MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2025, 21(9), em2694
ISSN:1305-8223 (online)
OPEN ACCESS **Research Paper** https://doi.org/10.29333/ejmste/16774

Predictive power of peer assessment: Correlations between evaluators' scores and academic success in a mathematics preparatory program

Yaniv Biton [1,2*]

[1] Shannan Academic College of Education, Haifa, ISRAEL
[2] Center for Educational Technology, Tel Aviv, ISRAEL

Received 15 May 2025 ▪ Accepted 16 July 2025

Abstract

This study examines the extent to which students' performance in peer assessment tasks predicts academic achievement in a mathematics preparatory program. Conducted at the Technion's Center for Pre-University Education, the study involved approximately 60 students participating in two intensive mathematics classes. A key innovation of this research lies in the structure of the peer assessment process: unlike conventional designs that provide expert-defined rubrics, this study empowered students to independently formulate their own evaluation criteria. Each peer group generated its own set of assessment standards based on internal consensus, without instructor intervention. Despite the variability in rubrics, the peer-generated scores showed statistically significant correlations with both interim and final examination results, as well as with teacher-assigned grades. These findings provide compelling evidence that peer assessment can yield reliable evaluative outcomes even when criteria are student-constructed. Moreover, stronger correlations were observed when tasks were aligned with the formal curriculum and administered later in the program, suggesting development in students' evaluative proficiency. The study contributes to the discourse on the reliability of peer assessment, challenges assumptions regarding the necessity of expert-driven rubrics and highlights the potential of student-authored frameworks to foster authentic engagement and reflective judgment in mathematical learning contexts.

Keywords: peer assessment, mathematics education, preparatory programs, academic achievement, evaluative judgment

LITERATURE REVIEW

Peer Assessment

The involvement of students in assessment processes in educational settings has gained increasing prominence in recent years. Involving students in peer assessment may serve formative or summative purposes. Formative peer assessment allows for early identification of errors and misconceptions. Furthermore, involving students in deciding on parameters for assessment–which are typically predefined and aligned with the curriculum goals–and then performing the assessment is a strategy that fosters greater understanding of teaching parameters and reflective learning, thus facilitating deeper understanding and knowledge retention.

Peer assessment is not universally defined. While the simple definition may be a process where individuals or groups rate their peers (Falchikov, 1995), a more detailed definition might be "an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status" (Topping, 1998, p. 250). In any case, peer assessment can position students as both assessors and learners in a collaborative framework and may focus on task outcomes and individual or collaborative efforts (de-Armas-González et al., 2023; Kollar & Fischer, 2010; Topping et al., 2020).

Peer assessment has been implemented across a variety of disciplines, including computer science (Venables & Summit, 2003), mathematics-teacher training (Zevenbergen, 2001), psychology (Smith et al.,

✉ yanivb@cet.ac.il **(*Correspondence)**

Contribution to the literature

- This study explores the potential of peer assessment scores—based on student-constructed evaluation criteria—as possible indicators of academic performance in mathematics.
- By focusing on student-generated rubrics rather than expert-designed ones, the research offers insights into the feasibility and educational value of more autonomous assessment processes.
- The findings may contribute to ongoing discussions on the reliability and pedagogical relevance of peer assessment in mathematics education.

2002), and economics (Brindley & Scoffield, 1998), highlighting the universality of its application across educational contexts.

Effectiveness of Peer Assessment

Carrying out peer assessment is an intellectually demanding task for students, however its overarching outcome is that it enhances the learning process and boosts motivation (Bedford & Legg, 2007; Biton, 2024; Black & Wiliam, 2006; Orsmond et al., 2000). For the learner, obtaining the perspectives of their peers regarding the topic also bestows a new way of understanding the material. Peer assessment supports cognitive and metacognitive development by providing students with immediate and detailed feedback at the time of undertaking the problem.

For the assessor, engaging in peer assessment requires them to thoroughly understand the task objectives and the criteria for success (Xiang et al., 2021; Yin et al., 2022), to engage in critical thinking and reflective practices (Zariski, 1996), and to gain insight into the thinking processes of their peers (Kollar & Fischer, 2010), thus fostering a greater sense of responsibility for their own learning. This dynamic also enables students to view the task from a teacher's perspective, which aids in grasping the intricacies of the evaluation process. It fosters interaction between students and their instructors due to the need to share evaluation strategies.

