Beyond the Accuracy Gap in Self-assessment: Exploring Reasons For (In)Accuracy and the Role of Individual Differences.

Leire Pinedo [1]; Ernesto Panadero [2, 1 & 3]; Javier Fernández-Ruiz[4] & Daniel García-Pérez[5]

[1] Education Regulated Learning and Assessment group (ERLA group), Facultad de Educación y Deporte, Universidad de Deusto, Bilbao, Spain.

[2] Centre for Assessment Research Policy and Practice in Education (CARPE), School of Policy and Practice, Institute of Education, St. Patrick's Campus, Dublin City University, Dublin, Ireland.

[3] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

[4] Departamento de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid, Madrid, Spain.

[5] Departamento de Investigación en Psicología y Educación, Facultad de Educación-Centro de Formación del Profesorado, Universidad Complutense de Madrid, Spain

Author Note

Leire Pinedo https://orcid.org/0000-0002-3046-5226

Ernesto Panadero https://orcid.org/0000-0003-0859-3616

Javier Fernández-Ruiz https://orcid.org/0000-0001-5419-7687

Daniel García-Pérez https://orcid.org/0000-0002-5697-1792

Recommended citation:

Pinedo, L., Panadero, E., Fernández-Ruiz, J., & García-Pérez, D. (2025). Beyond the Accuracy Gap in Self-assessment: Exploring Reasons for (In)Accuracy and the Role of Individual Differences. *Assessment in Education: Principles, Policy & Practice, 32*(4), 462–484. https://doi.org/10.1080/0969594X.2025.2565384

Funding: (1) Spanish Ministry of Universities (Ministerio de Universidades) National FPU Call (Formación de Profesorado Universitario), personal grant, reference: FPU19/04661. (2) Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad) National I+D Call (Convocatoria Excelencia) project reference EDU2016-79714-P. (3) Spanish National R+D call from the Ministerio de Ciencia, Innovación y Universidades (Generación del conocimiento 2020), Reference number: PID2019-108982GB-I00. (4) Basque Government Call for Grants to support the activities of research groups of the Basque University System (2022-2025) project reference: IT1624-22.

Correspondence concerning this article should be addressed to Leire Pinedo, ERLA Research Group, Universidad de Deusto, Avda. Universidades 24. 48007 Bilbao, España. Email: leire.research@gmail.com

The authors declare to not having any conflict of interest regarding this manuscript. This research has been approved by the ethics committee from Comité de ética de la investigación, Universidad Autónoma de Madrid. Reference: CEI-84-1557. PI: Ernesto Panadero.

Abstract

While accuracy is the most studied aspect in self-assessment, there is a need to understand why higher education students are inaccurate. This study undertakes a dual exploration, examining the effects of individual differences on self-assessment accuracy, and students' perspectives on the reasons for their inaccuracies. A total of 112 psychology students self-scored an essay, then received the instructor's score, and were interviewed. Results revealed: (1) a weak positive correlation between self- and instructor scores, with most students slightly underestimating; (2) a significant effect of essay score, as higher achievers tended to underestimate and lower achievers to overestimate; (3)a small but significant effect of learning-oriented self-regulation, with higher scorers tending to underestimate; (4) inaccuracies were rarely attributed to assessment criteria but instead reflected internal perceptions; (5) high achievers justified underestimation through pessimistic self-perceptions, whereas low achievers explained overestimation through optimism and overconfidence. The findings reveal that self-assessment is as much an affective process as it is a cognitive skill.

Keywords: self-assessment; self-assessment accuracy; academic performance; self-regulation; higher education;

Beyond the Accuracy Gap in Self-assessment: Exploring Reasons For (In)Accuracy and Effect of Individual Differences

Self-assessment is considered an essential learning strategy in higher education since it has a positive effect on academic performance and it plays an important role in self-regulated learning (Brown & Harris, 2013; Kostons et al., 2012). Among the research lines of self-assessment, self-assessment accuracy, usually evaluated through scoring accuracy (i.e., alignment between student self-score about her work compared to an expert's score) has been the dominant line of work (Andrade, 2019; León et al., 2023). The same reviews revealed an overestimation tendency in self-assessment accuracy, highlighting the need to explore the effect of individual variables such as academic performance and self-regulation on scoring accuracy. Moreover, a comprehensive understanding of the self-assessment accuracy requires complementing the quantitative approach with qualitative exploration of the internal mechanisms that shape students' judgement (Andrade, 2019; Panadero et al., 2024). By identifying the reasons students attributed to their inaccuracy, as well as the individual differences that affect accuracy, we could provide help to the students who may need personalized feedback, or additional aids tailored to address their specific challenges in self-assessment. Therefore, this study contributes to the understanding of the self-assessment accuracy from students' perspectives, by depicting the reasons behind the inaccuracy, and the effect of the individual variables.

Self-assessment Accuracy

Self-assessment encompasses "a wide variety of mechanisms and techniques through which students describe (i.e. assess) and possibly assign merit or worth to (i.e. evaluate) the qualities of their own learning processes and products" (Panadero et al., 2016, p. 804). Additionally, it can serve formative and/or summative purposes with a

focus on different elements such as competence, processes and/or products (Andrade, 2019).

Self-assessment accuracy, sometimes referred to as consistency (Andrade, 2019), usually involves comparing students' self-assessment with judgements from qualified raters (Brown et al., 2015; Leon et al., 2023). This often entails students assigning a score to their work and subsequently comparing it with an unbiased expert score to discern any discrepancies, referred to as scoring accuracy. Brown and Harris (2013) pointed out the importance of the scoring accuracy from psychometrics and learning theory perspectives, as it could help students know what went wrong and how to correct it.

Studies have shown that integrating self-scoring into the assessment process enhances students' performance in subsequent tests when compared to control groups (Sanchez et al., 2017) and promotes students' high level of deep approach to learning (Nieminen et al., 2021). However, as the Dunning-Krueger effect shows (Kruger & Dunning, 1999), humans tend to be unrealistically optimistic about their abilities and believe that they are above average (Dunning et al., 2004). This tendency of overestimation has already been proved by different meta-analyses (Falchikov & Boud, 1989; León et al., 2023), as well as the limited overlap between the student's self-score and the expert's score (Brown & Harris, 2013). Therefore, it is important not only to continue examining if higher education students can accurately predict their score in a task, but also to uncover the underlying factors contributing to their inaccuracies, as these have barely been investigated.

Additionally, there is also a call in the educational assessment field to understand the internal mechanisms of self-assessment such as cognitive processes (Andrade, 2019; Butler, 2018; Panadero et al., 2024). In this study, we go beyond

assessing students' scoring accuracy in predicting their performance; we aim to understand the reasons behind the inaccuracy from the students' perspective.

Reasons for (In)Accuracy

Given the considerable body of evidence suggesting that students tend to be inaccurate in their self-assessments (Brown & Harris, 2013; Falchikov & Boud, 1989; León et al., 2023), a key question arises: What are the reasons behind their tendency to overestimate?

