European Journal of Psychology of Education (2026) 41:67
https://doi.org/10.1007/s10212-026-01087-0

The affective cost of miscalibration: Effects of discovering self-assessment accuracy on emotions and self-efficacy

Leire Pinedo[1] · **Ernesto Panadero**[1,2,3] · **Javier Fernández-Ruiz**[4] · **Daniel García-Pérez**[5] · **Gavin T. L. Brown**[6]

Received: 27 March 2025 / Revised: 22 December 2025 / Accepted: 10 March 2026

Abstract
The alignment of self-assessment judgements to formal grading (i.e., accuracy) is an essential process by which students calibrate their thinking about work quality to disciplinary standards. However, extensive research shows that students are not well-calibrated in their self-assessments. Relatively little is known about how the realism of self-assessment affects positive or negative emotions and self-efficacy. In this study, we examined the relationships between self-assessment accuracy (over- or under-estimation) and students' emotions and self-efficacy. A total of 112 higher education students wrote an essay, self-assessed their performance, received feedback, self-assessed again, and subsequently received the tutor grades for their essays, creating a potential discrepancy between self-assessment and tutor assessment grades. Based on the sequence of events in the experiment, we used autoregressive path modeling to examine the repeated measures relationship of emotions and self-efficacy over time. Starting values for the two emotions and self-efficacy were strong predictors across time. The discrepancy between tutor and self-assessed grades was introduced after two rounds of self-reported emotions and self-efficacy and regressed onto the third round of those variables. Students who overestimated their performance experienced a decline in positive emotions, an increase in negative emotions, and a decrease in self-efficacy when confronted with their inaccuracies. In contrast, those who had underestimated their performance had a concomitant increase in positive emotions and self-efficacy, and a decrease in negative emotions. This study highlights the importance of fostering self-assessment accuracy for emotional well-being and self-efficacy. Future research should delve deeper into processes that support calibrated realism in student self-assessment.

Keywords Self-assessment · Self-assessment accuracy · Self-efficacy · Emotions · Higher education

Introduction

Extensive research in educational psychology and education has found that self-assessment has positive effects on students' academic performance, self- regulation and self-efficacy (e.g., Panadero et al., 2017; Yan et al., 2022). Student self-assessments tend to diverge considerably

Extended author information available on the last page of the article

from accepted disciplinary or professional standards, meaning that novices find realism in their self-assessments difficult (Boud & Falchikov, 1989). It is noteworthy that this problem is not restricted to students; it exists across many fields of knowledge (Dunning et al., 2004). Accuracy is essential in self-assessment, so students make appropriate decisions about strengths and areas for improvement, leading to appropriate adjustments to their learning strategies and efforts (León et al., 2023; Panadero et al., 2016a). Consistent with calls to examine self-assessment from new perspectives and with other constructs beyond score alignment (Andrade, 2019; Lui & Andrade, 2022; Panadero et al., 2016a), this study adds to our understanding of how accuracy in self-assessment relates to achievement emotions and self-reported self-efficacy by experimentally examining self-assessment practices in a longitudinal design.

Self-assessment definition and accuracy

Self-assessment "involves 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., 2016a, p. 804), thereby enabling students to correctly identify their strengths and weaknesses, and informing next steps in learning. It is considered a metacognitive activity, as students need to monitor, evaluate their own performance and interpret evidence about their progress (Zimmerman, 2008). These metacognitive processes enable students to generate informed evaluations rather than intuitive guesses. As such, the accuracy of self-assessment depends partly on the effectiveness of students' metacognitive monitoring and evaluation.

While research has established factors that lead to greater or lesser alignment of student self-assessments with instructors' expertise (Boud & Falchikov, 1989; Brown & Harris, 2013; León et al., 2023), the processes students use to self-assess have been less explored (e.g., Panadero et al., 2024; Yan & Brown, 2017). Because educational self-assessment usually happens at the request or instruction of teachers (Panadero et al., 2014), it is important to identify how those teacher-specified practices relate to self-regulated learning and emotional responses. Ideally, students whose self-assessments are well-calibrated with the judgments of their teachers will have greater self-efficacy and positive emotional responses in the process of self-assessment.

Some researchers have indicated that it is not important whether students can estimate properly their grades or not (Andrade, 2019). Other researchers have called for "content accuracy" which might be as important as scoring accuracy (Panadero et al., 2016a, 2016b). Others have advocated a student-centered learning approach for self-assessment in which students consider their work in light of their own learning interests rather than those of the curriculum or instructor (Bourke, 2017; Tan, 2008). Nevertheless, becoming a member of a discipline requires learning how to evaluate one's own work in accordance with norms and standards of that discipline; that means self-assessment can function as an approach to learning professional standards (Tai et al., 2018). Empirical research also shows that when students self-grade their work, their performance increased in subsequent testing against control groups (Sanchez et al., 2017). However, this improvement may depend on specific conditions,as León et al.'s (2023) meta-analysis showed that tendencies to over-estimate one's performance could be decreased if students received feedback, had more experience with self-assessment, had greater content knowledge, and embraced the formative purpose of self-assessment. Nevertheless, existing literature suggests that the consequences of discrepancies between student self-assessment and instructor evaluation

may depend on the direction of the miscalibration. Butler (2011) proposes that positive illusions about academic competence may be less problematic than negative ones; however, when overly optimistic self-assessments are confronted with lower-than-expected external evaluations, such discrepancies may generate emotional costs. This possibility remains underexplored empirically. Thus, understanding how under- and over-estimation of ability relate to self-efficacy and achievement emotions is of great interest.

It is important to acknowledge that instructor evaluations also entail subjective judgments, particularly in open-ended assessments such as essays. In this study, instructor scores are used as a reference benchmark, not as an unquestionable truth, in line with established quantitative approaches to examining self-assessment accuracy (e.g., Boud & Falchikov, 1989; Brown et al., 2015; Panadero et al., 2016a, b).

Impact of self-assessment accuracy on emotions

Achievement emotions, conceptualized by Pekrun's Control-Value Theory (CVT), are "emotions that are directly linked to achievement activities or achievement outcomes" (Pekrun et al., 2011, p. 37). According to this theory, achievement emotions arise from students' appraisals of control (i.e., their perceived ability to influence outcomes) and value (i.e., the importance they attribute to the task) (Pekrun et al., 2011). These emotions can be classified along three dimensions: valence (positive vs. negative), activation (activating vs. deactivating), object focus (activity vs. outcome) (Pekrun et al., 2023a). For instance, when students feel a high sense of control and value over a task, they are more likely to experience positive emotions such as pride (activating emotion), while negative emotions (e.g., activating anxiety or deactivating boredom) tend to decrease (Shao et al., 2020).

