Revista Românească pentru Educație Multidimensională
ISSN: 2066-7329 | e-ISSN: 2067-9270
Abstracting & Indexing | Clarivate Analytics - Web of Science (WoS); EBSCO; Google Scholar; ICI Journals Master List - Index Copernicus; Ideas RePeC; Econpapers; Socionet; CEEOL; Philpapers; KVK; WorldCat; CrossRef; J-GATE

2025, Volume 17, Issue 1, pages: 116-137 | https://doi.org/10.18662/rrem/17.1/943
Submitted: August 29, 2024 | *Accepted for publication:* December 11, 2024

Self-Assessment in Teacher Education: Impact for Educational Evaluation

Otilia CLIPA[1]

[1] "Stefan cel Mare" University of Suceava, Romania. otilia.clipa@usm.ro

Abstract: *Self-assessment has a significant place among modern assessment types. More importantly, self-assessment plays an essential role in pre-service teacher training, as it facilitates comprehensive learning and prepares future educators effectively. Thus, this paper aims to analyse the potential of self-assessment in teacher education. The study examines the extent to which two different assessment methods—self-assessment and hetero-assessment—yield similar results for students taking a test and whether these results are influenced by group size and academic performance levels. The formative outcomes of self-assessment in university are highlighted, as supported by existing literature. The study investigates the factors influencing students' self-assessment processes, including group size, academic performance and assessment methods. The findings indicate that students' self-assessment results are significantly better than those obtained from hetero-assessment. Thus, we argue that through self-assessment techniques, educators can foster a genuine culture of self-assessment, enabling students to develop an accurate self-image and take responsibility for their own outcomes. The study concludes that it is essential for teachers to recognise the importance of self-assessment in developing metacognitive abilities, deep learning and self-responsibility.*

Keywords: *self-assessment, teacher education, educational evaluation.*

How to cite: Clipa, O. (2025). Self-assessment in teacher education: impact for educational evaluation. *Revista Românească pentru Educație Multidimensională, 17*(1), 116-137. https://doi.org/10.18662/rrem/17.1/943

©2025 Published by In Manifest Network in partnership with LUMEN Publishing. This is an open access article under the CC BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Introduction

Current literature on pedagogy reveals that assessment is a didactic process that significantly influences other components of the academic curriculum. Assessment is described as 'the heart of the learning process' (Carless, 2013; Gibbs & Coffey, 2004; Hosein & Harle, 2018; Zeng et al., 2018) and 'the heart of student experience' (Brown et al., 2011). The assessment process allows students to reconfigure their entire learning experience; 'the student's experience about assessment determines the way they will approach the following learning' (Bale & Seabrook, 2021; Graham, 2005; McCarthy et al., 2018; Panadero et al., 2017; Struyven et al., 2008). It can be argued that 'assessment drives learning' (Norton, 2007). The learning style and its profundity depend on the assessment methods employed, but the process can also be the other way around, with assessment serving as a tool for learning and understanding (Black & William, 2009; Brown et al., 2011; Remesal, 2011; Zeng et al., 2018). Learning methods and the intensity of the process determine the performance reflected in assessments. In this context, inappropriate assessment procedures can lead to superficial learning (Struyven et al., 2008). Assessment plays a critical role in determining the quality of student learning (McCarthy et al., 2018, Ramsden, 2003). Norton (2007) states that 'students are more strategic and marks-oriented than they are'. If assessment and the learning styles are interdependent, then assessment becomes a context-dependent process. Students adjust their data processing methods according to the assessments they receive for the learning efforts (Segers et al., 2008). However, assessment for learning requires a context-free approach, promoting profound learning and the acquisition of complex, interdisciplinary skills (Hosein & Harle, 2018). Boud asserts that 'the purpose of assessment should be extended to include the students' preparation for sustainable assessment'. Sustainable assessment implies that assessments have long-term goals, fostering responsibility for one's own training and assessment involvement as a way of life, based on self-assessment and peer assessment. The results and perceptions of assessment among students were also identified by Brown et al. (2011). The authors noted that high-achieving students view assessment as a means of assuming learning responsibility and use assessment to improve their learning (Bale & Seabrook, 2021; Carless, 2013; Lowell & Ashby, 2018; Pope, 2005).

The role of self-assessment in teacher education – theoretical framework

If assessment plays a significant role in student training in general, it assumes an even greater importance in teacher training. The assessment methods employed will influence students' lifelong learning styles, their responsibility for their training to become teachers and the assessment methods they will utilise. Research has shown that becoming a lifelong learner necessitates the development of learning-how-to-learn abilities (Boud & Falchikov, 2006; Pope, 2005; Taras, 2010; Sridharan et al., 2019), as well as learning-how-to-evaluate-learning abilities and learning regulatory strategies (Panadero et al., 2017). These abilities are correlated not only with reaching performances in learning—when the student gives feedback on others' assessment—but also with fostering responsibility for their own results, which entails active participation in the assessment process. Consequently, we can discuss an assumed assessment, where future teachers can play the role of the assessee, as well as doing the assessing—assessing their own results and developing self-assessment abilities. Trainee teachers must know this type of assessment, as the responsibilities of the assessor are cultivated more through practice than through theoretical lectures on assessment. During their academic training, students participate in many theoretical and practical activities to prepare for teaching. Thus, they find themselves in the position of learning-how-to-evaluate. It has been suggested that the assessment methods employed during practical training have a more significant impact than what is learnt in assessment lectures (Brown et al., 2011; Struyven et al., 2008). Therefore, educators in teacher education should recognise the value of assessment and apply various types of assessment (peer-assessment and self-assessment), clarify assessment criteria at the beginning of the course and eventually discuss results with the students and provide feedback. This approach would encourage students to view the assessment process as a curricular partner necessary for training and promote learning through assessment and self-assessment. The introduction of peer- and self-assessment methods in teaching programmes has the potential to strengthen the link between tutor feedback and student learning (Clipa, 2015; Panadero et al., 2013; Panadero et al., 2013; Ross & Bruce, 2007; Sridharan et al., 2019). Moreover, self-assessment plays a significant role in pre-service teacher education, as it reflects what future teachers have learnt and, importantly, what they can improve in their training. They take ownership of their training and the results of their assessments.

