






Triadic assessment of oral presentations in higher education: Convergence between self-, peer- and teacher assessment according to academic performance

-  **Andrés Sánchez-García.** *Department of Physical Education and Sport. University of Sevilla. Sevilla, Spain. Research Group on Management and Innovation in Sports, Leisure and Recreation Services – GISDOR (SEJ-525). University of Sevilla. Sevilla, Spain.*
-  **Moisés Grimaldi-Puyana** . *Department of Physical Education and Sport. University of Sevilla. Sevilla, Spain. Research Group on Management and Innovation in Sports, Leisure and Recreation Services – GISDOR (SEJ-525). University of Sevilla. Sevilla, Spain.*
-  **Antonio Muñoz-Llerena.** *Department of Human Movement and Sports Performance. University of Sevilla. Sevilla, Spain. Research Group on Social Inclusion and Innovation through Physical Activity and Sport – INEFYD (HUM-1061). University of Sevilla. Sevilla, Spain.*
-  **Antonio Jesús Sánchez-Oliver.** *Department of Human Movement and Sports Performance. University of Sevilla. Sevilla, Spain. Studies Research Group in Neuromuscular Responses (GEPREN). University of Lavras. Lavras, Brazil.*

ABSTRACT

Formative assessment in higher education requires strategies that support the development of students' evaluative judgement. Triadic assessment, which integrates self-, peer- and teacher assessment, is considered a valuable practice to promote reflection on learning quality. Aiming to analyse differences between assessment modalities in university oral presentations and to examine whether students' academic performance is associated with greater convergence between student and teacher assessments, a cross-sectional observational study was conducted. 6,191 oral presentations assessments were carried out using an 8-item rubric (1–4 scale), and overall mean scores were calculated. Non-parametric procedures were used to analyse reliability, differences between assessment modalities and convergence with teacher criterion. The rubric showed adequate internal consistency ($\alpha = .85$). Significant differences were found between modalities ($p < .001$), with higher scores in self-assessment ($M = 3.75$), followed by peer assessment ($M = 3.58$) and teacher assessment ($M = 3.45$). Overall score was positively associated with the final course grade ($p = .16$; $p < .001$), with a stronger association in self-assessment ($p = .32$) than in peer assessment ($p = .12$). Students with higher performance tended to produce assessments closer to the teacher criterion, particularly in peer assessment ($p < .001$), although the pattern was not strictly linear. The largest discrepancies were observed in content, presentation, and time management. Findings support the value of triadic assessment as a strategy for developing students' evaluative judgement and provide empirical evidence of the role of academic performance in the progressive alignment with the teacher criterion in complex tasks.

Keywords: Physical education, Formative assessment, Rubric, Higher education, Communication skills.

Cite this article as:

Sánchez-García, A., Grimaldi-Puyana, M., Muñoz-Llerena, A., & Sánchez-Oliver, A. J. (2026). Triadic assessment of oral presentations in higher education: Convergence between self-, peer- and teacher assessment according to academic performance. *Journal of Human Sport and Exercise*, 21(4), 1238-1249. <https://doi.org/10.55860/tsmqhw98>

 **Corresponding author.** *Department of Physical Education and Sport. University of Sevilla. Sevilla, Spain.*

E-mail: mgrimaldi@us.es

Submitted for publication April 23, 2026.

Accepted for publication June 08, 2026.

Published June 23, 2026.

[Journal of Human Sport and Exercise](https://doi.org/10.55860/tsmqhw98). ISSN 1988-5202.

©Asociación Española de Análisis del Rendimiento Deportivo. Alicante. Spain.

doi: <https://doi.org/10.55860/tsmqhw98>

INTRODUCTION

Assessment in higher education has undergone a substantial shift in recent decades. It has gradually moved from approaches focused almost exclusively on grading towards models that emphasise its formative function (Pascual-Arias et al., 2023). This change responds to the need to promote deep, self-regulated and transferable learning. It also reflects the need to prepare students for academic and professional contexts in which the ability to judge the quality of one's own work and that of others is essential (Olague et al., 2025). In this framework, the concept of evaluative judgement has gained increasing attention.

Evaluative judgement is defined as the capacity to judge the quality of work based on explicit criteria and shared standards. It is considered a key competence for lifelong learning (Boud et al., 2018; Tai et al., 2018). From this perspective, assessment is no longer seen as a final event. Instead, it is understood as a continuous learning process (Sánchez-Oliver et al., 2025).

Self-assessment and peer assessment play a central role in the development of evaluative judgement. Many studies have shown that these practices contribute to improvements in self-regulation, metacognition and student engagement (Basurto-Mendoza et al., 2021; Bozkurt, 2020; Kambourova et al., 2021). However, it has also been repeatedly observed that student assessments differ from those made by teachers. In general, they tend to be less demanding and more variable in the application of criteria (Sánchez-García et al., 2026; Sánchez-Oliver et al., 2026).