A crucial moment affecting emotions is the reception of performance feedback, as it shapes students' appraisals of control and value (Goetz et al., 2018; Pekrun et al., 2023b). Receiving a score can be perceived as a sign of success or failure, influencing emotional responses. Empirical research reports that higher scores or grades are positively correlated with positive emotions and negatively correlated with negative emotions (Pekrun et al., 2023b). Likewise, students who received lower grades than anticipated often experienced negative emotions (Jonsson & Panadero, 2018; Pitt & Norton, 2017) like disappointment, frustration, and panic (Lipnevich & Smith, 2009). In parallel, Peterson et al. (2015) found that changes in emotions had meaningful relations to test scores and GPA once students knew their scores, with higher-performing students having greater levels of positive emotions and lower levels of negative emotions, compared to lower-performers. Negative emotions can have detrimental effects on student's engagement with the task, leading to disengagement from feedback, and reduced motivation (Lim et al., 2021).

Impact of self-assessment accuracy on self-efficacy

Self-efficacy refers to an individual's belief in their ability to succeed in a specific task (Bandura, 1997). Self-efficacy has shown to be one of the main predictors for academic performance (Richardson et al., 2012). As a dynamic component of the self-regulated learning cycle, self-efficacy is based on judgements of past experiences of one's ability, and that can change during a learning task (Bernacki et al., 2015). Unfortunately, changes in self-efficacy over time in an assessment context have not been extensively studied (Honicke et al., 2023).

Nonetheless, when students accurately assess their performance, they become more aware of their progress, which can lead to an increased confidence and a stronger sense of self-efficacy (Andrade, 2019). A meta-analysis (Panadero et al., 2017) concluded this relationship has a strong positive effect ($d = 0.73$); self-assessment interventions increase self-efficacy. Indeed, when students accurately self-assess their task performance their self-efficacy improves because it fosters a sense of mastery (Bandura, 1997). Simultaneously, students with higher self-efficacy are more likely to engage in self-assessment (Yan, 2020) because they trust their ability to monitor and evaluate their own performance.

The bidirectional effect of self-efficacy and performance is well established, with positive performance leading to increased self-efficacy and vice versa (Honicke et al., 2023). However, the positive effect of performance on self-efficacy is likely to be significantly larger than the reverse effect of self-efficacy on performance (Talsma et al., 2018). For instance, Villafane et al. (2016) found that positive exam experiences increased students' self-efficacy, while poor performance led to its decline. Similarly, Honicke et al. (2023) suggested that students rely on previous performance to develop self-efficacy beliefs, with successful task performance helping to calibrate efficacy judgements. This aligns with Bandura's (1997) theory of self-efficacy, which identifies prior performance and learning experiences as primary sources of efficacy beliefs.

According to Bandura, we therefore hypothesize that students' self-efficacy may be associated with their realization of the relative accuracy of their self-assessments. Specifically, when students discover that their actual performance is higher than their self-estimation, their self-efficacy will increase as they have a clearer understanding of their performance. Conversely, when students discover a lower performance than their self-assessment, their self-efficacy may decline. This hypothesis aligns with evidence showing that receiving lower grades negatively affects self-efficacy (Jonsson & Panadero, 2018; Lipnevich & Smith, 2009).

However, evidence that these principles apply equally to self-assessment is lacking. Exploring this relationship will contribute to our understanding of how students' perceptions of their own abilities (i.e., self-assessments) are related to the external measures, such as instructors' scores. Additionally, this could reveal critical factors shaping self-efficacy, including the role of feedback and the alignment (or misalignment) between self-perception and actual performance.

Aim

Our aim was to examine changes in students' emotions and self-efficacy following the discovery of discrepancies in the accuracy of their self-assessment. The hypotheses (H) are as follows:

H1: Discrepancy between instructor's score and self-score will be associated with positive emotions. Overestimating performance will decrease positive emotions, while underestimating actual performance will increase them.
H2: Discrepancy between instructor's score and self-score will be associated with negative emotions. Overestimating performance will increase negative emotions, while underestimating actual performance will decrease them.
H3: Discrepancy between instructor's score and self-score will be associated with self-efficacy. Overestimating performance will decrease self-efficacy, while underestimating

actual performance will increase it, as per Bandura's (1997) assertions about the impact of performance accomplishments.

Method

This study employed an experimental, between-subjects design with random assignment to one of three treatment conditions. It was approved by the University Ethics Committee at Universidad Autónoma de Madrid (Reference: CEI-84–1557).

Sample

A convenience sample of 126 undergraduate psychology students was recruited from the university where the second author was employed at the time. Due to missing data, the final sample comprised 112 students (89% females) from first ($n=39$), second ($n=38$), and third year ($n=35$) levels. Participants were randomly assigned to one of the three feedback conditions: instructor's feedback ($n=36$), rubric ($n=38$), and a combined condition of instructor's feedback and rubric ($n=38$). Participants received credits for participation. The final sample size ($n=112$) was appropriate for the subsequent path analysis. Although moderate, this sample size provides sufficient statistical power for the autoregressive path model employed here (Curran & Bollen, 2001), especially given that the study's primary aim is not the detection of small-magnitude effects or applied prediction, but rather the examination of longitudinal change patterns. A detailed justification for the sample size's adequacy regarding the model complexity is presented in the Data Analysis section.

Note, in this study we do not consider the effect of type of feedback, just the changes in students' emotions and self-efficacy that occur when they discover the discrepancy between their self-estimate and instructor grades. Other studies from those data can be found at Panadero et al. (2022; 2024).

Data collection and instruments

The data included the instructor's essay score, students' self-assessment accuracy, and longitudinal data consisting of three dependent variables (i.e., positive emotions, negative emotions, and self-efficacy). The dependent variables were measured three times: (1) before performing the self-assessment but with students knowing they were about to self-assess, (2) right after the self-assessment, and (3) right after receiving the instructor's score.