Dunning et al. (2004) proposed several reasons for the flawed self-assessment, such as (1) to be unrealistically optimistic about their own abilities, (2) to believe they are above average, (3) to ignore crucial information that should be used for assessing their work, and (4) to have deficits in the information. There is also the double handicap; if students lack domain knowledge, they will not be aware that they are not good in the domain (Dunning et al., 2004). Other factors that might influence accuracy include students' lack of awareness of assessment expectations, tendency to evaluate their own potential rather than real performance, and overconfidence or desire to impress (Evans et al., 2002). Rees and Shepherd's (2005) study explored why students tend to overestimate their performance from students' and assessors' views. Participants attributed their overestimation of performance to several reasons, such as insufficient feedback, lack of peer comparisons, and high expectations from assessors, while their assessors referred to students' previous academic success, lack of good quality feedback, downward comparison with peers, and gender.

Additionally, we should consider that self-scoring may be less accurate if it is not implemented under the right classroom assessment conditions. For example, the use of concrete, task-specific and well-understood assessment criteria (Andrade & Valtcheva, 2009; Butler, 2018) can lead to a calibrated self-scoring. Indeed, students

who are engaged in developing their criteria (Brown & Harris, 2013) and employ more specific standards (Kostons et al., 2012; Rust et al., 2003) seem to be more accurate. On the contrary, when students' assessment relies on internal criteria (i.e., effort) rather than objective standards, their self-assessment may become inaccurate (Taras, 2003).

However, the vast majority of research on factors influencing inaccuracy employed a quantitative approach, which might not explain why students find self-assessment accuracy challenging (Rees & Shepherd, 2005). Therefore, it is important to examine the possible reasons behind the self-assessment accuracy through students' lenses, and combine this with the quantitative approach in order to get a more accurate representation of the phenomenon and identify the effect of the individual differences.

Effects of Individual Differences on Self-assessment Accuracy

An additional aspect influencing accuracy is students' characteristics (Brown & Harris, 2013; Dunning & Krueger, 1999). In this section, we present the individual variables explored in this study: previous academic performance, year level, self-regulation and self-efficacy.

First, previous academic performance is related to self-assessment accuracy as students with higher academic performance seem to be more consistent with teacher scores and more severe in their ratings (Barnett & Hixon, 1997; Brown & Harris, 2013; Brown et al., 2015). This pattern is often attributed to their enhanced ability to monitor and judge their own learning processes (Lew et al., 2010). However, consistency does not necessarily imply accuracy, as higher achievers frequently underestimate their performance (negative bias), while lower achievers tend to overestimate (positive bias) (Boud et al., 2013, 2015; Dunning, 2011; Dunlosky & Rawson, 2012; Falchikov & Boud, 1989; Knof et al., 2024; Tejeiro et al., 2012). Such miscalibrations among lower

achievers may reflect difficulties in interpreting assessment criteria (Kostons, 2009) or a tendency to adopt more lenient standards (Brown & Harris, 2013; Pinedo et al., 2023).

Second, regarding year level, different authors claimed that self-assessment ability might be influenced by students' educational experience (Brown & Harris, 2013; Panadero et al., 2016; Ross et al., 2002). This has been probed in different studies which showed that older students' self-scores tend to be more aligned with teacher scores (Bradshaw, 2001; Elder, 2010), whereas younger students overestimate their self-scoring (Kaderavek et al., 2004; Ross et al., 2002). One reason for this difference could be that while higher education students seem to perform a more frequent and advanced use of self-assessment (Pinedo et al., 2023), participants with less expertise may find this process complex (Kostons et al., 2009; 2010) and perform it less regularly. Even though there is ample evidence of educational level effects on accuracy, it is unclear how accuracy develops during their university courses (Panadero et al., 2023). Thus, a study with consecutive years of study might provide insights into the influence of year level on the accuracy.

Third, self-assessment is an embedded process in the self-regulated models, with several studies demonstrating the positive effects of self-assessment on self-regulated learning (for a review, see Panadero et al., 2017). Self-assessment practices at various stages of self-regulated learning had a positive impact on academic achievement (Yan, 2020). However, to our knowledge, it has not been studied whether self-regulated learning skills can predict accuracy. Understanding the self-regulatory factors that contribute to accurate self-assessment could aid in developing interventions to improve self-regulatory skills and in turn, academic performance.

Fourth, self-efficacy has proven to be one of the strongest predictors of academic performance in university students (Richardson et al., 2012). While self-efficacy has

been explored as a dependent variable of self-assessment interventions (e.g., Panadero et al., 2017), there are only a few examples exploring how scoring accuracy and self-efficacy are related. For instance, Ng and Earl (2008) found no differences in self-efficacy among three accuracy groups (understimators, overestimators, and accurate). Here, we will continue investigating this area.

Aim and Research Questions

This study aims to investigate the accuracy of self-assessment among higher education students, explore the causes of any inaccuracies and compare them across various individual differences. The study is organised around four research questions (RQs):

RQ1. How accurate are higher education students in their self-assessment?

RQ2. Did individual differences influence accuracy? (i.e., year level, essay score, self-efficacy, and self-regulation)

RQ3. What were the reasons university students provided for the discrepancy between their score and the instructor's score?

RQ4. Did individual differences influence the reasons reported? (i.e., year level, essay score, self-efficacy, and self-regulation)

Method

Participants

A total of 112 university psychology students (89% females) from first (39), second (38) and third (35) year levels participated in this study. A convenience sampling was used at a public university located in Madrid (Spain), where two of the authors were affiliated. Participants received credits (ETCS) as a reward for participation.

Data Collection and Instruments

Essay Score

The participants' scores were based on an essay they wrote in the study. Two people, the second author and the seminar instructor (4th author), separately reviewed and scored one-third of the essays using an analytic rubric made for this study (See Appendix A), resulting in high alignment in general grades following the rubric categories (Intra-class Correlation .91). After that, the seminar instructor scored the remaining essays.

Student Self-score

Participants assigned a score to their own essays, stating it out loud, on a scale from 0 to 10, without access to a rubric or other standardized scoring aid. This approach aligns with common pedagogical situations where students are expected to evaluate their own performance without detailed criteria.

Academic writing self-efficacy scale

This ad-hoc instrument measures students' perceptions of efficacy when writing an academic text. It comprised eight items and used a 7-point Likert scale (See Appendix B). As our self-assessment accuracy is derived from the second measurement occasion, we used the self-efficacy measurement collected at that time (before receiving the score). The internal consistency was α=.80.

Emotion and Motivation Self-Regulation Questionnaire (EMSR-Q) (Alonso-Tapia et al., 2014)

This questionnaire measures the use of emotional and motivational strategies in educational contexts. It includes 20 items to be answered on a 5-point Likert scale. The EMSR-Q is composed of 20 items grouped into five first-order factors: (1) Avoidance-oriented self-regulation, (2) Negative self-regulation of stress, (3) Performance-oriented self-regulation, (4) Process-oriented self-regulation, and (5) Positive self-regulation of motivation. Each subscale consists of four items. These five factors are then organized into two second-order dimensions: Learning Self-regulation Style (Process-oriented SR,

Performance-oriented SR, and Positive motivation SR; 8 items, α = .86) and Avoidance Self-regulation Style (Avoidance-oriented SR, Negative stress SR, and Performance-oriented SR; 12 items, α = .82). For a full explanation of the first- and second-order scales, see the EMSR-Q manual (Panadero et al., 2021).