These discrepancies can be interpreted as a problem of reliability or validity. From a formative perspective, however, they can also be understood as a natural expression of the learning process involved in assessment. This process depends on factors such as previous experience, understanding of criteria, task complexity and the level of academic competence achieved (Sánchez-García et al., 2026; Sánchez-Oliver et al., 2025; Sánchez-Oliver et al., 2026).

In this context, triadic assessment, which integrates self-assessment, peer assessment and teacher assessment, emerges as a particularly valuable strategy (Gale et al., 2002). This approach makes it possible to compare different evaluative perspectives, identify patterns of convergence and divergence, and provide students with multiple reference points to calibrate their understanding of quality (Mendoza et al., 2021). It also offers an appropriate framework to analyse how evaluative judgement develops and which variables are associated with closer alignment with the expert judgement of teachers (Boud et al., 2018; Tai et al., 2018).

Among these variables, students' academic performance appears to be a potentially relevant factor. It is reasonable to assume that students with stronger academic and conceptual mastery may have a more accurate understanding of quality criteria and, therefore, produce assessments more closely aligned with those of teachers (Sánchez-García et al., 2026; Sánchez-Oliver et al., 2026). However, empirical evidence in this area is still limited, especially in complex tasks and in transversal competences such as oral presentations.

Oral presentations are particularly suitable for this type of analysis. They integrate conceptual, communicative, organisational and social dimensions, and require the simultaneous application of multiple quality criteria (Castejón et al., 2018; Chireac & Martín-Vegas, 2022; Fernández García et al., 2008). In addition, their assessment through analytic rubrics makes it possible to break down evaluative judgement and to examine in detail areas of agreement and disagreement between assessors (López-Pastor & Pérez-Pueyo, 2017; Velasco-Martínez & Tójar, 2017).

For the purposes of the present study, evaluative judgement is understood as the capacity to make informed decisions about the quality of work, both one's own and that of others, based on explicit criteria and evidence (Tai et al., 2018; Boud et al., 2018). This concept focuses specifically on the ability to interpret standards and apply them to concrete performances. In this context, convergence with teacher judgement is not interpreted as a direct measure of evaluative competence, but rather as an operational indicator of calibration with an institutionally legitimised reference point (McConlogue, 2020).

Based on the above, the aim of the present study is to analyse differences between self-assessment, peer assessment and teacher assessment in the evaluation of university oral presentations. More specifically, it seeks to examine whether students' academic performance is associated with greater convergence with teacher assessment, both in the overall score and in each rubric item. The main contribution of this study lies in the combined analysis of assessment modality and academic performance, and in the use of a distance-based index to estimate convergence with the teacher criterion both at overall level and item by item.

METHODS

Study design

An observational, descriptive-analytical and cross-sectional study was conducted. It was based on the secondary analysis of data derived from a triadic assessment system implemented in the context of university teaching. The ex post facto design made it possible to analyse relationships and differences between variables without any experimental manipulation. To analyse the role of academic performance, the students' final course grades were categorised into four ordered levels: pass, good, very good and excellent. This classification was used as the academic performance variable in all subsequent analyses.

Sample

The sample consisted of 6,191 assessment records corresponding to theoretical oral presentations given by university students. Each record represented a single assessment, either self-assessment, peer assessment or teacher assessment. Of the total number of assessments, 1,112 were self-assessments, 4,824 were peer assessments and 255 were teacher assessments. Consequently, 96.0% of the assessments were carried out by students and 4.1% by teachers. Among the assessors, 74.0% were men and 26.0% were women. All teacher assessments were carried out by male staff. Student assessments were linked to four levels of final academic performance in the course: pass, good, very good and excellent. Regarding academic performance, student assessments were distributed as follows: pass 2.7%, good 49.1%, very good 36.1%, and excellent 12.1%.

The lower number of assessments across modalities can be attributed to the reduced number of evaluators involved in teacher assessment compared to peer assessment, as well as in self-assessment relative to peer assessment.

Instrument

Assessment was carried out using a validated rubric composed of eight items, each rated on a scale from 1 to 4 points: content, organisation of information, presentation, oral expression, non-verbal language, time management, supporting materials and teamwork (Sánchez-Oliver et al., 2025). The final score for the oral presentation was calculated as the arithmetic mean of the eight items. The instrument showed adequate internal consistency, both overall and by assessment modality (see Results), which supports its use as a measure of oral presentation quality.