Academic writing self-efficacy questionnaire

This ad-hoc instrument measures students' perceptions of self-efficacy when writing an academic text. It included eight items and used a 7-point rating scale ranging from "absolutely disagree" to "absolutely agree" (See Appendix B). The internal consistency for the three measurement occasions was: before self-assessment ($\alpha=0.78$), before feedback ($\alpha=0.80$), and after feedback ($\alpha=0.87$). Confirmatory factor analysis (CFA) was conducted to confirm the unidimensional structure of the instrument. Due to the violation of multivariate normality (Mardia Skewness = 181.21, $p<.001$; Mardia Kurtosis = 1.89,

$p = .059$), diagonally weighted least squares (DWLS) was used. All fit indices met the recommended thresholds except for the SRMR that was slight above the recommended value (Schreiber et al., 2006): $\chi^2/df = 1.60$, $p = .21$; RMSEA = 0.075, 90% CI [0.013, 0.121]; CFI = 0.981; TLI = 0.974; SRMR = 0.097. Additionally, the model showed positive item factor loadings ($p < .001$) with a range from 0.38 to 0.81.

Emotions questionnaire

This ad-hoc instrument was based on 9 basic academic emotions from Pekrun et al. (2002). Four positive emotions (i.e., joy, hope, pride, and relief) and five negative emotions (i.e., anxiety, embarrassment, sadness, boredom, and anger) were included. In addition, we included "nervousness" as a sixth negative emotion, considering that in our study context -where participants were recorded and observed-it could be especially relevant. Therefore, the final instrument comprised 10 emotions in total. The rating scale ranged from 1, very low intensity, to 10, maximum intensity. We also used the DWLS estimation method as the multivariate distribution was non-normal (Mardia Skewness = 647.16, $p < .001$; Mardia Kurtosis = 6.99, $p < .001$). A two-factor solution had acceptable to good fit ($\chi^2/df = 1.30$, $p = .25$; RMSEA = 0.051, 90% CI [0.000, 0.090]; CFI = 0.975; TLI = 0.967; SRMR = 0.095). The positive emotions factor had statistically significant ($p < .001$), positive item factor loadings, ranging from 0.67 to 0.81, and the negative emotions factor also had statistically significant ($p < .05$), positive item factor loadings, ranging from 0.25 to 0.83. The reliability indices were $\alpha = 0.93$ for positive emotions, and $\alpha = 0.90$ for negative emotions. Consistent with the logically inverse factor meanings, the correlation between the two factors was negative, but not statistically significant ($r = -0.214$, $p = .065$).

Instructor's score

The instructor's score was based on the students' performance in the essay that they wrote for the study, using a rubric designed for the study. The scores were between 0 and 10. Two people, the second author and the seminar instructor (4th author), independently assessed and scored one-third of the essays using an analytic rubric specifically designed for this study (See Appendix A). This produced high agreement between markers and strong alignment with the rubric categories (Intra-class Correlation .91). Subsequently, the seminar instructor scored the remaining essays.

Self-assessment accuracy

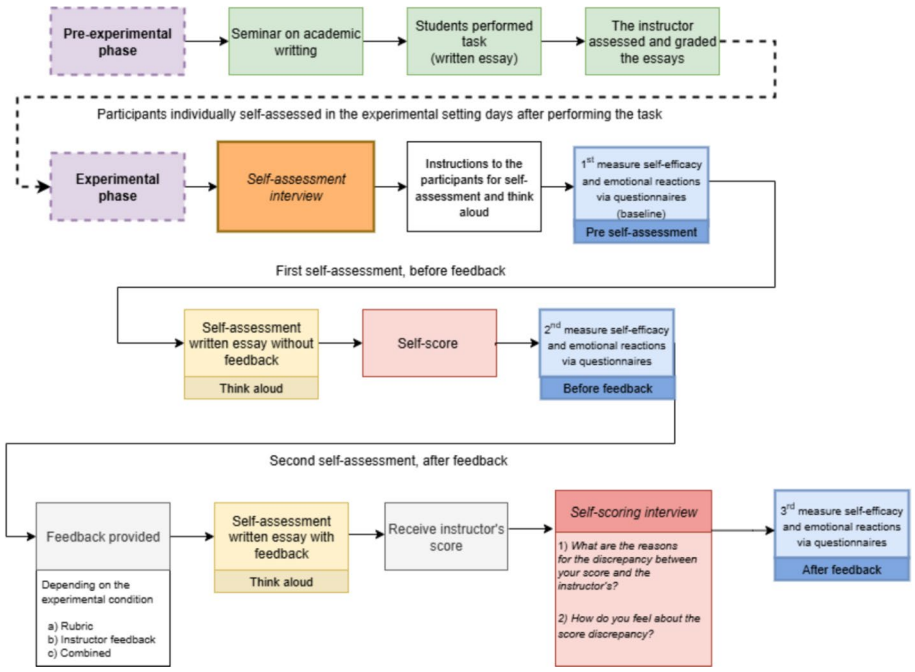
The self-assessment accuracy was calculated by computing the difference between the student's score and the instructor's score (Schraw, 2009). We considered not only the difference but also the directionality. A positive self-assessment accuracy score indicates that the student scored their own work higher than the instructor, whereas a negative score indicated an underestimated self-scoring. The accuracy range obtained was −3.8 to + 4.

Procedure

Participants attended a three-hour seminar on academic writing and wrote an essay. The essay was related to the students' psychology programme but did not have any implication for their academic record. After a few days, the participants individually

visited the laboratory. Findings derived from the interview (Step 1) and the experimental phase regarding the think-aloud protocols (Step 4) have been presented in previous publications (Panadero et al., 2022; 2024; Pinedo et al., 2023). Data for this study were obtained in Steps 2, 3, 5, and 6.

1. Participants were interviewed about their self-assessment practices (i.e., self-assessment training, processes, and use of feedback).
2. Participants received instruction in the experimental phase, and they completed the emotion and self-efficacy questionnaires for the first time.
3. Participants were asked to self-assess their own essay by thinking aloud and then, giving their essay a score from 0 to 10. Immediately after this self-assessment, they completed the emotion and self-efficacy questionnaires again.
4. Participants received feedback according to the feedback condition they were assigned to (i.e., rubric, instructor's feedback, or a combination of both). Then, they were asked to self-assess their own essay in light of the feedback, using a think-aloud method.
5. All participants received the instructor's score for their essays, together with the self-score.
6. Participants participated in an interview and completed the emotion and self-efficacy questionnaires for the third time (Fig. 1).