Interview data

The participants answered the questions after receiving the instructor's score in the last phase of the experiment, and the interview lasted an average of 10 minutes. The two questions were: (1) *What are the reasons for the discrepancy between your score and the instructor's?,* and (2) *How do you feel about the score discrepancy?*

Procedure

This study is part of a project that encompasses a larger data collection (Panadero et al., 2023; 2024; Pinedo et al., 2023). Here, we focus on students' self-assessment accuracy, as well as the reasons they reported in relation to the score discrepancy and the individual differences.

Participants attended a seminar on academic writing and wrote an essay. After a few days, the participants visited the laboratory individually. First, participants were interviewed one-on-one discussion with a researcher. They talked about: (1) how they learned to self-assess, specifically who taught them to self-assess and through which activities they developed this skill, (2) the process they follow for self-assessment, and (3) how they make use of feedback for self-assessment.

Following the interview, participants filled out various questionnaires for the first time (i.e., self-efficacy and emotions). Next, participants were asked to perform a self-assessment of their essays, and think-aloud their thoughts, emotions and motivational reactions while doing the self-assessment. After the self-assessment participants were

asked to give their essays a score on a scale of 0 to 10. Immediately after, they filled out the questionnaires again to document any changes in their self-efficacy and emotions.

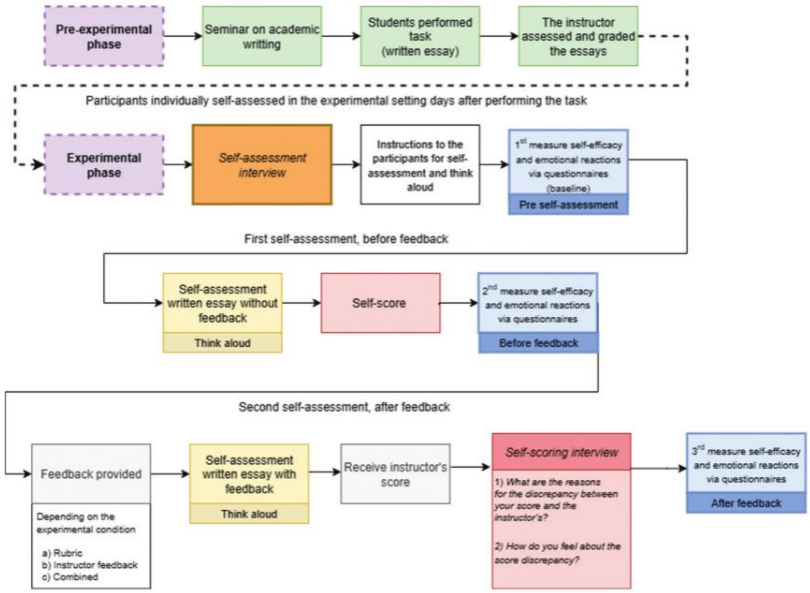
Subsequently, the participants received feedback on their essays, which varied according to the group they were assigned: (a) rubric (n=38), (b) feedback from an instructor (n=36), and (c) a combination of both (n=38). The rubric group received the rubric (Appendix A) and was instructed to use it for self-assessing their essay again. In the instructor's feedback group, participants received their essays with the instructor's comments. Lastly, the combined group received both. They were then asked to self-assess out loud their essays again with the feedback.

Finally, all participants were provided with their instructor's scores for their essays, and they participated in an interview. The participants then completed the questionnaires for the last time. The process was being video-recorded.

Results from the interview of the first phase and the experimental phase regarding the think-aloud protocols and questionnaires are presented in previous publications (Panadero et al., 2023; 2024; Pinedo et al., 2023).

Figure 1

Procedure of the study



Data coding and analysis

We employed a mixed-method approach, including qualitative data gathered through interviews and quantitative data derived from essay scores, self-regulated learning, and self-efficacy questionnaires. Interview data were transcribed and coded using content analysis. An open-coding process (e.g., Bazeley, 2013) was employed in which the second author read the transcripts and created categories according to the participants' answers and then discussed them with the rest of the authors. Then, the first author and a research assistant individually coded the transcriptions of 40 cases and reached a Cohen's Kappa of .93. The coding scheme is presented in Table 1.

Table 1

Coding scheme of student-reported reasons for self-assessment inaccuracy after receiving the instructor's score.

Coding categories	Explanation
Different Assessment Criteria Use	The participant and the instructor assessed the task using different evaluation criteria.
Error Identification	The participant and the instructor used the same evaluation criteria but identified different errors.
Differences in Scoring	The participant and the instructor used the same assessment criteria and identified the same errors but gave different scores.
Pessimistic Self-Perception	The participant relied on the pessimistic perception of their performance, often reported together with their critical nature.
Optimistic Self-perception	The participant self-scored higher than the real performance due to their overconfidence, and positive illusions about their performance.
Expert's Additional Resources	The participant lacked additional grading references, whereas the expert possessed all students' essays and additional resources, enabling them to make quality comparisons.
Lack of Knowledge	The participant lacked the content knowledge needed to self-score objectively.
Unexplained Discrepancy	The participant did not know or did not explain the reason for the difference.
No discrepancy	There was no difference in the scores, so the issue was not addressed.

Self-assessment accuracy was calculated from the difference between the student's score and the instructor's score (Schraw, 2009). Negative self-assessment accuracy scores indicate underestimation, while positive scores indicate overestimation.

Missing value analysis was conducted for the EMSR-Q data, as it is considered a valid procedure when the missing value proportion is small (i.e., 10%) (Little & Rubin, 2002). In our study, a small amount of missing values (1.79% in item 3, .89% in item 5,

item 12, and item 16) were identified and replaced using the expectation-maximization analysis.

To explore the effects of individual differences on the self-assessment accuracy, we created different groups based on the individual variables. These groups were incorporated as factors in our analysis to examine how these characteristics influence self-assessment accuracy, and the reasons participants reported. Regarding academic performance levels, participants were assigned to three groups according to their scores: low achievers (1-5.9), middle achievers (6-7.9) and high achievers (8-10). This scale was based on the common grading system implemented by the teachers in the country. For self-assessment accuracy, participants were assigned to three groups according to their score discrepancy: understimators (below -1.50), accurate (from -1.4 to +1.4), and overestimators (above +1.50).

Regarding self-efficacy, results indicate that participants reported moderate levels of self-efficacy in writing and assessing their work (M = 3.94, SD = .8). To facilitate group comparisons, we categorized participants into low self-efficacy group (scores ≤ 3.94, n = 50) and high self-efficacy group (>3.94, n = 62), based on the sample mean. Students in the high self-efficacy group perceived themselves to be skilled and confident in managing various aspects of academic writing, including coherence, cohesion, and adherence to APA standards. They also felt capable of effectively self-assessing their academic work.

Regarding self-regulation as measured by the EMSR-Q, participants primarily used learning-oriented strategies (M=31.66, SD= 7.02), which are beneficial for academic success. A total of 59 were classified as having high learning-oriented self-regulation (≥32), indicating the frequent use of emotional and motivational strategies that positively support learning (e.g., positive self-regulation of motivation, performance-oriented strategies, among others). Avoidance-oriented self-regulation was also present, with

moderate levels (M= 21.71, SD= 8.78), but with substantial variability in students' tendencies to engage in avoidance behaviors. Students with high avoidance self-regulation (≥ 23, n = 53) reflected a tendency to engage in behaviors and self-messages associated with avoidance of academic tasks or lack of regulation.