Procedure

Assessments were carried out in courses from the Degree in Physical Activity and Sport Sciences at the University of Seville during the 2024–2025 academic year. Presentations took place within the usual teaching context of the courses involved. After each presentation, students completed their self-assessment and the peer assessment of their classmates using the rubric. The presentations were delivered in small groups, and the assessment referred to the group performance as presented in class. At the same time, teachers carried out the teacher assessment using the same criteria. Teacher assessment was used as the expert reference point, not because it is assumed to be infallible, but because it represents the institutionally legitimised judgement within the context of higher education assessment.

Although teacher assessment is used here as a reference point for analytical purposes, it should not be interpreted as an absolute or infallible standard of quality. Rather, it represents a socially and institutionally legitimised benchmark within a given educational context. As such, convergence with teacher judgement should be understood as an operational indicator of alignment with shared standards, rather than as a direct measure of accuracy or evaluative competence. As previous authors have noted, no single judgement can be considered an absolute gold standard; however, teacher assessment provides a stable and socially recognised benchmark against which student judgements can be compared for analytical purposes (Boud & Falchikov, 2007; McConlogue, 2020). This approach makes it possible to examine whether students' judgements progressively align with shared standards of quality, while acknowledging that agreement with the teacher does not, by itself, guarantee high-quality evaluative judgement.

Data analysis

Given the ordinal nature of the scales and the heterogeneity of sample sizes, non-parametric procedures were used. Descriptive statistics were calculated (mean, standard deviation, median and sample size). Reliability was assessed using Cronbach's alpha. Differences between assessment modalities were analysed using the Kruskal–Wallis test, with Mann–Whitney post hoc tests and Bonferroni correction. Associations with academic performance were examined using Spearman's rho. Convergence with teacher assessment was estimated using an absolute distance index between student scores and the mean teacher score. The use of absolute distance allows a direct and comparable interpretation of the degree of convergence between assessors and facilitates the analysis of the progressive alignment of students' evaluative judgement with the expert reference.

Data was analysed using Python (version 3.11).

RESULTS

The evaluation rubric showed high internal consistency ($\alpha = .85$). When broken down by assessment modality, reliability was high in peer assessment ($\alpha = .85$), slightly lower in self-assessment ($\alpha = .82$), and acceptable in teacher assessment ($\alpha = .75$). These values indicate that the items consistently measure the overall construct of oral presentation quality across the three modalities.

Differences between assessment modalities

Statistically significant differences were found in the mean oral presentation score according to the type of assessment ($H = 250.54$; $p < .001$). Self-assessment produced the highest scores ($M = 3.75$, $SD = 0.32$), followed by peer assessment ($M = 3.58$, $SD = 0.41$) and teacher assessment ($M = 3.45$, $SD = 0.41$). Post hoc comparisons confirmed significant differences between all pairs, with small to moderate effect sizes. The largest difference was observed between self-assessment and teaching assessment (Figure 1).

A particularly relevant finding is that these differences between modalities appear in all items, although with varying intensity. Item-level analyses revealed statistically significant differences between modalities in all assessed dimensions ($p < .001$). The largest discrepancies were observed in the items content, presentation and time management, where teacher assessment was systematically more demanding than both self- and peer assessment. In contrast, items such as teamwork and organisation of information showed more moderate differences, suggesting greater convergence in dimensions that are more observable or more easily shared among assessors (Figure 1).