Recent studies have also pointed out how peer assessment exercises can develop critical problem-solving skills. For example, peer assessment activities that were introduced into courses for pre-service teachers enhanced the future teachers' mathematical understanding by exposing them to diverse problem-solving strategies, comprehensive evaluation skills, and deeper learning. This underscores the value of incorporating such practices into teacher-training programs (de-Armas-González et al., 2023).

In summary, peer assessment tasks can impact student learning across various educational contexts and, for the most part, result in significant improvement in student performance (Topping et al. [2020] showed an effect size of 0.291 standard deviation units over those who did not participate).

Challenges

Despite the potential benefits, peer assessment is not without its challenges. Peer bias, social dynamics, and students' limited assessment literacy may undermine reliability, particularly in non-anonymous or high-stakes settings. Lack of experience and familiarity with the process can influence their perceptions.

Additionally, the cognitive demand of interpreting complex rubrics can lead to variability in judgment. Potential exposure (of one's lack of understanding), fear of causing offense (by assigning a low score), or lack of trust in the fairness of the process may dissuade students from participating (de-Armas-González et al., 2023; Kirkpatrick & Fuller, 1995). This suggests that adequate training and clear guidelines are necessary: experience and familiarity with peer assessment can improve attitudes towards the practice (Cheng & Warren, 1997; Wen & Tsai, 2006). Another challenge was revealed in a study by de-Armas-González et al. (2023) of pre-service secondary school mathematics teachers who assessed their peers' solutions to open-ended mathematical problems. The study revealed that the teachers were often influenced by their own problem-solving approaches. This can challenge the objectivity of evaluations.

To mitigate these challenges, scholars recommend combining training protocols, iterative practice, anonymized reviews when appropriate, and calibration activities (Double et al., 2020; Panadero, 2016). These strategies support a shift from intuitive to criterion-referenced evaluation, thereby enhancing the trustworthiness and instructional value of peer-generated scores. Factors that influence effectiveness include the use of computer-mediated assessment methods and rater training (Topping et al., 2020). Studies have shown that explicit training enhances both the accuracy and educational benefit of peer assessment. Additionally, digitally based peer assessment methods were found to be more effective than paper-based methods, as it can facilitate clearer, more structured feedback (Topping et al., 2020).

Shaping Students Attitudes

Students may show ambivalence to the idea of peer assessment, especially if they are not sure of the process or feel that they lack sufficient understanding of the material to be evaluated. However, studies have shown that students who initially express reservations about peer assessment become more positive after experiencing it firsthand. Context is also important as was shown in a study involving computer science undergraduates (Venables & Summit, 2003). Here, students evaluated their peers' literature reviews and thus learned to appreciate the diverse perspectives of their peers and the value of varying levels of analysis and critique. This aligns with the principles of reflective learning, where the act of evaluation (of oneself or of others) deepened understanding of the subject.

Training and experience improve students' confidence, comfort levels and perceptions of fairness when assessing their peers, as was shown in a study of first-year electrical-engineering students, especially when it followed structured training and was considered a valid component of learning (Cheng & Warren, 1997). Another study with university students and using both traditional and online formats revealed that they generally held positive views about peer assessment, appreciating it as a tool to enhance interaction and understanding between peers and instructors (Wen & Tsai, 2006). Interestingly, male students seemed more positive about the process than female. However, while they recognized the logistical utility of online peer assessment, they did not necessarily see it as a beneficial learning tool. This study highlighted the need for clear guidelines and effective implementation strategies to ensure reliability and validity in the peer assessment processes.

Students Constructing Evaluation Criteria

One factor that allows peer assessment to become a powerful learning tool is having students actively involved in constructing the evaluation criteria themselves (Biton, 2025). In that study, Biton (2025) showed that when students collaboratively define what constitutes "quality" in mathematical problem solving instead of relying on externally imposed rubrics, they engaged in meaningful dialogue and exhibit deeper engagement. In fact, the act of negotiating criteria fostered metacognitive reflection and accurate jargon. It helped shift classroom discourse from procedural to conceptual. Through evaluative thinking and collaborative reasoning, the peer assessment process became not only a tool for measuring performance, but also a catalyst for learning.