In RQ1, two data analysis methods were used: contrast of means, and correlational analysis between student's and instructor's score. Univariate ANOVAs were conducted for RQ2, and for significant main effects, we used post-hoc testing with Bonferroni when the independent variables had more than two levels (i.e., essay score and year level). Descriptive statistics were performed to answer RQ3 and RQ4, and the differences between groups were examined using Fisher's exact test. We used SPSS 24 for data analysis.

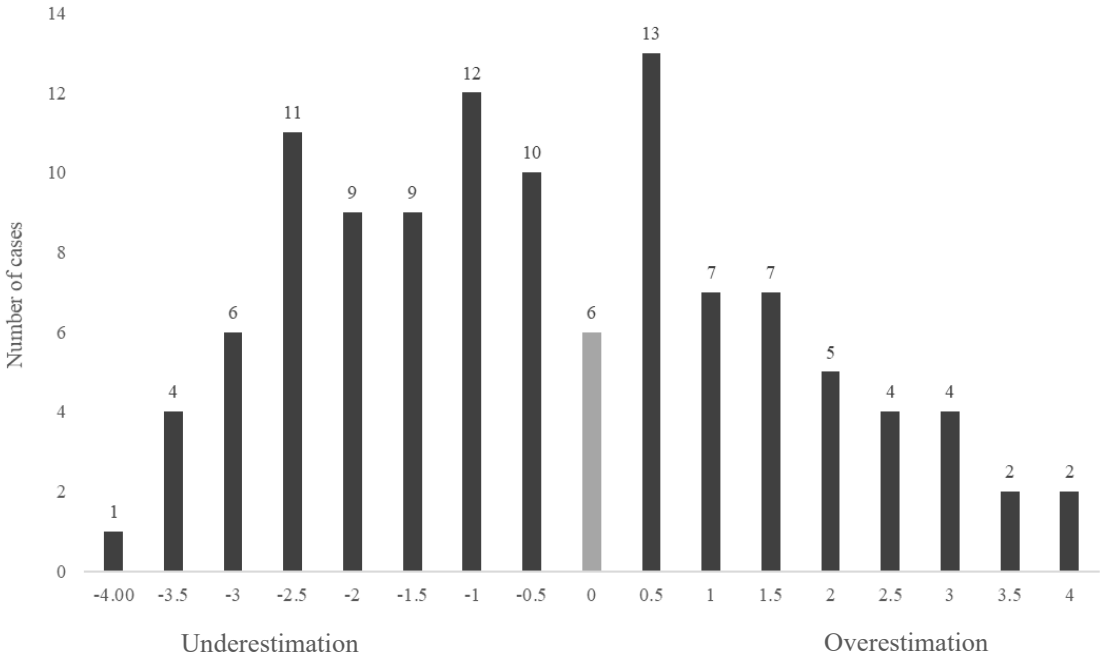
Results

RQ1. How Accurate Are Higher Education Students in Their Self-Assessment?

First, we conducted a correlation analysis to calculate the overlap between the students' and teacher's scores, finding out that they were significantly yet weakly and positively correlated (r=.246, $p < .05$). Then, we run a paired-samples t-test to find out if the instructor's scores and students' were significantly different, which they were (t_{111}=-*1.918,* $p \leq .05$, d= -.18). Finally, we explored the directionality of the deviation between students' (M=6.67, SD=1.16) and instructor's scores (M=7.01, SD=1.82). Thus, the average gap between self- and instructor's score was -0.345 (SD=1.90), suggesting a main tendency towards underestimation. However, when we examined the accuracy concerning both directionality and distance, it was found that 48 out of 112 participants were fairly accurate in their self-scoring, with score discrepancies falling within the range of -1 to +1 (Figure 2).

Figure 2

Frequencies for the score difference between student's and instructor's scores.



RQ2. Did Individual Differences Influence Accuracy (i.e., Year Level, Essay Score, Self-Efficacy, and Self-regulation)?

For the essay score, we found that it had a significant effect on self-assessment accuracy, F (2, 109) = [49.81], p = .00, partial η2 =.478. There were significant differences between groups ($p \leq .001$), with low achievers tending to assign higher scores than the instructor (M= 1.17, SD= 1.78), while middle (M= -.77, SD= 1.08) and high-performance group (M= -2.02, SD= 1.02) showed a tendency towards underestimation. Moreover, learning oriented self-regulation had a significant effect on self-assessment accuracy, F (1, 110) = [4.00], p = .048, partial η2 =.035, with high learning oriented self-regulation group presenting the more tendency towards underestimation (M= -.72, SD= 1.86), in comparison with the low learning oriented group (M= -.00, SD=1.89).

There were no significant results with regard to year level, F (2, 109) = [.018], p = .98, partial η2 =.000; self-efficacy, F (1, 110) = [2.10], p = .15, partial η2 =.019; and avoidance oriented self-regulation, F (1, 110) = [1.58], p = .211, partial η2 =.014.

RQ3. What Were the Reasons University Students Provided for the Discrepancy Between Their Score and Instructor's Score?

The participants reported a variety of reasons for the score discrepancy (Table 2). The main reason was related to their *pessimistic self-perception*, as participants expressed a negative view of their performance, leading them to assign lower scores than their actual achievement.

Table 2.

Reasons for explaining the score discrepancy between the student's and the instructor's scores.

	N	Pessimistic Self-Perception	Differences in Scoring	Error Identification	Different Assessment Criteria Use	Expert's Additional Resources	Lack of knowledge	Optimistic Self-perception	Unexplained discrepancy	No discrepancy
Total	112	44	37	29	23	13	12	11	7	5
Self-assessment accuracy										
Underestimator	40	31	25	5	4	5	5	0	2	0
Accurate	48	13	9	15	12	5	2	3	4	5
Overestimator	24	0	3	9	7	3	5	8	1	0
p		**0.000*****	**0.000*****	**0.034***	0.098	1.00	0.086	**0.000*****	0.784	**0.037***
Cramer's V		0.62	0.47	0.23	0.19	0.03	0.20	0.42	0.07	0.25
Feedback condition										
Rubric	38	22	18	7	4	7	4	1	1	1
Feedback	36	10	12	11	11	1	3	7	2	2
Combined	38	12	7	11	8	5	5	3	4	2
p		**0.018***	**0.025***	0.452	0.104	0.101	0.927	**0.050***	0.396	0.867
Cramer's V		0.27	0.25	0.12	0.20	0.20	0.06	0.23	0.13	0.06

Note. ***indicates<.001, **indicates p<.01, *indicates, p<.05

Next, we will analyse the interview answers comparing different accuracy groups. The majority of participants who underestimated their performance referred to the different approach of *scoring* between student and instructor (62.5% of underestimators referred to this). This is, they perceived that although they used the same criteria and identified the same errors as the instructor's, one of the judgements (i.e., student or instructor) was not

calibrated. Indeed, they also were more likely to attribute the score discrepancy to their *pessimistic self-perception* (77.5%) about their performance due to their perfectionism or to avoid the experience of receiving a lower score than expected, as the next extracts illustrate:

> I am very critical, and I do not want to overestimate myself. (…) And I do not like to praise myself, no, then you get bad surprises. (...) I prefer the teacher's scoring to be higher than my self-assessment. *Ana – 1st year level, and a score discrepancy of -2 points.*
>
> I am very perfectionistic and critical of everything. So maybe I have been really tough on the essay, but I do not know, I do not really think it is a 10; that seems like a high score to me. *Laura – 2nd year level, and a score discrepancy of -2.50 points.*

Conversely, overestimator cases referred significantly more to the *optimistic self-perception* (33.3%) as they explained that they had positive illusions about their performance due to their overconfidence or high self-esteem. Next, we present some extracts: “We tend to assess ourselves more positively whereas an external instructor can provide a more objective evaluation of the essay (…)” (Marcos*, 3rd year level, and a score discrepancy of 3.50 points)*; “Because I have high self-esteem, and in the end, the expert approaches it more objectively. I guess I evaluate myself subjectively. (…)” *(Ane, 3rd year level, and a score discrepancy of 2 points)*.