This type of assessment is regarded as 'an educative objective for the professor and a principle of the learning process adjustment for the students' (Arredondo & Diago, 2005a,b). Kitsantis et al. (2004) assert that self-assessment (self-judgment) is linked to the acquisition level and serves as an important predictive factor in learning motivation and its components (Clipa et al., 2011). The achievements of the best students result from a focus on the process objectives and a high degree of self-learning efficacy (McDonald & Boud, 2003; Panadero et al., 2017; Ross & Bruce, 2007). This acquisition of ability is perceived as a process of building, validating, applying and assessing the criteria used for evaluating students' learning outcomes. Students may choose how they will be assessed based on their material presentations (for example, assessing their work using activity evaluations) and the most suitable methods for improvement. Self-assessment is more akin to a coaching process regarding examination content, aimed at equipping students with skills applicable in real life, enabling them to take responsibility for their own training alongside the teacher, who can act as an intermediary (Boud & Falchikov, 2006; McDonald & Boud, 2003).

Boud also promotes inter-evaluation and self-evaluation at the university level in Australia. He states that students begin to assume responsibility for their learning and performance by evaluating the performances of others and providing feedback. He also claims that self-evaluation occurs when the assessor acts as their own 'agent and auditor'. Boud & Falchikov (2006) expands this definition, arguing that it involves 'the judgement in which one's results are compared against certain standards and criteria'.

McDonald & Boud (2003) note that self-assessment is 'students judging the quality of their work, based on evidence and explicit criteria for the purpose of doing better work in the future' (p. 3). Therefore, universities should prioritise assessment strategies, particularly self-evaluation, to enhance reflective skills and promote profound, lasting learning. Assessment strategies that encourage meta-assessment and self-assessment reflexes have been priority objectives for university education, employing portfolios and journals of alternative assessment methods.

The arguments for using self-assessment in teacher training can be both pedagogical and psychological. From a pedagogical perspective, self-assessment is important because it

- reduces evaluation time, as it is a natural process for students to judge their own results (Bein, 2004; Sadler & Good, 2006);
- aids students in understanding and internalising the criteria by which they are assessed (Panadero et al., 2013);

- facilitates deeper learning beyond specific content (Sadler & Good, 2006; Panadero et al., 2013);
- eliminates or diminishes the subjectivity of grading and the discrepancies between learning assessments and professor assessments (Day et al., 2018; Milne, 2008; Orsmond et al., 2000);
- provides essential feedback for professors in designing future teaching or practical activities;
- cultivates responsible assessors, who would know how to apply various evaluation types (Bannink, 2010; Clipa, 2015; Hosein & Harle, 2018);
- lays the groundwork for permanent education (Boud & Falchikov, 2006);
- establishes mechanisms for professional growth (Brown et al., 2011; Kearney, 2013; Ross & Bruce, 2007);
- extends student engagement within didactic activities, training them through their own example to develop self-assessment abilities (Remesal, 2021); and
- demystifies assessment while building awareness (Milne, 2008; Sadler & Good, 2006).

From a psychological standpoint, the arguments for utilising self-assessment can pertain to cognitive, affective and social development.

- Cognitive
  - Provides support points in learning, assisting students in identifying strategies prior to correction (Taras, 2016);
  - Facilitates adjustment to academic learning and internalisation of the teacher criteria (Andrade, 2010; Panadero et al., 2017; Taras, 2016)
  - Promotes good learning practices and the ability to learn how to learn (Panadero et al., 2015)
  - Enables predictions regarding their own performance (Hosein & Harle, 2018)
  - Develops metacognitive processes and enhances self-efficacy (Clipa et al., 2011; Panadero et al., 2013; Taras, 2016)
  - Stimulates deep-level learning and critical thinking (Knight, 2005; Struyven et al., 2008; Panadero et al., 2016)
- Affective
  - Increases self-confidence, particularly if the teacher positively values this type of assessment
  - Fosters collaborative learning abilities

- Positive feedback cultivates a positive attitude as opposed to low marks and encourages reflection on learning outcomes (Panadero et al., 2017)
- Is less intimidating than assessment through tests (Struyven et al., 2008)

➢ Social

- Addresses the need for self-awareness and social comparison
- Maintains individual autonomy,
- Develops moral reasoning and fairness (Clipa, 2015; Ion et al., 2024),
- Models self-awareness through received feedback and honesty (Stufflebeam & Wingate, 2005)
- Cultivates a critical attitude towards the results of peers and towards prior personal results

Through the process of self-assessment, students assume autonomy over their educational and evaluative processes. Consequently, the character traits of students' personalities and the way to increase responsibility for their own academic results are of paramount importance in the self-evaluation process. These traits are crucial for ensuring the accuracy of self-evaluation, thereby guaranteeing the reliability and validity of grading. If self-assessment is valid, it must align with the need for accurate self-knowledge and positive self-perception, serving as a strong motivation for learning.