Note. Data in this study obtained from self-reported inventory responses shown in blue boxes

Fig. 1 Procedure of the study

Data analysis

Given the large number of manifest variables in the model (18 emotions and self-efficacy items * three waves) and the modest sample size ($n = 112$), all factors were reduced to item parcels according to the confirmatory factor analysis. Parcelling into scale scores was deemed appropriate because the CFA results supported a clear and theoretically coherent factorial structure for each construct with robust scale reliability. Parcel scores are known to reduce item-specific random error and enhance the stability of parameter estimates in structural models (Little et al., 2013). This reduced the model to nine manifest variables for self-efficacy and the two emotions. With the addition of the discrepancy score, this approach produced a robust case to variable ratio (12.6:1; Costello & Osborne, 2005).

Because the self-efficacy and emotion questionnaires were administered together at the three time-points, their values were inter-correlated within each data wave. We used path model analysis to examine the relationship between variables over time, and the changes in the variables when the event of discovering a discrepancy score occurred. Specifically, we used autoregressive path modelling with manifest scores as it is considered an effective approach with small sample sizes (Curran & Bollen, 2001). Autoregressive path models assume each measure is predicted by itself immediately prior to the new measurement. Cross-lagged paths, which refer to the influence of one variable on another over time, were also tested. This means examining whether self-efficacy could predict changes in emotions (positive or negative) over time and whether emotions could predict changes in self-efficacy. However, none of the cross-lag paths were statistically significant, so they were excluded from the model (i.e., set to zero).

Models can be considered close to the data and not rejected when the most robust fit indices meet established standards: an acceptable model includes a non-significant normed χ^2 (i.e., $\chi^2/df < 3.80$), Comparative Fit Index (CFI) > .90, Root Mean Square Error of Approximation (RMSEA) < .08, and the Standardized Root Mean Residual (SRMR) is ≈ .06 (Fan & Sivo, 2007).

Additionally, we tested the equivalence of the model across feedback conditions through multigroup invariance testing. This requires constraining the model to be equivalent for each parameter, assessing the fit statistics, and moving on to the next parameter test when the fit indices showed equivalence across the parameter values (Brown et al., 2017). The sequence of invariance testing is (1) identical configuration of factors, items, and paths (configural equivalence), (2) factor to item regression weights (metric invariance), and (3) item intercepts (scalar invariance). This level of equivalence is required to compare means. However, because the model does not have latent variables, only metric invariance is required. A good indicator of equivalent regression weights is a small change in CFI (≤ .01) (Cheung & Rensvold, 2002). All analysis was carried out using AMOS software using maximum-likelihood estimation.

Results

The results are organized into three sections. First, we present the results from the autoregressive path model, examining the changes in emotions and self-efficacy over time. Next, we analyse the impact of discovering discrepancies in the accuracy of self-assessment on emotions and self-efficacy. Finally, we report the results of the invariance testing of the model across the feedback conditions.

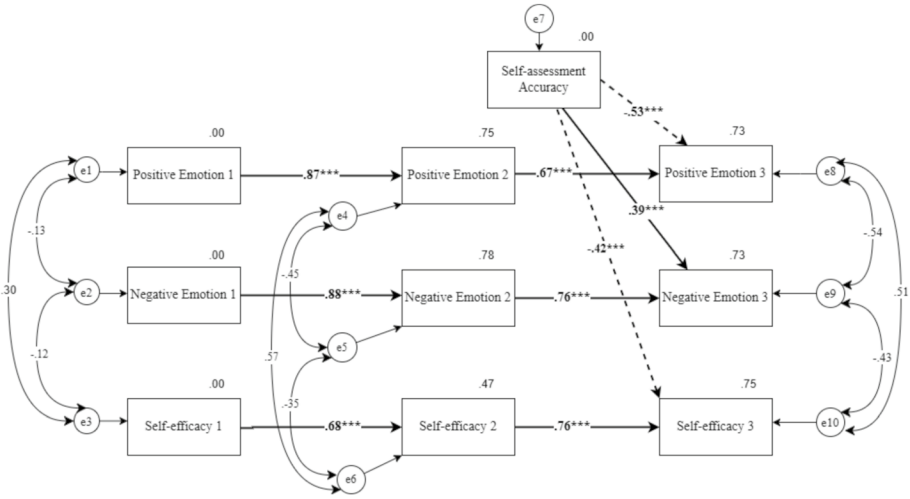
Emotions and self-efficacy over time

The autoregressive path model for describing changes in emotions and self-efficacy after discovering discrepancies in the accuracy of self-assessment is shown in Figure 2. It had an acceptable fit to the data ($\chi^2 = 21.27$; $df = 27$; $\chi^2/df = .78$, $p = .77$; CFI = 1.00; RMSEA = .000, [90% CI = .000–.052]; SRMR = .07). The path model revealed that both positive and negative emotions had autoregressive characteristics in their relationship to each other, with the strength of the paths decreasing across time.

For positive emotions, the path coefficient between Time 1 to Time 2 was β = .87, but the strength of the path decreased in the subsequent interval (i.e., Time 2 to Time 3 β = .67). Simultaneously, positive emotion means increased over time: Time 1 ($M = 4.92$ $SD = 1.64$), Time 2 ($M = 5.18$, $SD = 1.70$), and Time 3 ($M = 5.86$, $SD = 1.94$) (See Fig. 3). This suggests that, on average, participants reported greater positive emotions as the study progressed.