Lastly, accurate self-assessors were more likely to mention the *error identification* (31.3%). Even though they felt that they used the same assessment criteria as the instructor, they reported that the small discrepancy could be due to the identification of different errors. Specifically, these participants explained that they overlooked minor errors such as grammar, word repetition, and use of long sentences.

Notably, there were differences in participants' responses based on the feedback condition (See Table 2). Participants in the rubric condition reported more frequently *pessimistic self-perception* and *differences in scoring*. This suggests that providing a rubric before receiving their self-assessment accuracy score may have helped students identify the assessment criteria and recognise discrepancies in their scoring. In contrast, participants who received written feedback were significantly more likely to attribute their score discrepancies to an *optimistic self-perception*. However, these results should be interpreted with caution, given the weak relationship between these categories and the feedback condition, as found in the effect size results.

RQ4. Did Individual Differences Influence the Reasons Reported (i.e., Year Level, Essay Score, Self-Efficacy, And Self-regulation)?

In terms of differences by year level (Table 3), no significant differences were found. Regarding the essay score, low achievers tended to attribute the score difference to *error identification* (36.4%). This category refers to the discrepancy where the student and instructor used the same assessment criteria but identified different errors. For example, several cases pointed out the mistakes that they did not identify: "I did not realize that I repeated the same words frequently in the text." *(Tamara, 1st year level, and a score discrepancy of 0.50 point);* "There are long sentences, and I might have missed a punctuation mark here and there (…)" (*Rosa, 2nd year level, and a score discrepancy of -1 point*).

On the contrary, higher achievers were more likely to explain the discrepancy based on the *different scoring* system used by the student and instructor (60%), and this was often combined with their *pessimistic self-perception* (66.7%). As it can be seen in the following extracts, the words "self-demanding", or "not objective" were frequently used in their answers: "I'm quite pessimist with my own work (…)" (*Olivia, 1st year university, and a*

score discrepancy of -2.50 points); "I'm really demanding of myself and I cannot revise it objectively" *(Eva, 2nd year, and a score discrepancy of -1.50 points)*; "I think that it is because the self-demanding nature I always have. (…)" *(Carlota, 3rd year university, and a score discrepancy of -2.50 points).* However, lower achievers had the oppositive view and were more likely to use the self-score as a way to reinforce their ego and *optimistic self-perception* (20.5%): "when you do a task, you will always self-score higher than the instructor because you have done it, and you believe that your work is great (…)" *(Bea, 1st year level, and a score discrepancy of 4 points);* "I thought I had done better (…). I had a better opinion of myself than what it really is" *(Judith, 3rd year level, a score discrepancy of 3 points).*

Table 3.

Reasons for explaining the score discrepancy by individual differences and feedback conditions.

	N	Pessimistic Self-perception	Differences in Scoring	Error identification	Different Assessment Criteria Use	Expert's Additional Resources	Lack of knowledge	Optimistic Self-perception	Unexplained discrepancy	No discrepancy
Total	112	44	37	29	23	13	12	11	7	5
Year level										
1st year	39	15	15	12	9	5	4	4	0	1
2nd year	38	17	13	9	5	7	3	2	3	3
3rd year	35	12	9	8	9	1	5	5	4	1
p		0.673	0.502	0.739	0.394	0.099	0.692	0.450	0.081	0.527
Cramer's V		0.08	0.11	0.08	0.13	0.19	0.08	0.12	0.19	0.11
Essay score										
Low achievers	44	5	6	16	13	6	7	9	2	2
Middle achievers	38	19	13	10	6	4	2	0	3	3
High achievers	30	20	18	3	4	3	3	2	2	0
p		**0.000*****	**0.000*****	**0.032***	0.192	0.870	0.342	**0.005****	0.888	0.369
Cramer's V		0.47	0.39	0.24	0.18	0.05	0.14	0.30	0.06	0.14
Self-efficacy										
Low SE	50	17	14	10	13	6	6	7	4	4
High SE	62	27	23	19	10	7	6	4	3	1
p		0.335	0.322	0.278	0.242	1.00	0.764	0.214	0.698	0.170
Cramer's V		0.09	0.09	0.12	0.12	0.01	0.03	0.12	0.06	0.15
Learning oriented self-regulation										
Low LSR	59	17	15	17	17	5	5	6	4	2
High LSR	53	27	22	12	6	8	7	5	3	3
p		**0.021***	0.107	0.521	**0.034***	0.378	0.544	1.00	1.00	0.666
Cramer's V		0.22	0.17	0.07	0.21	0.10	0.07	0.01	0.02	0.05
Avoidance self-regulation										
Low AVSR	59	20	19	16	13	6	4	5	5	2
High AVSR	53	24	18	13	10	7	8	6	2	3
p		0.248	1.00	0.831	0.816	0.769	0.222	0.754	0.443	0.666
Cramer's V		0.11	0.01	0.03	0.03	0.04	0.13	0.04	0.09	0.05

Note. ***indicates<.001, **indicates p<.01, *indicates, p<.05

Regarding learning oriented self-regulation, two significant differences were found. First, highly learning oriented self-regulated students reported significantly more *pessimistic self-perception* (50.9%), often underestimating it due to their high self-standards and lack of confidence. Notably, even after receiving high instructor scores, some of these participants disagreed with the feedback, and mentioned the areas for further improvement, as evidenced by the following extracts:

> I do not see my text as good enough for a 9. (…) Maybe I do not have that much confidence in myself to say that I can do something good. *Carlos - 1st year level, a score discrepancy of -2.00 points.*
>
> I am quite critical of my essays (…). I do not think my self-assessment was incorrect because, as I perceived the text, it was not well-connected and had some poorly constructed sentences, so I think a 9 is a somewhat excessive score. *David - 3rd year level, a score discrepancy of -2.00 points.*

Conversely, low learning oriented self-regulated participants attributed more the discrepancy to the *different assessment criteria use* (29.3%) such as the use of non-specific content criteria or internal criteria (e.g., effort), as the following extracts illustrate: "I only focused on the structure, and not so much on if the content was aligned (…)" (*Clara– 1st year level, score discrepancy of 1 points*); "I self-scored based on my previous experiences. (…)" (*Lucía – 2nd year level, score discrepancy of 3 points*).

Lastly, in terms of self-efficacy levels and avoidance self-regulation, there were no remarkable differences among the groups.

Discussion

Our aim was to investigate the accuracy of self-assessment among higher education students. This included exploring the students' reasons for their inaccuracies and comparing them by individual differences.