Self-assessment represents a vital means of fostering personal autonomy and enhancing educational activities. It also serves as a psycho-pedagogical strategy for adjusting self-esteem and students motivation. 'With the help of self-evaluation, the students get accustomed to review their behaviour towards a diversity of learning aspects: awareness of one's efforts, difficulties, level of aspiration and their own learning efficiency' (Arredondo & Diago, 2005a,b). This process mediates the formation of self-awareness, self-esteem and personal value in relation to individual objectives and expectations. It is a means of reflective and responsible assessment. However, not all students can acquire self-assessment abilities, as this depends on their level of awareness and maturity.

Panadero et al. (2016) identify three reasons for self-assessment accuracy: students' understanding of self-assessment criteria, their performance or achievement level and their level of expertise. They demonstrate that when assessment criteria are stricter and well defined, students are more likely to self-assess accurately.

Academic research indicates that self-assessment develops from evaluations conducted by other colleagues or significant individuals in students' lives. Furthermore, high-achieving students often tend to underestimate their abilities (Clipa et al., 2011; Lowell & Ashby, 2018). This tendency may be linked to improvements in self-esteem and of a self-concept that may not accurately reflect reality. 'We learn to distinguish between assess and self-assess ourselves for the others' (Brown et al., 2011). The digital landscape presents new challenges for self-assessment; nonetheless, it becomes easier to utilise technology in the assessment process, such as synchronous self-assessment (Remesal, 2022), which is like a digital mirror - a rapid virtual tool for developing personal and professional competencies.

Much of the research on self-assessment has focused on comparing grades assigned by teachers, tutors, peers and even the students themselves. Gibbs & Coffey (2004) highlight the methodological weaknesses of these studies, as it is difficult to ascertain their relevance compared with the other assessment types, particularly when the curricula vary significantly, and clear assessment criteria cannot be established.

This study aims to determine whether we can trust the grades provided by students' self-assessments and to identify factors that can influence this complex process of self-estimation.

Methodology

Research objective and hypothesis

This study aims to analyse the factors influencing the modification (improvement) of grades within students' self-assessment. Specifically, it seeks to ascertain whether students tend to inflate their own grades during the self-assessment process and to what extent these self-assessments are influenced by variables such as group size and academic achievement level.

General Hypothesis. The variables assessment types, group size and academic achievement level influence students' results when taking rubrics tests.

Hypothesis 1. Students' results are influenced by the assessment types applied and the size of the group they belong to.

Hypothesis 2. Students' results in a rubrics test are influenced by the assessment form used and their level of academic achievement.

Sample and variables

The research involved a sample of 295 students. Differentiated by the environmental variable, the sample comprises 147 subjects from urban

areas and 148 from rural areas. In terms of high school type, 144 subjects graduated from pedagogical high schools, while 151 graduated from high schools covering other fields of study. Classified by group size, the sample includes 106 students who participated in larger groups and 189 students in smaller groups. In 2008 was tested 145 students and in 2018 was 150 students. The average age of the subjects was 31 years and 4 months.

The dependent variable was the rubrics test results, while the independent controlled variable was the assessment form, which has two categories: hetero-assessment (students assessed by the professor) and self-assessment (students assessing their own results). The independent variables include academic achievements level: high versus low. Students' academic achievement was measured using the final averages from the last study year. The median for the academic achievement variable was calculated, resulting in a value of 8.76. Subjects with averages above the median were classified as having a high level of academic achievement, while those with averages below the median were classified as having a low level of academic achievement. Subjects with averages equal to the median were excluded from the analysis. The second independent variable was group size, categorised as either small or large.

Instruments and procedures

To measure the students' performance in the rubrics test, a knowledge test on the assessment's theory and methodology was developed. The rubrics test comprised 10 objective and semi-objective items designed to assess the level of competencies acquired in the second-year lecture.

To obtain the students' grades through hetero-assessment, the test was evaluated and marked by the teacher according to a correction standard. For self-assessment, students were given the opportunity to self-correct their papers according to the same correction standard, with the teacher present, but without disclosing their grades. Groups were classified as small or large based on whether they contained fewer or more than 25 students.

The research was conducted over two periods (January–February 2008 and January–February 2018) during the students' exams at the Educational Sciences Faculty – Primary and Preschool Education about Assessment Methodology. These two periods highlight the stability of measure and validity of the self-assessment types and their contemporary applicability. This longitudinal study demonstrates that types of assessments over 10 years is valid and provide insight into the influence of assessment types and group size on authentic assessment. Following these longitudinal tests, self-assessment is now employed as part of summative assessment for students.

This study was conducted during the sophomore year of a three-year teacher training programme. Under the Bologna system, students obtain a teaching licence for primary and preschool levels after three years of study. Initially, the rubrics test was administered, after which the exam papers were collected by the teacher. Subsequently, students were invited to take part in the research by writing a letter to an academic assessor. Meanwhile, all papers were copied and returned to the students for self-correction and grading. The papers were returned, and students could assign themselves points for each item. The correct results were discussed between the teacher and students item by item. Students wrote the solution on the right side of the paper, and some of them modified the answer during the self-correction. Students were informed about the research aims and their right to withdraw from participation.

To quantify the number of changes made, we compared the exam copies with the original papers, noting any modifications.

Research design

We employed two multiple ANOVA variance analysis models, incorporating a repeated measures variable (assessment method) and an inter-group variable.