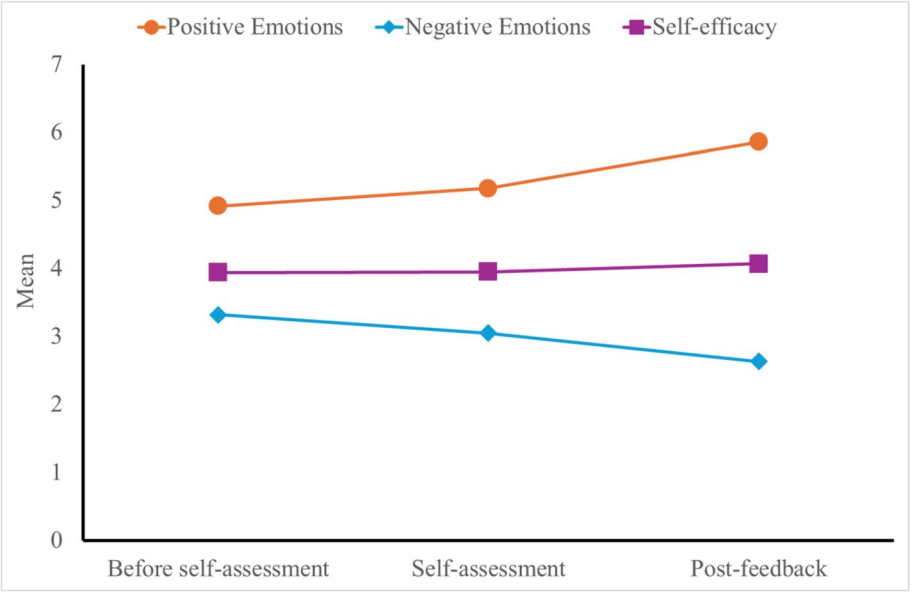
Similarly, for negative emotions, the path coefficients also demonstrated a decline across time (Time 1 to Time 2, β = .88; Time 2 to Time 3, β = .76). The means for negative emotions declined over time: Time 1 ($M = 3.32$, $SD = 1.37$), Time 2 ($M = 3.05$, $SD = 1.33$), and Time 3 ($M = 2.63$, $SD = 1.27$) (Fig. 3). This decline suggests that participants generally experienced fewer negative emotions as they progressed through the study.

However, the path coefficients for self-efficacy increased over time (i.e., Time 1 to Time 2, β = .68; Time 2 to Time 3, β = .76). The means for self-efficacy remained reasonably stable: Time 1 ($M = 3.94$, $SD = 0.73$), Time 2 ($M = 3.95$, $SD = 0.84$), and Time 3 ($M = 4.07$, $SD = 1.01$). This suggests that the self-efficacy was stable, with a trivial increase ($d = .15$) between Time 1 and Time 3.



Note: All values are standardized beta regression weights; positive values for self-assessment accuracy indicate positive illusions of competence, while negative values indicate negative illusions of competence; the negative values of the beta regression weights are represented by dashed lines; correlational values > .90 are statistically significant ($p < .001$) within the sample size.

Fig. 2 Autoregressive path model for predicting changes in emotions and self-efficacy after discovering discrepancies in the accuracy of self-assessment



Note. Before self-assessment refers to the phase before the self-assessment task began, when participants completed the initial questionnaire. Self-assessment refers to the phase in which students assess their own essays before receiving any feedback. Post-feedback refers to the phase after students received the scores on their essays

Fig. 3 Mean scores for self-efficacy, and emotions questionnaires by phase

Changes in emotions and self-efficacy following self-assessment discrepancy

The average self-assessment score ($M = 6.67$, $SD = 1.16$) was slightly lower than the instructor's score ($M = 7.01$, $SD = 1.82$), resulting in an average self-assessment accuracy of -0.35 ($SD = 1.90$). This suggests a general tendency toward underestimation, meaning that, on average, students rated their own performance slightly lower than the instructor did. Nevertheless, the discrepancy is well within tolerances as reported by Boud and Falchikov (1989) for reasonably accurate self-assessment.

The impact of discrepancy on the emotions and self-efficacy of participants were robust. Because the model includes autoregressive paths for emotions and self-efficacy, prior levels of each variable are controlled. As a result, the estimated impact of self-assessment discrepancy can be interpreted as measuring changes in emotions and self-efficacy after controlling for prior levels of emotions and self-efficacy.

Participants with over-estimation scores had reduced positive emotions ($\beta = -.53$) and self-efficacy ($\beta = -.42$), while also exhibiting increased negative emotions ($\beta = .39$). Conversely, these path values indicate that students who underestimated their performance experienced the opposite effect; more positive emotions, less negative emotions, and more self-efficacy. Therefore, the results align with hypotheses H1, H2, and H3, revealing that the direction of self-assessment inaccuracy (overestimation vs. underestimation) is significantly associated with changes in students' emotions and self-efficacy beliefs.

Invariance across feedback conditions

Although the present study does not aim to model between-group differences by feedback condition, invariance testing was conducted to examine whether the structural relations of interest were equivalent across conditions. The model had statistically equivalent regression weights (i.e., metric equivalence) (ΔCFI = .002) across all the feedback conditions. This means that the type of feedback provided (rubric, feedback, or combined) did not significantly influence the results seen in Fig. 2. Specifically, whatever form the feedback took, the impact of discrepancy (over-estimation vs under-estimation) was stable.

Discussion

We investigated how discovering discrepancies in the accuracy of self-assessment affects students' emotions and self-efficacy. Although the associations observed were statistically robust, the average magnitude of change was modest, which is to be expected given that they followed a single academic task. Importantly, even relatively small changes are theoretically meaningful in this context, given the central role of discrepancy discovery in shaping students' emotional responses and self-efficacy beliefs.

Impact on emotions

The autoregressive patterns of emotions found in our study align with previous research, showing considerable stability of emotions over time (Pekrun et al. 2017; 2023b). The gradual decrease in the negative emotion and the increase in positive emotions suggests that as students engage repeatedly in self-assessment, their emotional state became more positive. This could be attributable to greater familiarity with the experimental setup, or possibly greater familiarity with the self-assessment process.

Additionally, our findings highlight the emotional consequences of the discrepancy between self-assessment and external teacher scores. Students who overestimated their performance and faced this discrepancy experienced a decrease in positive emotions and an increase of negative emotions. This aligns with Pekrun's Control Value Theory (2011), which proposes that emotions are directly related to perceptions of control and value. Realizing that you received a lower score than expected could affect students' perceived control over academic outcomes and trigger negative emotions such as frustration or disappointment (Lipnevich & Smith, 2009; Pitt & Norton, 2017). This relationship between performance expectations and emotions is further supported by research showing a positive correlation between grades and positive emotions and a negative correlation between grades and negative emotions (Pekrun et al., 2023b). Similarly, Peterson et al. (2015) found that emotions have a meaningful relationship with achievement once students received their test scores, with lower achievers reporting lower levels of positive emotions and higher levels of negative emotions.