Reliability and Limitations of Peer Assessment

An overarching question that overshadows the possibility of imposing peer assessment on studies, especially if the assessment is to be used to compile a student's final grade, is whether it is reliable and fair, and devoid of any overt bias, particularly in disciplines that demand subjective evaluation such as mathematics or writing. For this reason, studies have been done in which peer-generated scores are compared with "professional" scoring to assess inter-rater reliability. The results indicate that proper scaffolding is a must, otherwise peer assessments can yield biased or inconsistent results (Topping, 1998; van Zundert et al., 2010). Li et al.'s (2020) meta-analysis also suggests that reliability is not inherent to peer assessment, but requires explicit criteria, structured task design, and targeted assessor training, which significantly enhances judgment accuracy and instructional value of peer feedback.

These claims were further substantiated in a study comparing two peer-assessed tasks in an undergraduate accounting course. One task required subjective judgment and the other had more objective criteria. As could be predicted, the task with higher objectivity yielded substantially stronger inter-rater reliability (ICC = 0.79) compared to the subjective task (ICC = 0.50). Averaging the scores given by two peer assessors of the subjective context improved reliability, supporting the argument for multi-rater strategies. The students also reported greater comfort and confidence when given clear rubrics and well-defined expectations, thereby reinforcing the importance of designing assessment tasks that reduce ambiguity and guide evaluators through structured processes (Kruger, 2025).

As mentioned above, studies indicate that, despite initial hesitation, students generally develop favorable attitudes towards peer assessment activities (Cheng & Warren, 1997; Strachan & Wilcox, 1996; Venables & Summit, 2003; Wen & Tsai, 2006). To increase the development of peer assessment exercises as an instructional tool and to solidify its role as a meaningful, reliable, and equitable practice in modern education, educators need to explore in greater depth how student attitudes evolve, what factors contribute to students' acceptance or resistance of peer evaluation, the long-term effects of peer assessment on learning, and which methodological frameworks can better link the assessment process to specific learning outcomes (van Zundert et al., 2010).

RESEARCH OBJECTIVE AND METHODOLOGY

This study aimed to investigate the extent to which students' performance in peer assessment tasks correlates with their academic success in a mathematics preparatory program. In other words, the research sought to explore whether the grades assigned through peer evaluation corroborated with the instructor's "official" grades and thus could serve as valid indicators or predictors of students' achievements in interim and final mathematics examinations.

The main research question guiding this inquiry was: To what extent do preparatory-school students' scores in

$f(x) = \frac{\cos x}{\sin^2 x}$ is a function at $-\frac{\pi}{2} \leq x < 0$.

a. Find all the asymptotes, if any, which are parallel to the axes.
b. Find the points of intersection, if any, of the function with the axes.
c. Prove that the function is increasing.
d. Find the equation of the tangent to the graph of $f(x)$ at its point of intersection with x-axis.
e. Find the area of the figure constrained by $f(x)$, the above tangent, and $x = -0.5$.

Figure 1. Sample problem (Source: Field study)

peer assessment tasks corroborate with their formal mathematics assessments?

Research Locale and Participants

The study took place at the Center for Pre-University Education of the Technion–Israel Institute of Technology. The participants comprised approximately 30 students in each of two classes that taught an intensive, three -semesters preparatory course in mathematics to bring the students up to level-five-matriculation proficiency. Such preparatory courses are offered in many universities worldwide to close any gaps between students' achievements in high school and prerequisites for academic study. One teacher taught both classes.

Typically, the students' progress is assessed systematically using traditional methods, i.e., exams, but during this course, students worked in groups and were asked to assess their peers' work. Six such peer assessment activities (three from each class) were recorded using a video camera and audio recordings.

Peer Assessment in the Current Study

Each of the six peer assessment activities was 90 minutes long. In the first stage (25 minutes), the students were asked to individually solve a mathematical problem randomly chosen from the course textbook (see **Figure 1** for a sample problem).

In the second stage (15 minutes), the students were randomly divided into groups of two or three and asked to formulate criteria for the evaluation of the other students' solutions to the problem and to weigh the criteria according to what they believed were their relative importance. Each group was given a four-page booklet in which they were to write their evaluation criteria on the first page and use the remaining pages to comment and score the other students' solutions.