(a) Research design type 2 x 2 (2 assessment method categories x 2 group categories), with four independent groups: [Figure 1]

Figure 1 *Research Design: Group and Assessment Types*

Variable		**Assessment Types**	
		Hetero-assessment	Self-assessment
Group	large	G1	G3
	small	G2	G4

(b) Research design type 2 x 2 (2 assessment method categories x 2 academic achievement levels), with four independent groups: [Figure 2]

Figure 2 *Research Design: Academic Achievement and Assessment Types*

Variable		**Assessment types**	
		Hetero-assessment	Self-assessment
Academic achievement levels	High level	G1	G3
	Low level	G2	G4

Analysis and Results Interpretation

Data processing was conducted using SPSS 19.0 for Windows. The statistical analyses included the following:

1. Frequency analysis for visualising the sample composition, median for academic achievement (the average from the previous study year) and average for the age variable

2. Multiple ANOVA variance analysis to study the influence of the variables assessment method and group on the test result variable (Hypothesis 1)

3. T-tests for paired samples to confirm the influence of the assessment method variable on the test result variable (Hypothesis 1)

4. ANOVA multiple variance analysis to study the influence of the assessment method and performance levels on test results (Hypothesis 2)

Hypothesis 1. Students' test results are influenced by the assessment form applied and the size of the group they belong to.

To test this hypothesis, we performed multiple ANOVA variance analysis, with assessment types and group size as independent variables and test results as the dependent variable.

According to the statistical analysis results, there is notably a main effect of the assessment form variable on the test results variable [$F (1, 293) = 106.746$, $p < 0.05$], and there is no main effect of the group size variable on the test results variable [$F (1, 293) = 3.557$, $p = 0.060$]. There is also an effect of the variables assessment types and group size on the test results variable [$F (1, 293) = 78.691$, $p < 0.05$] (Table II.2). Table 1 presents the averages, standard deviations and subject samples for the scores regarding the test results variable in relation to the variables assessment types and group size.

Table 1 Averages, Standard Deviations and Subject Samples for the Scores Regarding the Test Results Variable in relation to the Assessment Types and Group Size

		Assessment types					
		Hetero-assessment			Self-assessment		
		M	SD	N	M	SD	n
Group	large	3.65	0.94	106	4.21	0.65	106
	small	4.08	0.78	189	4.12	0.72	189
Total		3.92	0.86	295	4.15	0.70	295

Table 2 Results of ANOVA Multiple Variances for Influence of Variables Assessment Types and Group Size on Test Results Variable

Variance source	SS	df	MS	F	p
Assessment types (FE)	12.661	1	12.661	106.746	0.000
Group (G)	3.852	1	3.852	3.557	0.060
FE x G	9.333	1	9.333	78.691	0.000
Residual (error) (FE)	34.752	293	0.119		
Residual (error) (G)	317.320	293	1.083		

Main effect of the assessment types of variables on the test results variable. There is a main effect of the assessment types of variables on the test results variable [F (1, 293) = 106.746, $p < 0.05$] (Table II.2). To further investigate this main effect, we conducted a t-test for paired samples. The results indicate significant differences between the results of the rubrics test concerning assessment types [t (294) = 7.262, $p < 0.05$], revealing that the grades from the self-assessment method are significantly higher than those obtained through hetero-assessment.

Thus, in the self-assessment process, students tended to inflate their own grades. The grades obtained through self-assessment are significantly better than those resulting from teacher assessment. There is no significant main effect of the group size variable on the test results variable [F (1, 293) = 3.557, $p = 0.060$] (Table 2).

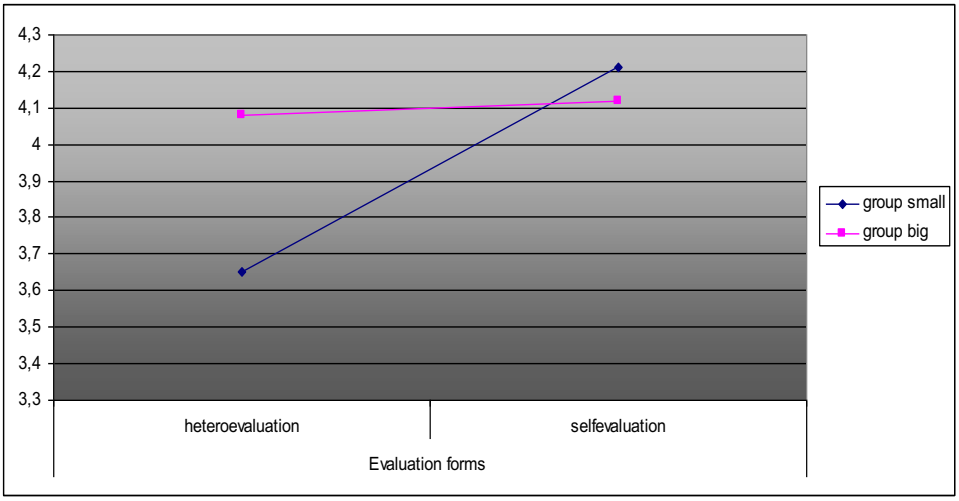
Interaction effect of assessment types and group size on test results variable. There is a significant interaction effect of the assessment types and group size variables on the test results variable [F (1, 293) = 78.691, $p < 0.05$] (Table 2). To analyse this effect, we conducted t-tests for paired and independent samples. The results are presented below.

- For subjects in the large group, significant differences were observed between the grades from hetero-assessment and those from self-assessment [t (105) = 8.421, $p < 0.05$], with self-assessment grades being superior to those from hetero-assessment.
- For subjects in the small group, there were no significant differences between the grades from hetero-assessment and those from self-assessment [t (188) = 1.869, $p = 0.063$].
- Regarding the results obtained through hetero-assessment**,** significant differences were noted in terms of group size [t (293) = 3.997, $p < 0.05$]; grades taken by the students from the large group were lower than those of the students from the small group.