Self-Assessment Accuracy (RQ1 & RQ2)

In RQ1 and RQ2 we explored the self-assessment accuracy. The correlation of students' self-scores and instructor's scores was weak and positive, and, regarding the direction of the deviation, students assigned a lower score to their essay than the instructor. Specifically, most participants were fairly accurate in their self-assessment, with 48 out of 112 showing a score discrepancy within the range of -1 to +1. A smaller group tended to underestimate their performance (40 out of 112). This direction towards accuracy and underestimation found in our study is contrary to previous empirical evidence claiming that students tend to overestimate their performance (Falchikov & Boud, 1989; León et al., 2023). There could be several explanations for this result. First, as suggested by Brown and Harris (2013), older or more experienced students tend to assign lower or less optimistic scores. Given that our participants have been part of the educational system for many years, they have likely had frequent opportunities to engage in self-assessment (Pinedo et al., 2023). This extended experience may have helped them develop competencies that lead to a more modest or accurate self-assessment. In this line, Boud et al. (2013) argued that the improvement in judgement requires opportunities to practice self-assessment over semesters, and receive feedback from an experienced marker, among others. Second, our results could also be explained by the low-stakes context of our study, where self-assessment did not have any consequences for the participants' scores. The lack of high-stakes consequences likely reduced the incentive to inflate their self-assessment. This aligns with research suggesting that self-assessment should be used in formative contexts, without counting towards the final grade, to promote accuracy (Andrade, 2019; Brown & Harris, 2013; Brown et al., 2015). Future studies should examine whether engaging in formative low-stakes self-assessment, where students receive ongoing feedback and

opportunities for self-reflection, leads to long-term improvements in self-assessment accuracy.

Regarding the effect of individual differences in self-assessment accuracy, we found that the essay score had a significant effect. Specifically, higher achievers showed a tendency towards underestimation while low achievers overestimated their performance. This finding goes in line with several studies supporting the Dunning-Kruger effect, in which students with lower ability or expertise overestimate their performance, whereas those with higher academic performance are more likely to underestimate (Boud et al., 2013; Boud & Falchikov, 1989; Dunning, 2011; Dunlosky & Rawson, 2012; Knof et al., 2024; Kruger & Dunning, 1999; Nepal et al., 2020). These discrepancies may also reflect differences in metacognitive awareness and knowledge of one's own skills. Lower achievers tend to be overconfident about their performance due to their lack of metacognitive skills (Panadero et al., 2016) and their limited level of proficiency (Dunning et al., 2004). Moreover, they tended to exhibit poor self-judging skills and demonstrate minimal to modest improvements in self-assessment over time (Boud et al., 2015). Indeed, they may rely on motivational strategies, self-protection, or optimistic self-presentation when judging their performance (Dunlosky & Rawson, 2012), while higher achievers may underestimate their abilities to appear humble or to protect themselves in case their performance falls short of expectations. Further research should explore the cognitive and metacognitive processes of students with different levels of performance to determine what self-assessment processes lead students to exhibit accurate judgement. By identifying the most efficient processes, teachers can design interventions to help students acquire the metacognitive skills needed for calibrated self-judgment.

Additionally, there was a significant but small effect of learning oriented self-regulation on self-assessment accuracy, with the high group presenting an underestimation pattern. Although, to our knowledge, there are no studies exploring this, extensive evidence supports the positive relationship between self-regulated learning and academic performance (Broadbent & Poon, 2015). Given this association, it is possible that students with greater self-regulatory skills, who often are academically successful, may underestimate their performance due to their focus on continuous improvement and learning. However, as the effect was small, future studies should examine the relationship between self-assessment accuracy and self-regulation to determine the effective self-regulatory strategies for accurate self-assessment.

Reasons for Self-Assessment (In)accuracy (RQ3 & RQ4)

Most participants attributed the self-assessment inaccuracy to their *pessimistic self-perception*, and at a lower frequency to the *differences in scoring* strategy. This suggests that even when students used the same assessment criteria as the instructor and identified the same errors, often their *pessimistic self-perception* led them to assign different scores to their work. In this line, there is evidence that students often feel unable to be objective in observing their own knowledge (Nieminen & Tuohilampi, 2020). This was especially the case for higher achievers and underestimators who scored the essays differently than the instructor due to their negative and more critical perception of themselves. This tendency might be explained by their greater metacognitive awareness, which makes them more sensitive to their own limitations and more cautious when judging their performance (Brown & Harris, 2013; Panadero et al., 2016). In addition, high-achieving students may apply more demanding internal standards, which can intensify self-criticism and lead to underestimation (Andrade, 2019).

On the contrary, lower achievers, who tended to overestimate, may have attributed their inaccuracy to an *optimistic self-view* that prevents them from recognizing their actual level of competence. This aligns with Dunning et al. (2004), who argued that individuals with limited domain knowledge are "doubly handicapped": they not only perform poorly but also lack the metacognitive insight needed to recognize their errors. In the absence of such diagnostic insight, optimism and reliance on subjective impressions, such as effort or perceived potential, replace objective evaluation (Evans et al., 2002). These findings suggest that self-assessment is not only a cognitive monitoring process dependent on knowledge and evaluation skills but also an affective act shaped by self-perception biases. Accuracy in self-evaluation, therefore, may be shaped by the interaction between what students know about the task and how they feel about their own competence.

Differences also emerged in relation to learning-oriented self-regulation. Participants with high learning oriented self-regulation were more likely to explain their underestimation based on their *pessimistic self-perception*. Such underestimation may reflect defensive pessimism (Butler, 2011), which can carry an affective cost, as these students report lower satisfaction with their performance compared to overestimators (Narciss et al., 2011). Conversely, low learning oriented self-regulated participants attributed their score discrepancy to the *use of different assessment criteria* compared to the instructor, although this relationship was weak. These students may engage less frequently in motivational strategies aimed at achieving learning goals (Alonso-Tapia et al., 2014), which could hinder their ability to accurately calibrate self-assessment. Therefore, there is a need to provide clear and task-specific criteria (Andrade & Valtcheva, 2009; Brown et al., 2015; Butler, 2018), as well as a pre-specified marking guides that help students evaluate essays consistently. An efficient instructional aid

could be a rubric, as it clarifies the assessment criteria, offers quality levels to compare the work with (Andrade, 2001) and enhances self-assessment accuracy (Krebs et al., 2022).

Importantly, differences also emerged across feedback conditions, although these interpretations should be made cautiously given the small effect sizes (Cramer's V ≤ .20). Participants in the rubric condition reported more frequent pessimistic self-perceptions, that is, reported to assign themselves lower scores compared to those in other conditions. However, they also reported that they identified the same errors and criteria as the instructor's grading - that is, they did not perceive themselves as being less accurate in content terms (see Panadero et al., 2016). This suggests that the difference lies not in evaluative accuracy but in the interpretive framing of their performance in terms of the score. The rubric, by providing explicit performance standards, may have made students more aware of subtle gaps between their work and the highest achievement levels, activating a more self-critical stance. In other words, the rubric might have enhanced metacognitive accuracy while simultaneously reducing affective confidence through greater standard awareness. Conversely, participants in the instructor's written feedback condition reported more optimistic self-perceptions than those in the rubric and combined conditions. A plausible explanation is that narrative comments, without the normative precision of a rubric, left more room for subjective interpretation and self-protective attributions. Students may have framed discrepancies as temporary or externally caused ("I didn't explain myself well") rather than structural ("my performance did not meet this level"). Moreover, the personalised tone of instructor feedback could have increased perceived support, promoting a more benevolent self-view even when recognising errors. Taken together, these findings illustrate how different feedback formats may differentially engage cognitive and

affective components of self-assessment: rubrics sharpen evaluative awareness but can elicit more self-critical appraisals, whereas personalised comments foster a sense of encouragement that sustains optimism but may dilute precision in self-judgement.

