Langan et al., 2005). As Dochy et al. (1999) argue, different forms of assessment have to be integrated into the study plans, linking them to learning quality through consequential validity (Boud, 2007), given that assessment can affect learning and other educational aspects. Self-assessment can be quite effective in preparing students to integrate various aspects of their learning, demonstrate their achievements, and explore implications for their later training. Therefore, the usefulness of self-assessments would lie in their dimension of formative assessment, that is, as a way to improve skills and capabilities (Birenbaum & Dochy, 1996), in addition to their capacity to energize the class and make the learning process more dynamic.

In order to guide self-assessment experiences in the framework of university teaching, several lines of action are proposed. First, to improve self-assessment accuracy, it would be necessary to: (1) increase students' training in self-assessment; (2) increase the number of self-assessment experiences, as they facilitate improvements in students' capacity to evaluate themselves (Birenbaum & Dochy, 1996; Boud & Falchikov, 1989); (3) involve the students in designing assessment scales (Falchikov, 2005), given that this process increases their commitment to the system; and (4) warn students about the possibility of applying correction factors, to the extent that their self-assessment differs greatly from the assessment of the referents.

Second, and to continue this line of study, it would be desirable to increase the number of degrees studied in order to draw conclusions common to all of them, given possible differences in demographic composition and learning styles (Cela-Ranilla & Gisbert Cervera, 2013). One relevant question is whether using a rubric can minimize the differences produced by the context.

Third, it would be necessary to analyze whether the results are maintained when self-assessment is not used for summative purposes. By eliminating the pressure to obtain a good grade, we might assume that students who have shown that they can assess others would also be able to accurately assess their own work.

Finally, although we examined the influence of gender on self-assessment accuracy, we did not explore other personal differences among students that could explain the divergence between self-assessments and teacher and peer assessments, and they should be addressed in future studies.

References

- Al-Fallay, I. (2004). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32, 407–428.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3). Retrieved from <http://PAREonline.net/getvn.asp?v=10&n=3>
- Archer, J. (1992). Sex bias in evaluations at college and work. *The Psychologist: Bulletin of the British Psychological Society*, 5(5), 200–204.
- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5), 960–970.
- Birenbaum, M., & Dochy, F. (Eds.). (1996). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston, MA: Kluwer.
- Bolívar-Cruz, A., Dávila-Cárdenes, N., Fernández-Monroy, M., Galván-Sánchez, I., González-Betancor, S. M., López-Puig, A. J., et al. (2013). Evaluación y fomento de la competencia de comunicación oral de los estudiantes universitarios. *Revista Iberoamericana de Evaluación Educativa*, 6(1), 17–28.
- Boud, D. (1989). The role of self-assessment in student grading. *Assessment and Evaluation in Higher Education*, 14, 20–30.
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term*. London: Routledge.

- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529–549.
- Campbell, K. S., Mothersbaugh, D. L., Brammer, C., & Taylor, T. (2001). Peer- versus self-assessment of oral business presentation performance. *Business Communication Quarterly, 64*(3), 23–42.
- Cela-Ranilla, J. M., & Gisbert Cervera, M. (2013). Learning patterns of first year students. *Revista de Educación, 361*, 171–195.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.
- De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(2), 129–142.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in Higher Education: a review. *Studies in Higher Education, 24*(3), 331–350.
- Falchikov, N. (2005). *Improving assessment through student involvement: practical solutions for aiding learning in Higher and Further Education*. New York, NY: Routledge Falmer.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research, 59*, 395–430.
- Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment and Evaluation in Higher Education, 22*(4), 385–396.
- García-Ros, R. (2011). Análisis y validación de una rúbrica para evaluar habilidades de presentación oral en contextos universitarios. *Electronic Journal of Research in Educational Psychology, 9*(25), 1043–1062.
- Gessa-Perera, A. (2011). La coevaluación como metodología complementaria de la evaluación del aprendizaje. Análisis y reflexión en las aulas universitarias. *Revista de Educación, 354*, 749–764.
- Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer-assessment: the students' views. *Higher Education Research and Development, 20*(1), 53–70.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130–144.
- Kwan, K. P., & Leung, R. W. (1996). Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education, 21*(3), 205–214.
- Langan, A. M., Shuker, D., Cullen, R., Penney, D., Preziosi, R., & Wheeler, P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment and Evaluation in Higher Education, 33*(2), 179–90.
- Langan, A. M., Wheeler, C., Shaw, E., Haines, B., Cullen, W., Boyle, J., et al. (2005). Peer assessment of oral presentations: Effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment and Evaluation in Higher Education, 30*(1), 21–34.
- Lew, M. D., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education, 35*(2), 135–156.
- Magin, D., & Helmore, P. (2001). Peer and teacher assessments of oral presentation skills: how reliable are they? *Studies in Higher Education, 26*(3), 287–298.
- Marín-García, J. A. (2009). Los alumnos y los profesores como evaluadores. Aplicación a la calificación de presentaciones orales. *Revista Española de Pedagogía, 242*, 79–98.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles, 48*, 265–276.
- Patri, M. (2002). The influence of peer feedback on self- and peer assessment of oral skills. *Language Testing, 19*(2), 109–131.
- Regehr, G., Hodges, B., Tiberius, R., & Lofchy, J. (1996). Measuring self-assessment skills: An innovative relative ranking model. *Academic Medicine, 71*(10 supplement), 52–54.

- Sellnow, D., & Treinen, K. (2004). The role of gender in perceived speaker competence: an analysis of student critiques. *Communication Education, 53*(3), 286–296.
- Stefani, L. A. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69–75.
- Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education, 15*(2), 199–209.
- Tejeiro, R. A., Gómez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., & Emberley, E. (2012). Summative self-assessment in higher education: Implications of its counting towards the final mark. *Electronic Journal of Research in Educational Psychology, 10*(2), 789–812.
- Topping, K. J. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. S. R. Segers, F. J. R. Dochy, & E. C. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, The Netherlands: Kluwer Academic.
- Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20–27.
- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: current state of the art. *Advances in Health Sciences Education, 7*, 63–80.