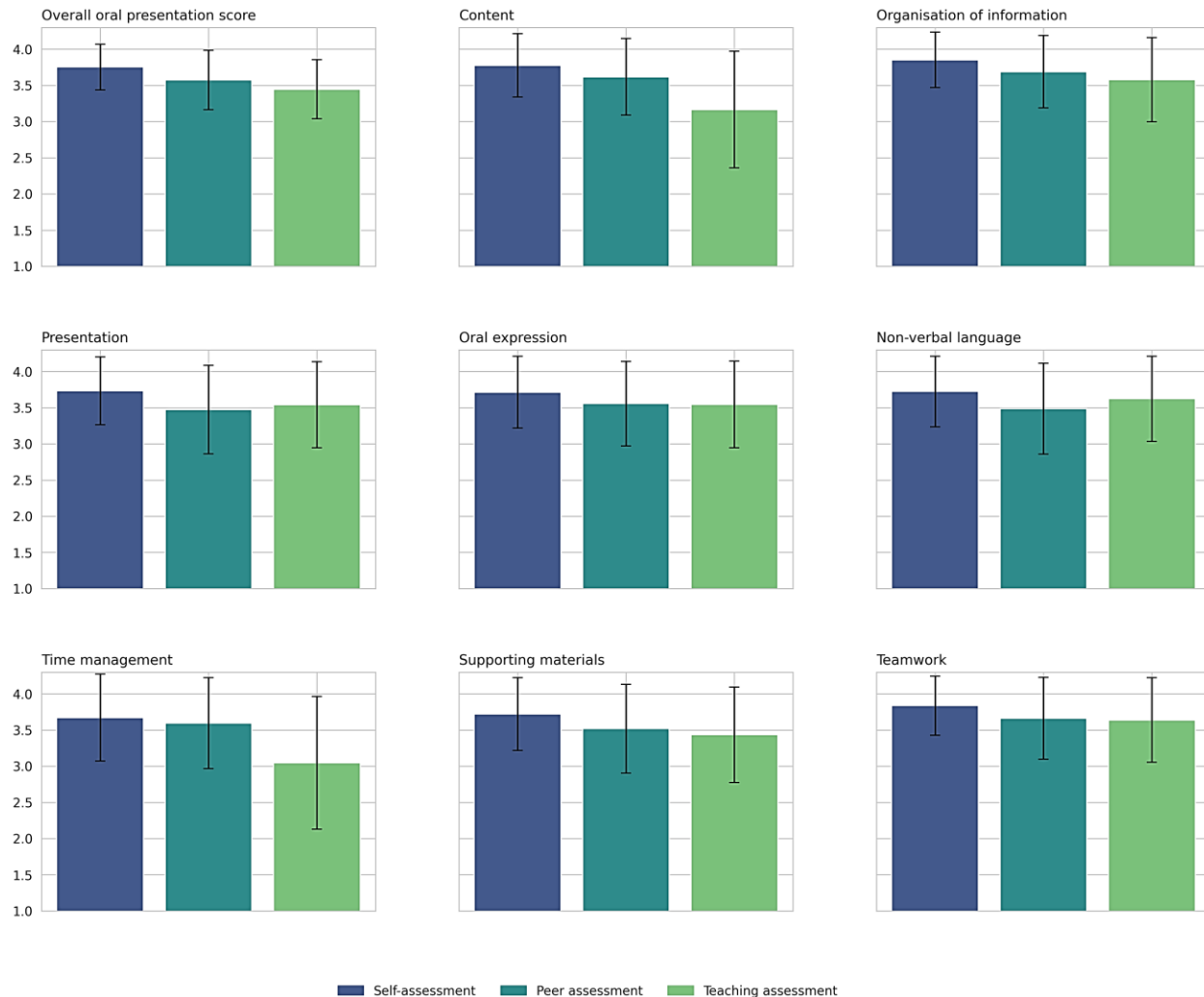


Figure 1. Overall and item-level mean scores by evaluation modality (self-assessment; peer assessment; teacher assessment) (Mean \pm SD; Mean score: 1-4).

Relationship between presentation assessment and academic performance

Of particular interest is the fact that the mean presentation score was positively associated with the final course grade. This association was stronger in self-assessment than in peer assessment. The mean oral presentation score showed a positive and significant correlation with the final course grade ($\rho = .16$; $p < .001$). This pattern was maintained when analysed by modality, although with different magnitudes. The correlation was moderate for self-assessment ($\rho = .32$) and low for peer assessment ($\rho = .12$). In addition, significant differences were found in oral presentation scores according to the final course grade ($H = 195.85$;

$p < .001$). Students with very good and excellent grades generally obtained higher presentation scores than those with pass or good grades (Table 1).

Table 1. Oral presentation scores by final course grade (pass, good, very good and excellent).

Course mark	M ± SD	Median	n
Pass	3.72 ± 0.40	4.00	161
Good	3.55 ± 0.41	3.62	2.915
Very Good	3.66 ± 0.39	3.75	2.142
Excellent	3.69 ± 0.37	3.75	718

Note. M: Mean; SD: Standard Deviation.

Regarding convergence with teacher assessment, a key result is that this convergence varies according to the final course grade, with significant differences in both self- and peer assessment. Analysis of the distance between student and teacher assessments showed significant differences according to the final course grade, both for self-assessment ($H = 36.94$; $p < .001$) and for peer assessment ($H = 63.38$; $p < .001$). Overall, students with higher final course grades tended to produce assessments closer to the teacher criterion, particularly in peer assessment. However, the pattern was not strictly linear. In some cases, greater distances were observed at the highest levels of academic performance (Figure 2). It should be noted that, despite their statistical significance, these differences show small to moderate magnitudes. This is consistent with the progressive and formative nature of the development of evaluative judgement.



Figure 2. Convergence with teacher assessment according to final course grade (pass, good, very good and excellent).