Limitations and Future Lines of Research

The first limitation in our study is the type of data, derived from interviews and questionnaires, which could be affected by the limitations of self-report. However, as our focus was to understand the reasons behind students' inaccuracies in their self-scores, the interview data provided valuable insights into their perceptions. Second, our participants were mainly female, which limits the generalizability of our findings. However, this reflects the typical gender distribution in Psychology programs in Spain, where women tend to be the majority. Third, our participants lacked standardized scoring aids (i.e. rubric) before self-assessment, and this potentially could lead to individual biases. Future studies could address this limitation by providing rubrics with the scoring criteria before the self-assessment, or implementing a two-stage self-assessment where participants first score without aids, receive feedback, and then revise and re-score their essays. These approaches align with evidence from Jönsson et al. (2025), who identify the characteristics of successful rubric based interventions that positively influence student learning. Lastly, the study only measured scoring accuracy in one task, but it could be interesting to explore if students improve their judgement in similar tasks after receiving the instructor's score or feedback.

Conclusion

This study shows that self-assessment accuracy is not only about comparing students' and instructors' scores, but also about how students make sense of the discrepancies between them. By prompting students to reflect explicitly on why their self-assessments did not always align with external evaluations, we gained unique insight into

the cognitive and metacognitive processes that shape their judgments. The findings reveal a consistent pattern: high achievers tend to underestimate their performance, while lower achievers are more likely to overestimate. Importantly, students were aware of these tendencies, attributing inaccuracies either to internal factors (e.g., self-critical attitudes, lack of knowledge) or to external factors (e.g., assessment criteria, expert resources).

These results underline that self-assessment inaccuracies are not random errors, but reflections of broader self-perception patterns linked to performance and self-regulation. Recognizing these biases is a first step toward designing interventions that help students calibrate their judgments more accurately. Future research should continue to explore how emotions, self-efficacy, and other individual differences influence these processes, ideally through longitudinal approaches that capture how self-assessment accuracy develops over time.

References

Alonso-Tapia, J., Panadero, E., & Ruiz, M. A. D. (2014). Development and validity of the Emotion and Motivation Self-regulation Questionnaire (EMSR-Q). *The Spanish Journal of Psychology*, *17*, E55. https://doi.org/10.1017/sjp.2014.41

Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education, 4*(4).

Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00087

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into practice, 4*8(1), 12-19. https://doi.org/10.1080/00405840802577544

Barnett, J. E., & Hixon, J. E. (1997). Effects of Grade Level and Subject on Student

Test Score Predictions. *Journal of Educational Research, 90(*3), 170-174. https://doi.org/10.1080/00220671.1997.10543773

Bazeley, P. (2013). *Qualitative data analysis. Practical strategies*. Sage.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time*?*. *Assessment & Evaluation in Higher Education, 38*(8), 941-956. https://doi.org/10.1080/02602938.2013.769198

Boud, D., Lawson, R., & Thompson, D. G. (2015). The calibration of student judgement through self-assessment: disruptive effects of assessment patterns. *Higher education research & development, 34*(1), 45-59. https://doi.org/10.1080/07294360.2014.934328

Bradshaw, B. K. (2001). Do students effectively monitor their comprehension*? Reading Horizons, 41*(3), 143–154

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The internet and higher education*, *27*, 1-13. https://doi.org/10.1016/j.iheduc.2015.04.007

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy and Practice*, *22*(4), 444–457. https://doi.org/10.1080/0969594X.2014.996523

Brown, G., & Harris, L. R. (2013). Student self-assessment. SAGE handbook of research on classroom assessment. In J. H. McMillan (Ed.), *The SAGE handbook*

of research on classroom assessment (pp. 367–393). Sage.

Butler, R. (2011). Are positive illusions about academic competence always adaptive, under all circumstances: New results and future directions. *International Journal of Educational Research*, *50*(4), 251-256. https://doi.org/10.1016/j.ijer.2011.08.006

Butler, Y. G. (2018). Young learners' processes and rationales for responding to self-assessment items: Cases for generic can-do and five-point Likert-type formats. In J. Davis, J. Norris, M. Malone, T. McKay, & Y. Son (Eds.), *Useful assessment and evaluation in language educatio*n (pp. 21–39). Georgetown University Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed). Routledge.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://psycnet.apa.org/doi/10.1016/j.learninstruc.2011.08.003

Dunning, D. (2011). The dunning-kruger effect: On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology, 44*, 247–296. https://doi.org/10.1016/B978-0-12- 385522-0.00005-6

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Elder, A. D. (2010). Children's self-assessment of their school work in elementary school.

Education, 38(1), 5–11.

Evans, A. W., Leeson, R. M. A., & Newton-John, T. R. O. (2002). The influence of self-deception and impression management on surgeons' self-assessment scores. *Medical education, 36*(11), 1095-1095. https://doi.org/10.1046/j.1365-2923.2002.134612.x

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of educational research, 59*(4), 395-430.

Jönsson, A., Panadero, E., Pinedo, L., & Fernández-Castilla, B. (2025). Using rubrics for formative purposes: identifying factors that may affect the success of rubric implementations. *Assessment in Education: Principles, Policy & Practice, 32*(2), 192–211. https://doi.org/10.1080/0969594X.2025.2486947

Kaderavek, J. N., Gillam, R. B., Ukrainetz, T. A., Justice, L. M., & Eisenberg, S. N. (2004). School- age children's self-assessment of oral narrative production. *Communication Disorders Quarterly, 26*(1), 37–48. https://doi.org/10.1177/15257401040260010401

Knof, H., Berndt, M., & Shiozawa, T. (2024). Prevalence of Dunning-Kruger effect in first semester medical students: a correlational study of self-assessment and actual academic performance. *BMC Medical Education, 24*(1), 1210. https://doi.org/10.1186/s12909-024-06121-7

Kostons, D., Van Gog, T., & Paas, F. (2009). How do I do? Investigating effects of expertise and performance-process records on self-assessment. *Applied Cognitive Psychology, 23,* 1256–1265. https://doi.org/10.1002/acp.1528

Kostons, D., Van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner-controlled instruction: Differences between effective and ineffective learners. *Computers and Education, 54*, 932e940. https://doi.org/10.1016/j.compedu.2009.09.025

Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning & Instruction, 22*(2), 121–132. https://doi.org/10.1016/j.learninstruc.2011.08.004

Krebs, R., Rothstein, B., & Roelle, J. (2022). Rubrics enhance accuracy and reduce cognitive load in self-assessment. *Metacognition and Learning, 17*(2), 627-650. https://doi.org/10.1007/s11409-022-09302-1

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology, 77*(6), 1121. https://doi.org/10.1037/0022-3514.77.6.1121

León, S. P., Panadero, E., & García-Martínez, I. (2023). How Accurate Are Our Students? A Meta-analytic Systematic Review on Self-assessment Scoring Accuracy. *Educational Psychology Review, 35*(4), 106. https://doi.org/10.1007/s10648-023-09819-0

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. https://doi.org/10.1080/02602930802687737

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). John Wiley & Sons.