- In contrast, self-assessment results showed no significant differences based on group size [t (293) = 1.103, p = 0.271].

The following is a graphical representation of the interaction effect.

Figure 3. Influence of the Variables Assessment Types and Group Size on the Test Results



Thus, for students in the large group, the self-assessment results were significantly better than those from hetero-assessment. Conversely, self-assessment results of the students in the small group were not significantly better than those from hetero-assessment. These results may stem from the fact that students in smaller groups tend to be more responsible and demonstrate better academic performances. The students who are part of large groups tend to make more modifications, as the teacher's attention cannot be equally distributed among all students, making it easier to make changes in assessment tests. Individual responsibility tends to diminish with larger groups. It is thus advisable to conduct assessments in smaller groups to enhance responsibility for learning and evaluation. In universities, examinations are typically conducted with smaller groups to accurately reflect students' competency development.

In the case of hetero-assessment, students in the smaller group tended to have better results than those in the larger group. However, self-assessment results did not significantly differ based on group size.

Hypothesis 2. Students' results in the rubrics test are influenced by assessment types and their academic achievement levels.

To verify this hypothesis, we applied ANOVA multiple variance analysis, with assessment types and academic achievement levels as independent variables and test results as a dependent variable.

The statistical analysis revealed a significant main effect of assessment types on test results [$F (1, 289) = 53.607$, $p < 0.05$] and a main effect of academic achievement levels on test results [$F (1, 289) = 77.286$, $p < 0.05$]. There is also an effect of assessment types and academic achievement levels on test results variable [$F (1, 289) = 12.183$, $p < 0.05$] (Table 3). Table 3 presents the averages, standard deviations and subject samples, for the scores regarding the test results variable in terms of assessment types and performance levels.

Table 3. Averages, Standard Deviations and Subject Samples for the Scores Regarding the Test Results Variable in terms of the Assessment Types and Performance Levels

		Assessment types					
		Hetero-assessment			Self-assessment		
		M	SD	N	M	SD	n
Levels of achievement	Low	3.52	0.89	146	3.86	0.73	146
	high	4.31	0.63	145	4.43	0.53	145
Total		3.92	0.89	291	4.15	0.70	291

Table 4. The Results of the ANOVA Multiple Variance Analysis for Effect of the Assessment Types and Performance levels on the Test Results Variable

Variation source	SS	Df	MS	F	P
Assessment types (FE)	7.828	1	7.828	53.607	0.000
Levels of achievements (NP)	67.215	1	67.215	77.286	0.000
FE x NP	1.779	1	1.779	12.183	0.001
Residual (error) (FE)	42.204	289	0.146		
Residual (error) (NP)	251.339	289	0.870		

Main effect of the assessment types variable on the test results variable. The assessment types variable demonstrates a significant main effect on the test results variable [$F (1, 289) = 53.607$, $p < 0.05$] (Table 4). This effect was analysed in Hypothesis 1.

Main effect of the performance level variable on the test results variable. A significant main effect of performance levels on grade modification was observed [$F (1, 289) = 77.286$, $p < 0.05$] (Table 4). To verify this effect, we applied the Bonferroni t-test. The results indicate that

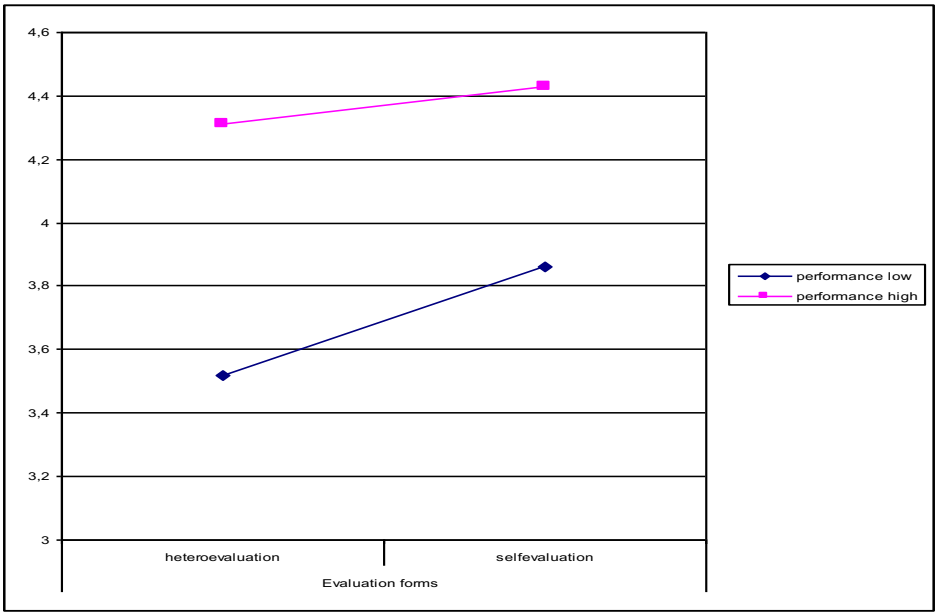
high-achieving students consistently outperform their low-achieving counterparts, regardless of the assessment method employed [Bonferroni $t = 8.831$, $p < 0.05$]. This outcome was anticipated, as students with excellent grades tend to achieve similar results in this test and receive high scores both in teacher's assessment and self-assessments.