Figure 3 shows the mean distance from the teacher assessment for each item regarding academic performance. A general trend towards greater convergence as academic performance increases can be observed, although with variations depending on the dimension assessed. The largest distances were found in the items content, presentation and time management. In contrast, organisation of information and teamwork showed closer alignment with the teacher criteria across all academic levels.

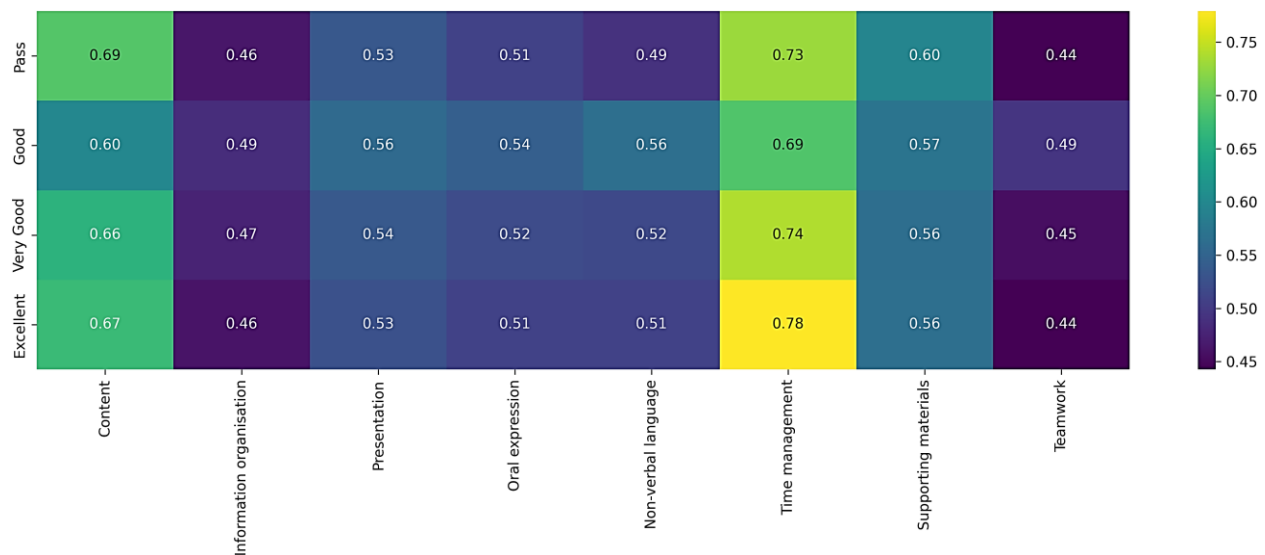


Figure 3. Distance to teacher assessment by item according to final course grade (pass, good, very good and excellent).

DISCUSSION

The findings, in this context and in similar higher education settings, confirm that the assessment of university oral presentations varies systematically according to the assessor. A clear and robust pattern shows that self-assessment produces the highest scores, peer assessment occupies an intermediate position, and teacher assessment is the most demanding. Although the observed relationships and differences are statistically significant, their magnitudes are small to moderate and should be interpreted in educational rather than deterministic terms.

Specifically, the findings show a clear gradient in triadic assessment, with higher scores in self-assessment ($M = 3.75$) than in peer assessment ($M = 3.58$), and higher scores in peer assessment than in teacher assessment ($M = 3.45$; $H = 250.54$; $p < .001$). This reproduces both the pattern observed in earlier studies and the one described in the literature on self-, peer- and teacher assessment in higher education (Topping, 1998, 2009; Falchikov & Goldfinch, 2000; Boud et al., 2018). This gradient is consistent with the idea that evaluative judgement may be associated with repeated exposure to explicit criteria, assessment practice and the use of quality evidence, rather than through grading alone (Boud & Falchikov, 2007; Boud et al., 2018).

The internal consistency of the rubric ($\alpha = .85$) supports its use to discriminate between levels of quality and suggests that the observed differences are more likely to reflect assessment processes than measurement error. This is in line with reviews showing that well-designed rubrics with clear descriptors support shared interpretations of quality (Reddy & Andrade, 2010; Panadero & Jonsson, 2013). The slightly lower alpha in teacher assessment is compatible with the application of more demanding and more differentiated criteria in dimensions with a higher conceptual load, such as content and presentation. This has also been noted in studies that point to limitations in the design and use of competence-based rubrics in higher education and to the need to improve their descriptive precision (Velasco-Martínez & Tójar, 2018; Reddy & Andrade, 2010).