A possible explanation for this emotional response lies in self-assessment accuracy. When students discover that the performance they expected does not align with reality as represented by the instructor's score, this inaccuracy leads to understandable emotional reactions. For some students, receiving a lower score than anticipated may be interpreted as unfair or unexpected, which could evoke emotions such as frustration, disappointment,

and anger (Jonsson & Panadero, 2018; Pekrun et al., 2023a, b). These negative emotions have been shown to lead to disengagement with feedback (Lim et al., 2021), which could, in turn, reduce the opportunity students have to use the feedback to learn what they need to do to improve their performance (Jonsson & Panadero, 2018). To mitigate these negative emotional consequences, future studies should explore interventions designed to (a) help students become more realistic or calibrated in their self-assessments (Brown & Harris, 2014) or (b) support students in coping with discovering their own self-assessment inaccuracy so that they have greater ability to engage with feedback constructively (i.e., increase student feedback literacy; Yan & Carless, 2021). In classroom settings, instructors might consider providing formative assessment across the learning process, allowing students to practice self-assessment, revise their work, and reduce the difference between their self-assessment and actual performance.

Impact on self-efficacy

Regarding self-efficacy, because mean scores remained almost identical across time, this suggests that students did not recalibrate their self-competence perceptions as they engaged in the experimental process. However, the introduction of the accuracy or discrepancy information seems to have changed the basis of their self-efficacy ratings. Self-efficacy for writing was impacted when the self-assessment of their own essay did not align with the teacher grading, both positively or negatively. This is consistent with research that shows academic performance consistently predicted subsequent levels of self-efficacy (Honicke et al., 2023). What is surprising is how strongly self-efficacy shifted after discovering what was on average a relatively small discrepancy across the sample between self-assessed marks and teacher marks.

Clearly, students who over-estimated the quality of their writing had their self-efficacy for writing dented by a considerable margin (i.e., 2/5 of a standard deviation), discovering that they had over-estimated their performance decreased their sense of mastery and, consequently, their confidence in their ability to succeed similar tasks. Conversely, when students experienced a mark higher than their own self-assessment, their self-efficacy was increased by reinforcing their perceptions of competence. This aligns with Bandura (1997) who argued that self-efficacy is sensitive to mastery experiences and feedback, with mastery experience being the strongest predictor of self-efficacy (Britner & Pajares, 2006; Usher & Pajares, 2008).

This result contributes to the literature on factors shaping students' self-efficacy by providing empirical evidence for the relationship between self-assessment accuracy and self-efficacy. Particularly, we found that knowing how accurate one's self-assessment had been helped students to become more aware of their actual performance. This awareness prompted them to adjust their self-efficacy beliefs towards a more realistic and hopefully adaptative state. While previous studies have highlighted the need of feedback to improve self-assessment accuracy (Brown et al., 2015; Panadero et al., 2016a, b), our results extend this understanding by showing that gaining knowledge of how accurate their self-assessment had been also helped calibrate their self-efficacy. Feedback from the teacher helps students become more realistic in their self-efficacy estimates. This result is relevant for teachers designing self-assessment practices, as delivering feedback about the realism of self-assessments can help students better understand their performance and develop more realistic beliefs.

Overall, our findings provide a nuanced perspective on Butler (2011), who suggested that optimistic illusions are generally less problematic than negative ones. Our results indicate, however, that overestimation of performance can have tangible costs, affecting students' emotional and self-efficacy responses. This highlights that even optimistic perceptions of competence may not always be beneficial, particularly when discrepancies with actual performance are revealed. Importantly, the consequences of positive self-evaluative biases are likely to vary across contexts, suggesting that their impact depends on factors such as task type, learning environment, and individual differences.

Fostering self-assessment accuracy to support emotions and self-efficacy

Given our findings on the emotional and motivational costs of overestimation, it is crucial to consider ways that teachers can foster self-assessment accuracy and, in doing so, help mitigate negative emotional responses and support students' self-efficacy. First, understanding the reasons students have to over- or under-estimate their performance is essential for effective intervention. Overestimation may stem from optimistic biases, limited domain knowledge, or a reduced understanding of assessment criteria, while underestimation is often linked to low self-efficacy, perfectionism, or conservative scoring tendencies shaped by prior academic experiences (Brown et al., 2015; Dunning et al., 2004; Panadero et al., 2016a, b; Pinedo et al., 2025).To mitigate these biases and promote better calibration, teachers can help students develop more accurate self-assessments by focusing on three key areas. Teachers should ensure criteria are explicit and supported by exemplars to create a shared understanding of quality, reducing uncertainty and emotional stress. Second, implementing formative assessment practices that allow students to practice self-assessment and compare their judgments with external standards (e.g., the teacher's score or a model answer) provides essential opportunities for calibration. Finally, teachers should focus on constructive feedback and reflection, which means providing feedback that specifically addresses the accuracy of the self-assessment itself, encouraging reflection on the discrepancies observed. In sum, fostering accurate self-assessments goes beyond helping students understand their performance: it helps them understand when their expectations differ from feedback, manage their emotions, and build their confidence, turning each assessment into a chance to learn and grow.

Limitations and future lines of research

Several methodological limitations should be acknowledged. The study took place in a specific domain (academic writing in psychology), educational level (higher education), and country (Spain), and did not consider how sociocultural or gender-related factors may influence students' self-assessment processes. Conducting similar studies across various subjects, task types, and levels of education could provide valuable insights about the generalisability of these results. This is especially needed in secondary education, where research is limited (León et al., 2023), or in domains like science, which have previously been found to yield more accurate self-assessments (Falchikov & Boud, 1989).

Although the study employed a longitudinal design and autoregressive modelling, causal inferences should be interpreted with caution and within the specific context of this study. While the sample size was adequate for the autoregressive path model, larger samples would allow the estimation of more complex models, including fully latent variable models and additional longitudinal relations. The use of item parcelling maintained an

appropriate ratio between cases and parameters, given the modest sample size. Future studies with larger samples could model emotions and self-efficacy as latent constructs using individual items, avoiding the need for parcelling.

Additionally, since emotional data was collected through a questionnaire, our ability to understand the specific reasons behind students' emotional experiences is limited. Additionally, this study did not investigate whether the students questioned the accuracy or fairness of the instructor's assigned scores. This may be an important factor shaping emotional responses and self-efficacy, as perceiving a score as unfair could amplify negative emotions or reduce engagement with feedback. Finally, data regarding task value or perceived control during the self-assessment were not collected; however, these are essential factors to understand relationship between appraisals, emotions and self-assessment (Pekrun et al., 2023a). To address these limitations, subsequent studies could collect additional measures to capture these contextual factors and include qualitative methods to better understand students' experiences and underlying reasons for their emotional responses during self-assessment. In addition, examining whether feedback that reveals inaccuracies in self-assigned scores can improve both self-assessment accuracy and self-efficacy in later tasks would shed light on how feedback facilitates calibration of performance and confidence over time. Investigating interventions designed to enhance self-assessment accuracy and examining their broader impact on psychological variables such as emotions and motivation could further advance our understanding of these processes.