Interaction effect of the assessment types and performance level variables on the test results variable. There is a significant interaction effect of the assessment types and performance level variables on the test results variable [$F (1, 289) = 12.183$, $p < 0.05$] (Table 4). To analyse this effect, we conducted t-tests for independent samples. The results are presented below.

- For students with low performance levels, significant differences were observed in test results concerning assessment types [$t (145) = 6.411$, $p < 0.05$], with self-assessment results being significantly better than those from hetero-assessment.
- For students with high performance levels, significant differences were also noted regarding assessment types [$t (144) = 3.583$, $p < 0.05$], with self-assessment results being significantly superior to those from hetero-assessment.
- Regarding hetero-assessment results, significant differences were found based on performance levels [$t (289) = 8.697$, $p < 0.05$], with lower test results for students with low academic achievement compared with those with high academic performance levels.
- Similarly, for self-assessment results, significant differences were noted based on performance levels [$t (289) = 7.554$, $p < 0.05$], with lower test results for students with poor academic performance compared with those with better academic performance levels. Therefore, even though the students with low academic performance tended to improve their grades more through the self-assessment process contrary to those with high performance levels, the test results from the self-assessment of the students demonstrating low academic performance remain significantly lower than those of their higher-achieving peers.

The below graph illustrates the interaction effect (Figure 4).

Figure 4. Influence of Assessment Types and Performance Levels on Test Results



Conclusions and Discussions: Educational benefits?

The present study analysed the extent to which the use of two different assessment methods—self-assessment and hetero-assessment—yield similar results for students undertaking the rubrics test and whether these results are influenced by the students' group size and academic performance levels. The results support the hypothesis. The students' self-assessment results were significantly better than those derived from hetero-assessment. Consequently, even in the presence of teachers, students tend to inflate their own grades (Panadero et al., 2013; Panadero et al., 2017). These results align with previous studies indicating that students often value their results more highly than their teachers, with a difference of nearly half a standard deviation (Falchikov & Boud, 2006).

The differences between the test results from the two assessment methods (hetero- and self-assessment) are the same only when students are part of large study groups; this is not the case for smaller groups. Meanwhile, the hetero-assessment results are superior for students in small groups compared with those who belong to large groups. This means that those in small groups had better performances than those from large groups. The differences appeared due to not only the higher responsibility of those from the small group towards their learning process but also the lack of

responsibility of those from the large group who had better grades in self-assessment than those assigned by teachers. Thus, 'a self-assessment culture' should be created to familiarise students with accurately assessing their own work, initially under teachers' guidance. This will enable them to develop a correct self-image and take responsibility for their results (Kearney, 2013; Mabe & West, 1982; Panadero et al., 2013; Panadero et al., 2017). Teachers can act as effective coaches, empowering students with greater responsibility for their self-training (Noonan & Duncan, 2005). Sullivan & Hall (1997) discuss methods for integrating self-assessment into training, including introducing schemes that provide opportunities for discussion about assessment and self-assessment, developing a course philosophy (self-assessment strategies proposed together with the course philosophy, objectives, contents, explanation, etc.), providing training in self-assessment (proposing numerous peer-assessment and self-assessment exercises) (Panadero et al., 2013), developing a marking guide (for a deeper understanding of the assessment process), analysing student insights (encouraging students to reflect on their strengths and weaknesses through journals or similar metacognitive activities) and evaluating the impact of self-assessment on their learning activities.

Our study also demonstrated that, regardless of students' academic performance levels, self-assessment results are significantly higher than those from hetero-assessment. Additionally, students with a high-performance level will score well in the test, whether self- or hetero-assessment is used. Therefore, unlike students with lower academic performance who overestimate their grades more in self-evaluation, high-performing students exhibit less overestimation. Nevertheless, the test results for low-performing students in self-assessment remain significantly lower than those of their high-performing counterparts. These findings corroborate previous research highlighting the tendency of lower-performing students to overestimate their abilities compared with their higher-performing peers (Panadero et al., 2013; Panadero et al., 2016; Topping, 2003). This phenomenon is influenced by the development of self-image and self-presentation. We learnt to distinguish between 'assessing oneself for oneself and assessing oneself for others' (Brown et al., 2011).

Similar results were observed regarding self-assessment accuracy and student performance. Research indicates that the better a student's performance, the more accurate their self-assessment, thereby reducing the disparity between self-assessment grades and teacher grades (Clipa et al., 2011). Similarly, statistically significant differences were also found between age and self-assessment accuracy, with teachers' grades aligning more closely

with those of students (Clipa et al., 2011; Lindblom-Ylanne et al., 2006; Orsmond et al., 2000; Panadero et al., 2016; Toping, 2003).

Teachers with high self-assessment accuracy are likely to exhibit greater efficacy, as they can establish realistic expectations based on past experiences. The teacher's efficacy significantly influences student performance and their belief in students' learning potential. Consequently, teachers who maintain high expectations regarding their ability to teach produce higher student achievement (Hosein & Harle, 2018; Panadero et al., 2013; Panadero et al., 2016; Ross & Bruce, 2007), and they have better capacity to adapt to the workplace and performance (Klassen & Chiu, 2011). It is crucial to incorporate mechanisms in teacher training that promote responsible learning training, enabling students to reflect on their methods for enhancing professional efficacy and learning autonomy (McCarthy et al., 2018; Panadero et al., 2013; Ramsden, 2003; Taras, 2016).