From an educational validity perspective, the presentation score shows a positive association with the final course grade ($\rho = .16$; $p < .001$), with a stronger relationship in self-assessment ($\rho = .32$) than in peer

assessment ($\rho = .12$). This suggests that the task captures relevant transversal competences, such as communication, organisation and time management, without being reduced to a purely summative function (Sánchez-García et al., 2025). The moderate size of this association also supports the use of complementary assessment instruments to reflect the complexity of learning in academic communication tasks (Panadero & Jonsson, 2013). Although statistically significant, the magnitude of this association is modest, suggesting that academic performance explains only a limited proportion of the variance in assessment convergence.

A central contribution of this study is the finding that convergence with the teacher criterion is positively associated with academic performance, as reflected in the final course grade. Significant differences were found in the distance to teacher assessment in both self- and peer assessment, with a more consistent pattern in peer assessment ($H_{\text{self}} = 36.94$; $H_{\text{peer}} = 63.38$; $p < .001$). This pattern of alignment can be interpreted as indicative of calibration of evaluative judgement. This interpretation should be treated with caution. Convergence with teacher assessment does not necessarily reflect more accurate evaluative judgement, as teacher judgement is context-dependent and variable. Thus, alignment should be understood as calibration with an institutional reference rather than as a definitive indicator of evaluative quality, while divergence may also reflect alternative interpretations of the criteria. It is also consistent with the perspective of sustainable assessment, which aims to develop transferable evaluative capacities, and with recent proposals that integrate rubrics, exemplars, self- and peer assessment, and feedback to support the development of students' judgement (Boud & Soler, 2016; Tai et al., 2018). In this sense, the triadic structure facilitates the comparison of perspectives and the explicit discussion of criteria, as already suggested in early work on triadic assessment (Gale et al., 2002).

However, the pattern is not strictly linear. In some cases, students with excellent grades show distances similar to, or even greater than, those observed in other groups. This finding deserves a cautious and theoretically informed interpretation. Students with higher levels of mastery may tend to apply more demanding personal standards, interpret the criteria in a more critical or expert-like way, or show calibration effects associated with highly self-regulated profiles. In this sense, greater divergence from teacher judgement does not necessarily reflect poorer evaluative judgement but may be related to the adoption of stricter or more autonomous quality standards (McConlogue, 2020; Nicol & Macfarlane-Dick, 2006; Hattie & Timperley, 2007). This non-linear pattern reinforces the idea that the relationship between academic performance and evaluative alignment is complex and likely influenced by additional cognitive, motivational and contextual factors.

The item-level analysis provides useful guidance for teaching practice. The largest discrepancies are found in content, presentation and time management, whereas organisation of information and teamwork show greater convergence. This pattern is consistent with the meta-analysis by Falchikov and Goldfinch (2000), which showed that agreement between peer and teacher marks tends to decrease when assessment is multidimensional and analytic, and with reviews that stress the importance of precise descriptors to support the validity and reliability of rubrics (Reddy & Andrade, 2010). In the Spanish higher education context, limitations in the design and use of competence-based rubrics have been reported, which may partly explain item-level variability and point to the need for further teacher training in this area (Velasco-Martínez & Tójar, 2018).

Oral presentations provide a particularly suitable context for studying evaluative judgement. They integrate conceptual, communicative, organisational and social dimensions, and their assessment through analytic rubrics makes it possible to identify specific areas of agreement and disagreement that are useful for teaching (Castejón et al., 2018). In addition, Spanish research on oral communication competence has been relatively

limited in recent years, which reinforces the interest of studies such as the present one and the need for clearer pedagogical models to guide practice (Chireac & Martín-Vegas, 2022). Taken together, these results suggest that triadic assessment may support the development of students' capacity to judge quality in complex academic tasks (McConlogue, 2020).

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the design is observational and cross-sectional, which allows the identification of associations but does not support causal or developmental inferences regarding the relationship between academic performance and convergence with the teacher criterion.

Second, although the dataset is large, there are important constraints related to representativeness. The distribution of assessment records is markedly unbalanced, with the vast majority of evaluations conducted by students (self- and peer-assessment) and a comparatively small proportion by teachers. This asymmetry reflects the pedagogical design but may influence variability, dispersion and comparability across assessment modalities.

Third, the teacher assessment used as the expert reference comes from a limited group of teachers within the same institutional context. In addition, the study is confined to a single university and a specific degree programme. These factors may reduce the generalisability of the findings to other institutional settings, disciplines or assessment cultures, and future research would benefit from including more diverse samples and greater inter-teacher variability.

Finally, the study focuses on theoretical oral presentations, which represent a specific type of complex task. Caution should therefore be exercised when extrapolating the results to other forms of assessment or learning outcomes. Despite these limitations, the large sample size and the adequate reliability of the rubric strengthen the internal robustness of the findings.