Conclusion

The main contribution of this study lies in its revelation that feedback as to the realism of self-assessments matters to both those with optimistic and pessimistic illusions of performance. Overly optimistic students experienced negative feelings and reduced self-efficacy, while more pessimistic students felt better and had greater task confidence. This raises doubts about the general benefit of having optimistic illusions about competence (Butler, 2011). Mitigating overestimation, in particular, may help reduce negative emotional responses and foster the development of more stable self-efficacy beliefs. This study shows that students who want to feel better about assessed activities when they get their marks should be more modest in their self-assessments. This modesty bias does not imply undervaluing one's abilities but encourages aligning self-assessments with external standards, reducing emotional disappointment, and supporting more stable self-efficacy. Importantly, the role of teachers is critical in scaffolding these processes. By providing clear assessment criteria, constructive feedback, and opportunities for formative assessment, instructors can guide students to practice self-assessment, reflect on performance discrepancies, and use these experiences as a resource for learning and self-regulation (Andrade, 2019; Brown et al., 2015).

Further, this study shows that feedback that reveals an over-estimation does have a negative emotional and motivational impact on learners. The study does not reveal if students used this apparently negative feedback to improve performance, which would be consistent

with students who understand feedback exists to support improved learning (Brown et al., 2016). Teachers cannot deny learners access to the truth about their performance to protect their psycho-emotional state. Hence, all learners must learn how to be realistic and modest in their self-evaluations. This, of course, is not easy given human predilections to overestimating their abilities and work (Dunning et al., 2004). Herein lies the challenge for both teacher and learner; the former must know how to give truthful feedback and help students cope with the inevitable emotional reaction, while the latter has to learn how to embrace disappointing results as a learning process.

These insights contribute to a deeper understanding of the psychological mechanisms underlying the impact of accuracy or realism on self-assessment. By demonstrating how score discrepancy impacts emotions and self-efficacy, our findings highlight the emotional challenges students face when self-assessment is contrasted with external assessments.

Appendix A

Rubric

Category	Low quality	Average quality	High quality
Writing process	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

Category	Low quality	Average quality	High quality
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 appro- priately. However, there may be one or more of these flaws: Same paragraph includes different unorganized ideas Same idea in two paragraphs 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, the 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

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									

	1st Occasion	2nd Occasion	3rd Occasion
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			

Appendix C

Table 1 Descriptive statistics for scores of positive and negative emotions, and self-efficacy

	Pre-assessment		Before feedback		After feedback	
	M	*SD*	*M*	*SD*	*M*	*SD*
Positive Emotions	4.92	1.64	5.18	1.70	5.86	1.94
Negative Emotions	3.32	1.37	3.05	1.33	2.63	1.27
Self-efficacy	3.94	0.73	3.95	0.84	4.07	1.01

N=112

Authors' contributions Leire Pinedo: conceptualization, data curation, formal analysis, investigation, visualization, writing – original draft, and writing- review & editing. Ernesto Panadero: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, and writing- review & editing. Javier Fernández-Ruiz: data curation, methodology, resources, and data collection. Daniel García-Pérez: data curation, investigation, methodology, and data collection. Gavin T. L. Brown: data curation, formal analysis, supervision, and writing- review & editing.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. (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: PID2024-156609NB-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.

Availability of data and material The data and materials from this study are not publicly available.

Declarations

Ethics approval and consent to participate 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.

Consent for publication The authors listed in the by-line have agreed to the order of authorship and consent to the publication of the manuscript upon acceptance.

Competing interests The authors declare to not having any conflict of interest regarding this manuscript.

References

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*, Article 87. https://doi.org/10.3389/feduc.2019.00087

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. Freeman.

Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2015). Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning. *Metacognition and Learning, 10*, 99–117. https://doi.org/10.1007/s11409-014-9127-x

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*(5), 529–549. https://doi.org/10.1007/BF00138746

Bourke, R. (2017). Self-assessment to incite learning in higher education: Developing ontological awareness. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2017.1411881

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*(5), 485–499. https://doi.org/10.1002/tea.20131

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Sage.

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research, 3*, 22–30. https://doi.org/10.14786/flr.v2i1.24

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

Brown, G. T., Peterson, E. R., & Yao, E. S. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *British Journal of Educational Psychology, 86*(4), 606–629. https://doi.org/10.1111/bjep.12126

Brown, G. T. L., Harris, L. R., O'Quin, C., & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: Identifying and understanding non-invariance. *International Journal of Research & Method in Education, 40*(1), 66–90. https://doi.org/10.1080/1743727X.2015.1070823

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

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, *10*(7). https://doi.org/10.7275/jyj1-4868

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107–135). APA.

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

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395–430. https://doi.org/10.3102/00346543059004395

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Goetz, T., Lipnevich, A. A., Krannich, M., Gogol, K. (2018). Performance feedback and emotions. In Lipnevich A. A. & Smith, J. K. (Eds.) *Cambridge Handbook of Instructional Feedback*. Cambridge University Press.

Honicke, T., Broadbent, J., & Fuller-Tyszkiewicz, M. (2023). The self-efficacy and academic performance reciprocal relationship: The influence of task difficulty and baseline achievement on learner trajectory. *Higher Education Research & Development, 42*(8), 1936–1953. https://doi.org/10.1080/07294360.2023.2197194

Jonsson, A., & Panadero, E. (2018). Facilitating students' active engagement with feedback. *The Cambridge Handbook of Instructional Feedback, 10*(2), 531–553. https://doi.org/10.1017/9781316832134.026

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

Lim, L. A., Dawson, S., Gašević, D., Joksimović, S., Pardo, A., Fudge, A., & Gentili, S. (2021). Students' perceptions of, and emotional responses to, personalised learning analytics-based feedback: An exploratory study of four courses. *Assessment & Evaluation in Higher Education, 46*(3), 339–359. https://doi.org/10.1080/02602938.2020.1782831

Lipnevich, A. A., & Smith, J. K. (2009). I really need feedback to learn: " Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability, 21*, 347–367. https://doi.org/10.1007/s11092-009-9082-2

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18*(3), 285–300. https://doi.org/10.1037/a0033266

Lui, A. M., & Andrade, H. L. (2022). Inside the next black box: Examining students' responses to teacher feedback in a formative assessment context. *Frontiers in Education, 7*, Article 751549. https://doi.org/10.3389/feduc.2022.751549

Panadero, E., Brown, G. T. L., & Courtney, M. G. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice, 21*(4), 365–383. https://doi.org/10.1080/0969594X.2014.919247

Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016a). the future of student self-assessment: A review of known Unknowns and potential directions. *Educational Psychology Review, 28*(4), 803–830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E., Jonsson, A., & Strijbos, J. W. (2016b). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the challenge of implementation* (pp. 311–326). Springer.