Future teachers must recognise the utility and necessity of self-assessment in developing metacognitive abilities, deep learning and personal responsibility. Self-assessment serves as a constructive tool (Clipa, 2015; Lowell & Ashby, 2018; Panadero et al., 2013; Panadero et al., 2015; Ross & Bruce, 2007), enabling individuals to assess themselves for their own benefit rather than for others (Brown et al., 2011).

Limitations and Future Directions

One limitation of this study pertains to the preparation of the rubrics test. Designed to assess abilities, it could have included subjective items. A qualitative study on students' self-assessment would have been valuable. We chose the objective and semi-objective items because their modifications during self-assessments were easier to track and quantify, and well-defined criteria were necessary for assessment accuracy.

Another limitation is that the investigated sample should have been larger and more diverse, encompassing various study types and academic centres both nationally and internationally. We could not organise a broader sample due to the need for teachers' consent and availability throughout the process, as well as the requirement for rapid implementation to avoid 'contamination' of subjects by revealing the procedure and undermining the study's objectives. We aim to carry out the study again with a well-built sample, allowing for comparisons with the current results. Validation of the findings could be achieved through alternative methods, such as filming student behaviour during tests or employing other research techniques. Future studies should include other factors to explore self-assessment accuracy and its connections with other personality traits.

In terms of future research directions, we intend to investigate other aspects of academic self-assessment. One avenue would be to explore the development of this 'self-assessment culture' during academic instruction. Self-assessment could serve as a formative objective for teachers (Arrendondo & Diago, 2005a,b). By promoting self-assessment, teachers can cultivate assessors responsible for their own learning and the authenticity and accuracy of their assessment (Boud & Falchikov, 2006; Day et al., 2018; Milne, 2008; Bannink, 2010; Panadero et al., 2017). 'If you want to change a student's learning, you should change the assessment methods' (Brown et al., 2011). We also seek to examine the links and influences between different assessment types and self-assessment across various academic years and curricula.

Another direction could be data correlation obtained from test information, which investigates students' character using various personality assessment questionnaires. It was noticed that self-assessment fosters moral development and personal autonomy, shaping individuals' conscience and honesty based on the feedback received (Gibbs & Coffey, 2004; Stufflebeam & Wingate, 2005).

A third direction could focus on investigating students' perceptions of self-assessment and its impact on their overall average. This is particularly relevant as self-assessment stimulates metacognitive abilities (Taras, 2016), deep-level learning and critical thinking (Lowell & Ashby, 2018; Struyven et al., 2008) while also fostering positive attitudes towards other assessment methods (Kearney, 2013). Students' attitudes towards self-assessment and self-grading were analysed by many authors, with opinions ranging from negative (due to perceived lack of objectivity, deep understanding and maturity on the part of the students) to positive, appreciating the objectivity and the opportunity to learn from their mistakes (Sadler & Good, 2006).

We aspire for students to take responsibility for their own learning and development. It would be beneficial for teachers to implement various assessment methods (including peer-assessment and self-assessment), clarify marking criteria, analyse students' results and promote learning and professional development through self-assessment mechanisms.

References

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. J. Andrade & G. J. Cizek (Eds.), *Handbook of Formative assessment*. Routledge. https://digitalcommons.lib.uconn.edu/cgi/viewcontent.cgi?article=1007&context=nera_2010

Arredondo, C. S., & Diago, C. J. (2005a). *Formación del profesorado en educación superior: didactica y curriculum (1st volume)* McGraw-Hill.

Arredondo, C. S., & Diago, C. J. (2005b). *Formación del profesorado en educación superior: desarrollo curricular y evaluación (2nd volume)*. McGraw-Hill.

Bale, R., & Seabrook, M. (2021). *Introducing to university teaching*. Sage Publication.

Bannink, A. (2010). How to capture growth? – Video narratives as an instrument for assessment in teacher education. *Teaching and Teacher Education*, *25*(2), 244–250. https://doi.org/10.1016/j.tate.2008.11.009

Bein, K. (2004). *What the best college teachers do*. Harvard University Press.

Black, P., & William, D. (2009). Developing the theory of formative assessment. Educational assessment. *Education and Accountability*, *21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*(4), 399–413. https://doi.org/10.1080/02602930600679050

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*(1), 210–220. https://doi.org/10.1016/j.tate.2010.08.003

Carless, D. (2013). Sustainable feedback and the development of student self-evaluative capacities. In S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education: Developing dialogue with students* (pp. 113-122). Routledge.

Clipa, O. (2015). Roles and strategies of teacher evaluation: Teacher perspectives. *Procedia- Social and Behavioral Sciences, 180*, 916–923. http://www.sciencedirect.com/science/article/pii/S187704281501589X

Clipa, O., Ignat, A.-A., & Rusu, P. (2011). Relations of self-assessment accuracy with motivation level and metacognitive abilities in pre-service teacher training. *Procedia Social and Behavioral Sciences*, *30*, 883–888. http://www.sciencedirect.com/science/article/pii/S1877042811019963

Day, I., van Blakenstain, F. M., Westemberg, M., & Admiraal, W. F. (2018). Teacher and student perceptions of intermediate assessment in higher education, *Educational Studies, 44*(4), 449–467. https://doi.org/10.1080/03055698.2017.1382324

Gibbs, G., & Coffey, M. (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education*, *5*(1), 87–100. https://doi.org/10.1177/1469787404040463

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through pre-service teacher education. *Teaching and Teacher Education, 21*(6), 607–621. https://doi.org/10.1016/j.tate.2005.05.001

Hosein, A., & Harle, J. (2018). The relationship between students' prior mathematical attainment, knowledge and confidence on their self-assessment accuracy. *Studies in Educational Evaluation, 56*, 32–41. https://doi.org/10.1016/j.stueduc.2017.10.008