Implications and future research

Several lines of work appear particularly promising for further progress in this field. First, longitudinal or quasi-experimental studies are needed to examine the development of evaluative judgement over time and to test whether sustained participation in triadic assessment increases convergence with teacher assessment. Second, it would be useful to evaluate specific training interventions, such as calibration with anchored examples, commented rubrics or comparative feedback, and to analyse their impact at item level. This seems especially relevant for content, presentation and time management, where the largest discrepancies were observed. Third, future studies should include samples with greater diversity of teachers and courses, in order to estimate inter-teacher variability and improve external validity. Finally, it would be valuable to incorporate psycho-educational variables, such as self-efficacy, goal orientation, previous experience with rubrics or motivation. These variables may help to explain when and why certain profiles, including high-performing students, diverge from the teacher criterion. It would also be useful to replicate this approach in other assessable products, such as written assignments, practical tasks, projects or competence-based assessments.

CONCLUSIONS

Triadic assessment is confirmed as a valid strategy for the assessment of university oral presentations. Academic performance, operationalised here as the final course grade, is associated with closer alignment

with the teacher criterion, particularly in peer assessment. However, this convergence is neither linear nor automatic. These findings reinforce the formative value of assessment and highlight the need to design specific strategies to support the development of students' evaluative judgement. Triadic assessment makes it possible to make discrepancies visible, to promote reflection on quality, and to create learning opportunities that are unlikely to emerge in exclusively teacher-led assessment models.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship in accordance with established ethical guidelines. Contributions are specified according to the CRediT (Contributor Roles Taxonomy) as follows:

Conceptualization: M.G.-P.; A.J.S.-O. Methodology: A.S.-G.; M.G.-P.; A.J.S.-O. Formal analysis: A.S.-G.; A.J.S.-O. Investigation: A.S.-G.; M.G.-P.; A.M.-L. Data curation: A.S.-G.; M.G.-P.; A.M.-L. Writing – original draft: A.S.-G.; A.J.S.-O. Writing – review & editing: A.J.S.-O.; M.G.-P.; A.M.-L. Supervision: A.J.S.-O. All authors have critically reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

FUNDING

This study was developed within a project funded by the IV Own Teaching Plan of the University of Seville (2023 call), under the framework of faculty activities for educational innovation (L.2.2), specifically Action 221 devoted to supporting teaching innovation. Moreover, it forms part of the doctoral dissertation of Andrés Sánchez García.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

AI USE DISCLOSURE

In accordance with current publishing ethics and transparency recommendations, during the preparation of this manuscript, the authors utilised Microsoft Copilot, Gemini, and Grammarly to translate and proofread the final manuscript before sending it to a native English translator, who further checked and proofread the final text. After employing these tools, the authors reviewed and edited the content as necessary. The authors retain full responsibility for the content of the manuscript and confirm its originality, integrity, and accuracy.

DATA AVAILABILITY STATEMENT

In alignment with Research Data Policies promoting open science, the dataset from this study is accessible to interested researchers via the Figshare repository: [10.6084/m9.figshare.31268428](https://doi.org/10.6084/m9.figshare.31268428).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the collaboration of the faculty and students of the Degree in Physical Activity and Sports Sciences at the University of Seville, whose active involvement was instrumental in advancing communicative competence and enhancing formative assessment practices.