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

Panadero, E., Pérez, D. G., Fernández-Ruiz, J., Fraile Ruiz, J., Sánchez-Iglesias, I., & Brown, G. T. L. (2022). University students' strategies and criteria during self-assessment: Instructor's feedback, rubrics, and year level effects. *European Journal of Psychology of Education, 38*, 1031–1051. https://doi.org/10.1007/s10212-022-00639-4

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

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child development, 88*(5), 1653–1670.

Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023a). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology, 124*(1), 145–178. https://doi.org/10.1037/pspp0000448

Pekrun, R., Marsh, H. W., Suessenbach, F., Frenzel, A. C., & Goetz, T. (2023b). School grades and students' emotions: Longitudinal models of within-person reciprocal effects. *Learning and Instruction, 83*, Article 101626. https://doi.org/10.1016/j.learninstruc.2022.101626

Peterson, E. R., Brown, G. T., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology, 42*, 82–96. https://doi.org/10.1016/j.cedpsych.2015.05.002

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

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

Pitt, E., & Norton, L. (2017). 'Now that's the feedback I want! 'Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education, 42*(4), 499–516. https://doi.org/10.1080/02602938.2016.1142500

Richardson, M., Bond, R., & Abraham, C. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin 138*(2), 353–387. https://doi.org/10.1037/a0026838

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://doi.org/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

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural equation modeling and confirmatory factor analysis results: A Review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Shao, K., Pekrun, R., Marsh, H. W., & Loderer, K. (2020). Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction, 69*, Article 101356. https://doi.org/10.1016/j.learninstruc.2020.101356

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*, 467–481. https://doi.org/10.1007/s10734-017-0220-3

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Tan, K. H. (2008). Qualitatively different ways of experiencing student self-assessment. *Higher Education Research & Development, 27*(1), 15–29. https://doi.org/10.1080/07294360701658708

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*(4), 751–796. https://doi.org/10.3102/0034654308321456

Villafane, S. M., Xu, X., & Raker, J. R. (2016). Self-efficacy and academic performance in first-semester organic chemistry: Testing a model of reciprocal causation. *Chemistry Education Research and Practice, 17*(4), 973–984. https://doi.org/10.1039/C6RP00119J

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education, 45*(2), 224–238. https://doi.org/10.1080/02602938.2019.1629390

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education, 42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

Yan, Z., & Carless, D. (2021). Self-assessment is about more than self: The enabling role of feedback literacy. *Assessment & Evaluation in Higher Education, 47*(7), 1116–1128. https://doi.org/10.1080/02602938.2021.2001431

Yan, Z., Lao, H., Panadero, E., Fernandez-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review, 37*, Article 100484. https://doi.org/10.1016/j.edurev.2022.100484

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166–183. https://doi.org/10.3102/0002831207312909

"Current themes of research" and "most relevant publications" in the field of psychology of education. Leire Pinedo.

Current themes of research: Educational assessment (self-assessment, and teacher's feedback). Rubrics. Self-regulation. Psychology of assessment.

Most relevant publications in the field of Psychology of Education:

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

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.

Ernesto Panadero.*Current themes of research:* Self-regulated learning. Formative assessment (self-assessment, peer assessment, and teacher's feedback). Rubrics. Socially shared regulated learning.*Most relevant publications in the field of Psychology of Education:*

Panadero, E. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. *Educational Psychologist, 58(*3), 193-204. https://doi.org/10.1080/00461520.2023.2223642

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*(4), 803–830.

Javier Fernández-Ruiz.

Current themes of research: Higher education. Formative assessment. Teacher education. Assessment design.

Most relevant publications in the field of Psychology of Education:

Fernández-Ruiz, J., Panadero, E., & García-Pérez, D. (2021). Assessment from a disciplinary approach: Design and implementation in three undergraduate programmes. *Assessment in Education: Principles, Policy & Practice, 28*(5–6), 703–723.

Fernández Ruiz, J., Panadero, E., García-Pérez, D., & Pinedo, L. (2021). Assessment design decisions in practice: Profile identification in approaches to assessment design. *Assessment & Evaluation in Higher Education,* 1–16.

Daniel García-Pérez.*Current themes of research:* Educational research. Democratic education. Assessment. Learning strategies.

Most relevant publications in the field of Psychology of Education:

García-Pérez, D., Fraile, J., & Panadero, E. (2021). Learning strategies and self-regulation in context: How higher education students approach different courses, assessments, and challenges. *European Journal of Psychology of Education, 36*(2), 533–550.

Panadero, E., García-Pérez, D., Fernández Ruiz, J., Fraile, J., Sánchez-Iglesias, I., & Brown, G. T. L. (2023). Feedback and year level effects on university students' self-efficacy and emotions during self-assessment: positive impact of rubrics vs. instructor feedback. *Educational Psychology, 43*(7), 756–779. https://doi.org/10.1080/01443410.2023.2254015

Gavin T.L.Brown.

Current themes of research: Educational assessment. Psychology of assessment. Psychometrics.

Most relevant publications in the field of Psychology of Education:

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. McMillan (Ed.), The SAGE handbook of research on classroom assessment (pp. 367–393). SAGE.

Authors and Affiliations

Leire Pinedo[1] · Ernesto Panadero[1,2,3] · Javier Fernández-Ruiz[4] · Daniel García-Pérez[5] · Gavin T. L. Brown[6]

✉ Leire Pinedo
leire.pinedo@opendeusto.es

[1] Education, Regulated Learning and Assessment Group (ERLA Group), Faculty of Education and Sport, ERLA Research Group, Universidad de Deusto, Avda. Universidades 24, 48007 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 y Psicología en Educación, Facultad de Educación-Centro de Formación del Profesorado, Universidad Complutense de Madrid, Madrid, Spain

[6] Faculty of Arts & Education, The University of Auckland, Auckland, New Zealand