Ion, G., Díaz-Vicario, A., & Mercader, C. (2024). Making steps towards improved fairness in group work assessment: The role of students' self- and peer-assessment. *Active Learning in Higher Education, 25*(3), 425-437. https://doi.org/10.1177/14697874231154826

Kearney, S. (2013). Improving engagement: The use of 'authentic self-and peer-assessment for learning' to enhance the student learning experience. *Assessment & Evaluation in Higher Education, 38*(7), 875–891. https://doi.org/10.1080/02602938.2012.751963

Kitsantis, A., Reisner, R. A., & Doster, J. (2004). Developing self-regulated learners: Goal setting, self-evaluation, and organisational signals during acquisition of procedural skills. *The Journal of Experimental Education*, *72*(4), 269–288. https://doi.org/10.3200/JEXE.72.4.269-287

Klassen, R., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36*(2), 114–129. http://dx.doi.org/10.1016/j.cedpsych.2011.01.002

Knight, P. (2005). *El profesorado de educación superior. Formacion para la excelencia.* Narcea.

Lindblom-Ylanne, S., Pihlajamaki H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, *7*(1), 51–62. https://doi.org/10.1177/1469787406061148

Lowell, V. L., & Ashby, I. V. (2018). Supporting the development of collaboration and feedback skills in instructional designers. *Journal of Computing in Higher Education, 30*, 72-92. https://doi.org/10.1007/s12528-018-9170-8

Mabe, P., & West, S. (1982). Validity of self-evaluation of ability. *Journal of Applied Psychology, 67*(3), 280–296. https://doi.org/10.1037/0021-9010.67.3.280

McCarthy, K., Likens, A., Johnson, A., Guerrerro, T., & McNamara, D (2018). Metacognitive overload!: Positive and negative effects of metacognitive prompts in an intelligent tutoring system. *International Artificial Intelligence in Education Society, 28*, 420–438. https://doi.org/10.1007/s40593-018-0164-5

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education, 10*(2), 209–220. http://dx.doi.org/10.1080/0969594032000121289

Milne, C. (2009). Assessing self-evaluation in a science methods course: Power, agency, authority and learning. *Teaching and Teacher Education*, *25*(5), 758–766. https://doi.org/10.1016/j.tate.2008.11.008

Noonan, B., & Duncan, R. (2005). Peer and self-evaluation in high school. *Practical Assessment, Research & Evaluation*, *10*(17), 1–8. https://doi.org/10.7275/a166-vm41

Norton, L. (2007). Using assessment to promote quality learning in higher education. In A. Campbell & L. Norton (Eds.), *Learning, teaching and assessing in higher education: Developing reflective practice* (92–101). Learning Matters Ltd.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment & Evaluation in Higher Education, 25*, 23–28. https://doi.org/10.1080/02602930050025006

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. http://dx.doi.org/10.1016/j.edurev.2017.08.004

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551e576. http://dx.doi.org/10.14204/ejrep.30.12200

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

Pope, N. (2005). The impact of stress in self- and peer assessment. *Assessment & Evaluation in Higher Education*, *30*(1), 51–63. https://doi.org/10.1080/0260293042003243896

Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). Routledge Falmer.

Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education, 27*(2), 472–482. https://doi.org/10.1016/j.tate.2010.09.017

Remesal, A. (2021). Synchronous self-assessment. Assessment from the other side of the mirror. In *Assessment as learning. Maximising opportunities for student learning and achievement* (218–231). Routledge.

Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education, 23*(2), 146–159. https://doi.org/10.1016/j.tate.2006.04.035

Sadler, P. M., & Good, E. (2006). The impact of self and peer grading on student learning. *Educational Assessment, 11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1

Segers, M., Martens, R., & Van den Bossche, P. (2008). Understanding how a case-based assessment instrument influences student teachers' learning approaches. *Teaching and Teacher Education*, *24*(7), 1751–1764. https://doi.org/10.1016/j.tate.2008.02.022

Sridharan, B., Tai, J., & Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education, 77*(5), 853–870. https://doi.org/10.1007/s10734-018-0305-7

Struyven, K., Dochy, F., & Janssens, S. (2008). The effects of hands-on experience on students' preferences for assessment methods. *Journal of Teacher Education*, *59*(1), 69–88. https://doi.org/10.1177/0022487107311335

Stufflebeam, D. L., & Wingate, L. A. (2005). A self-assessment procedure for use in evaluation training. *American Journal of Evaluation*, *26*(4), 544–561. https://doi.org/10.1177/1098214005279730

Sullivan, K. & Hall, C. (1997) Introducing students to self-assessment. *Assessment & Evaluation in Higher Education*, *22*(3), 289-305, https://doi.org/10.1080/0260293970220303

Taras, M. (2010). Student self-assessment: Process and consequences. *Teaching in Higher Education*, *15*(2), 199–209. https://doi.org/10.1080/13562511003620027

Taras, M. (2016). Situating power potentials and dynamics of learners and tutors within self-assessment models, *Journal of Further and Higher Education, 40*(6), 846–863, https://doi.org/10.1080/0309877X.2014.1000283

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising New Modes of Assessment: In Search of Qualities and Standards* (Vol. 1). Springer. https://doi.org/10.1007/0-306-48125-1_4

Zeng, W., Huang, F., Yu, L., Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—A critical review of literature. *Educational Assessment, Evaluation and Accountability*, 30, 211-250. https://doi.org/10.1007/s11092-018-9281-9