REFERENCES

- Basurto-Mendoza, S. T., Moreira-Cedeño, J. A., Velásquez-Espinales, A. N., & Rodríguez-Gámez, M. (2021). Autoevaluación, Coevaluación y Heteroevaluación como enfoque innovador en la práctica pedagógica y su efecto en el proceso de enseñanza-aprendizaje. *Polo del Conocimiento*, 6(3), 828-845.
- Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. Routledge. <https://doi.org/10.4324/9780203964309>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>
- Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (Eds.). (2018). *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*. Routledge. <https://doi.org/10.4324/9781315109251>
- Bozkurt, F. (2020). Teacher candidates' views on self and peer assessment as a tool for student development. *Australian Journal of Teacher Education*, 45 (1). <https://doi.org/10.14221/ajte.2020v45n1.4>
- Castejón, F. J., Santos Pastor, M., & Cañadas, L. (2018). Desarrollo de competencias docentes en la Formación Inicial del Profesorado de Educación Física. Relación con los instrumentos de evaluación. *Estudios Pedagógicos (Valdivia)*, 44(2), 111-126. <https://doi.org/10.4067/S0718-07052018000200111>
- Chireac, S. M., & Martín-Vegas, R. A. (2022). A bibliometric analysis of the Spanish scientific production on teaching oral communication between 2010 and 2020. *TEM Journal*, 11(4), 1620-1625. <https://doi.org/10.18421/TEM114-24>
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. <https://doi.org/10.3102/00346543070003287>
- Fernández García, C. M., Torío López, S., Viñuela Hernández, M. P., Molina Martín, S., y Bermúdez Rey, T. (2008). La comunicación oral como competencia transversal de los estudiantes de pedagogía y magisterio: Presentación del diseño metodológico de una innovación para su trabajo en el aula. *Revista de Enseñanza Universitaria*, 31, 26-38.
- Gale, K., Martin, K., & McQueen, G. (2002). Triadic assessment. *Assessment & Evaluation in Higher Education*, 27(6), 557-567. <https://doi.org/10.1080/0260293022000020318>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Kambourova, M., González-Agudelo, E. M., & Grisales-Franco, L. M. (2021). La autoevaluación del estudiante universitario: revisión de la literatura. *Teoría de la Educación: Revista Interuniversitaria*, 33, 2, 2021, 217-264. <https://doi.org/10.14201/teri.23672>
- López Pastor, V. M., & Pérez Pueyo, Á. (2017). Introducción a la creación y el uso de escalas descriptivas y rúbricas. En V. M. López Pastor & Á. Pérez Pueyo (Coords.), *Evaluación formativa y compartida en educación: Experiencias de éxito en todas las etapas educativas* (pp. 92-115). Universidad de León.
- McConlogue, T. (2020). *Assessment and feedback in higher education: A guide for teachers*. UCL Press.
- Mendoza, S. T. B., Cedeño, J. A. M., Espinales, A. N. V., & Gámez, M. R. (2021). Autoevaluación, Coevaluación y Heteroevaluación como enfoque innovador en la práctica pedagógica y su efecto en el proceso de enseñanza-aprendizaje. *Polo del Conocimiento: Revista científico-profesional*, 6(3), 828-845.

- Nicol, D. J., & Macfarlane Dick, D. (2006). Formative assessment and self regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Olague, M. Z., Cañadas, L., & Manso, J. (2025). Percepción del profesorado de educación básica sobre la participación de los diferentes agentes (profesorado y alumnado) en los procesos de evaluación. *Revista Fuentes*, 27(3), 289-300. <https://doi.org/10.12795/revistafuentes.2025.27856>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Pascual-Arias, C., López-Pastór, V. M., & Hórtiguera-Alcala, D. (2023). Estudio longitudinal sobre los efectos del desarrollo de la Evaluación Formativa y Compartida en la Formación Inicial del Profesorado. *Cultura, Ciencia y Deporte*, 18(55), 5-31. <https://doi.org/10.12800/ccd.v18i55.1904>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>
- Sánchez-García, A., Sanchez-Oliver, A. J., & Grimaldi-Puyana, M. (2026). Triadic formative assessment in higher education: interaction effects between assessment modality and evaluator sex. *Retos*, 77, 346-357. <https://doi.org/10.47197/retos.v77.118382>
- Sánchez-Oliver, A. J., Carnero-Díaz, Á., Muñoz-Llerena, A., & Bianchi, P. (2025). Proyecto de innovación docente para la implementación de la evaluación formativa en la universidad. In *Acciones educativas innovadoras en el ámbito universitario* (pp. 85-103). Dykinson.
- Sánchez-Oliver, A. J., Rodríguez Rojas, F. J., Feria-Madueño, A., Muñoz-López, A., Carnero Díaz, Ángel, Muñoz-Llerena, A., Sañudo-Corrales, B., Oviedo-Caro, M. Ángel, Grimaldi-Puyana, M., Bianchi, P., Domínguez, R., & Angosto, S. (2025). Validation of a tool to evaluate theoretical-practical presentations in Higher Education. *Retos*, 67, 903-916. <https://doi.org/10.47197/retos.v67.112987>
- Sánchez-Oliver, A. J., Sánchez-García, A., Feria-Madueño, A., Muñoz-López, A., Carnero Díaz, Á., Muñoz-Llerena, A., Sañudo-Corrales, B., Oviedo-Caro, M. Á., Grimaldi-Puyana, M., Bianchi, P., & Domínguez, R. (2026). Evaluating oral presentations in university students: Self, peer, and hetero evaluation. *Journal of Sport and Health Research*. In press.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467-481. <https://doi.org/10.1007/s10734-017-0220-3>
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276. <https://doi.org/10.3102/00346543068003249>
- Topping, K. J. (2009). Peer assessment. *Theory Into Practice*, 48(1), 20-27. <https://doi.org/10.1080/00405840802577569>
- Velasco-Martínez, L. C., & Tójar Hurtado, J. C. (2018). Uso de rúbricas en educación superior y evaluación de competencias. *Profesorado, Revista de Currículum y Formación del Profesorado*, 22(3), 183-208. <https://doi.org/10.30827/profesorado.v22i3.7998>