Narciss, S., Koerndle, H., & Dresel, M. (2011). Self-evaluation accuracy and satisfaction with performance: Are there affective costs or benefits of positive self-evaluation bias?. *International Journal of Educational Research, 50*(4), 230-240. https://doi.org/10.1016/j.ijer.2011.08.004

Nepal, K., Sharma, R., Ghimire, K., & Thapa, M. (2020). Self-assessment behavior in mathematics courses: A function of their performance on academic assignments. *Journal of Excellence in College Teaching, 31*(3), 29–49. https://celt.miamioh.edu/index.php/JECT/article/view/229

Ng, J. R., & Earl, J. K. (2008). Accuracy in self-assessment: The role of ability, feedback, self-efficacy and goal orientation. *Australian Journal of Career Development, 17*(3), 39-50. https://doi.org/10.1177/103841620801700307

Nieminen, J. H., Asikainen, H., & Rämö, J. (2021). Promoting deep approach to learning and self-efficacy by changing the purpose of self-assessment: A comparison of summative and formative models. *Studies in Higher Education, 46*(7), 1296-1311. https://doi.org/10.1080/03075079.2019.1688282

Nieminen, J. H., & Tuohilampi, L. (2020). 'Finally studying for myself'–examining student agency in summative and formative self-assessment models. *Assessment and Evaluation in Higher Education, 45*(7), 1031–1045. https://doi.org/10.1080/02602938.2020.1720595

Panadero, E., Alonso-Tapia, J., García-Pérez, D., Fraile, J., Sánchez Galán, J. M., & Pardo, R. (2021). Deep learning self-regulation strategies: Validation of a situational model and its questionnaire. *Revista de Psicodidactica, 26*(1), 10–19. https://doi.org/10.1016/j.psicoe.2020.11.003

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational psychology review, 28*, 803-830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E., Fernández, J., Pinedo, L., Sánchez, I., & García-Pérez, D. (2024). A self-feedback model (SEFEMO): Secondary and higher education students' self-assessment profiles. *Assessment in Education Principles, Policy & Practice*, 1–33. https://doi.org/10.1080/0969594X.2024.2367027

Panadero, E., García-Pérez, D., Fernández-Ruiz, J., Fraile Ruiz, J., Sánchez-Iglesias, I., & Brown, G.T. L. (2023). University students' strategies and criteria during self-assessment: Feedback and year level effects. *European Journal of Psychology of Education, 38,* 1031–1051. https://doi.org/10.1007/s10212-022-00639-4

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review, 22,* 74-98. https://doi.org/10.1016/j.edurev.2017.08.004

Pinedo, L., Panadero, E., Fernández-Ruiz, J., & Rodríguez-Hernández, C. (2023). Students' experiences in self-assessment: Training, processes and feedback use in secondary and higher education. *Assessment in Education Principles, Policy & Practice, 30*(5–6), 448–480. https://doi.org/10.1080/0969594X.2023.2284630

Rees, C., & Shepherd, M. (2005). Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviours. *Medical education, 39*(1), 30-39. https://doi.org/10.1111/j.1365-2929.2004.02030.x

Richardson, M., Bond, R., & Abraham, C. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138,* 353–387. https://doi.org/10.1037/a0026838

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Influences on student cognitions about evaluation. *Assessment in Education: Principles, Policy & Practice, 9*, 81–95. https://doi.org/10.1080/09695940220119201

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*(2), 147-164. https://doi.org/10.1080/02602930301671

Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology, 109*(8), 1049. https://psycnet.apa.org/doi/10.1037/edu0000190

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education, 28*(5), 549-565. https://doi.org/10.1080/02602930301678

Tejeiro, R. A., Gomez-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.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education,* 1-15. https://doi.org/10.1080/02602938.2019.1629390

Appendix A.

Rubric.

Category	Low quality	Average quality	High quality
Writing process[a]	I started writing the text without planning what I wanted to write. I have hardly reread what I was writing and, when I finished, I have not reviewed the text or I have only looked for misspellings.	2 options: a) Before writing, I have planned what I wanted to communicate. At the end, I have hardly reviewed the text or I have only looked for misspellings b) I started writing without thinking much about what I wanted to tell. However, I reviewed the text several times, looking for all or some of these factors: Text structure, coherence and connection between paragraphs, clarity of the message, style, and spelling	Before writing, I thoroughly planned what I wanted to tell and how I was going to do it. I reviewed while I was writing and, at the end, I also reviewed the full text at least once. While reviewing, I looked for all or some of these factors: Text structure, coherence and connection between paragraphs, clarity of the message, style, and spelling.
Text components: Structure and coherence/ connection between paragraphs	There is no clear structure, with an introduction, a crux, and a closing Lack of incorrect use of text connectors and/or discourse markers Regarding paragraphs, one of these two happens: a) The text has only one or two paragraphs, without clear internal and external coherence b) The text has many very short paragraphs, which makes it difficult to follow the argument line	A structure is somehow present (introduction, crux, and closure) but could be more clearly delimited Connectors are most of the times used appropriately. However, there may be one or more of these flaws: Same paragraph includes different unorganized ideas Same idea in two para- graphs when it could be in one The paragraph where the argument is developed is too long; it could be divided Connector/text markers are misused	There is a very clear structure in the text: including opening, argument crux and closing Ideas are connected and presented in well-organized paragraphs Connectors and/or discourse markers are effectively used

Text components: Sentences, vocabulary, and punctuation	Sentences are too long (over 40 words) or too short. Excessive use of text insertions within sentences. Punctuation is incorrect (e.g., lack of commas, they break the sentence) Too many colloquial expressions Abuse of passive or impersonal tenses	Most sentences are of adequate length, with a few too long or short or incomplete Punctuation is correct, although there may be a few mistakes The vocabulary is adequate, but different terms are used to refer to the central concept of the text Some colloquial expression may appear	The sentences are well constructed, usually following a simple structure, in an active language and a coherent use of the verbs Punctuation is correct The vocabulary is adequate, and the main terms are used with precision

[a] The "Writing Process" criterion was included only in the version of the rubric provided to students in the rubric condition group after their self-assessment. It was important due to the seminar content, but this criterion could not be evaluated by the instructor.

Appendix B.

Self-efficacy Questionnaire

WRITING OF ACADEMIC TEXTS SELF-EFFICACY QUESTIONNAIRE

Express how capable or competent you feel about the following statements, using the scale shown below:

0	1	2	3	4	5	6
Very strongly disagree	Strongly disagree	Rather disagree	Indifferent	Rather agree	Strongly agree	Very strongly agree

	1st Occasion	2nd Occasion	3rd Occasion
I usually feel capable to write academic texts			
When I write an academic text I know what steps to follow			
I can write different types of texts following specific communicative purposes			
When I write an academic text, I am able to apply the principles of coherence and cohesion between paragraphs			
I am able to write clear texts, considering elements such as the length of the sentences and the vocabulary used			
Normally, I am able to quote and make the list of references following APA standards			
I am able to self-assess my own tasks			
I am able to self-assess my own academic texts			