For the third stage (50 minutes), each group received the solutions of three other students for assessment. This was not anonymous (i.e., all the students knew whose solutions they were reviewing and who was checking theirs). Inter-group conversation was allowed and encouraged at all stages of the activity. After evaluation, the annotated work was returned to the assessees who were given the opportunity to respond if they wished to explain their work or question the assessment.

The students were advised that this was merely an exercise, and their assessment would not affect their peers' final marks. This was to prevent them from feeling inhibited about giving a low mark or negative comment.

Designing Peer Assessment Criteria For This Study

The way the students designed their peer assessment criteria was affected by the way they engaged in social interaction, took responsibility, and depended on each other. The nature of the activity meant that the learners were required to provide clear explanations and ask questions alongside their role as learners. That is, they needed to become fully familiar with the correct solution to be able to assess their peers' work appropriately. This evoked discussion and active listening throughout the phases of the assignment.

As noted, the students in each group were chosen at random, which led to a heterogeneous group. This allowed stronger students to teach, help, and enrich the weaker. The questions or misunderstandings shown by the weaker students enlightened the group regarding possible misunderstandings. As a whole, therefore, this environment was conducive to active involvement of all. Each student was able to show their skills and contribute knowledge to the group. This social interaction in collaborative learning, where students are exposed to multiple perspectives and there is an exchange of knowledge among members of the group is known to positively impact learners' achievements (Kollar & Fischer, 2010; Leikin & Zaslavsky, 1999; Radford, 2011).

The first stage, in which the students solves the assignment on their own, may be considered a kind of "test writing," and later contributed to the discourse since each described their individual way of solving the problem and a description of the difficulties they encountered. The second stage enabled "brainstorming" the criteria for assessing the solutions as well as discussing the "correctness" of each of the solutions.

As part of the exploratory study, various options for conducting the peer assessment tasks were explored. In one, pairs were formed and one student solved the problem and the other assessed it. However, we noted that students who failed to solve the problem on their own refrained from assessing their partner's, claiming that they could not assess a solution they had failed to find themselves. In another, a group were formed and they evaluated each of the solutions together. In this case, we noted that students who had difficulty solving the assignment individually still took an active part in the assessment phase, e.g., asking questions, attempting to probe deeper and understand the phases of the solution offered, taking a stand on its validity, and even voicing their opinions concerning the grade they had to give. This was encouraging since the phenomenon of social avoidance during collaborative learning is well documented in the literature (Kerr & Tindale, 2004) but was not observed in the present study.

Another issue that concerned us was how students would react given that the peer assessment was done openly. In such cases, fear of exposure and personal hurt, fear of hurting one's peers, and difficulty of being objective are well-known phenomena recognized in the literature. However, we made a point to assure the students that their evaluations would not affect the formal grade, thus reducing these concerns. As a result, they spontaneously initiated assessor-assessee meetings.

Challenges in Integrating Peer Assessment Activities

Alongside the advantages of collaborative learning are the three challenges described below. These may be overcome by designing peer assessment activities as in the current study.

Group pressure

As part of a collaborative group, students need to feel that they are members of a team. However, this may lead them to feel pressured by the group and as a result, refrain from expressing their true thoughts.

Advanced students

A small collaborative group reduces the students' average to that of the group. The need for conformity in the group harms advanced students and limits their thinking skills and creativity. However, we observed that, similar to other studies on collaborative learning that point to group heterogeneity as an effective tool for enhancing collaborative learning (Davidson, 1990; Johnson & Johnson, 1985), the heterogeneity paid off. When groups were composed of students who succeeded in solving the assignment and those who could not, the successful students (including those who only partly succeeded) very quickly took on the role of "teacher" within the group and started to answer their peers' questions concerning the solutions. This situation, where one member of the group explains the material to their peers, is well known and supported in the literature. Not only does it provide the less successful members with better understanding, but it is also an advantage for the "teacher," as the act of explaining may shed light on aspects which they themselves had been previously unaware of and they also obtain a better understanding of what was unclear to their peers (Webb, 1991).

Less motivated students

In a collaborative setting, weak, lazy, or uninterested students take advantage of the collaboration with more diligent ones, thus relinquishing their responsibility for their own learning. This is a phenomenon that is hard to mitigate, however, in this study we observed that, even the less motivated students took part in the discussions, perhaps because the groups consisted of only a few individuals and thus "sitting out" was more noticeable.

Research Tools and Data Analysis

Data analysis employed a quantitative approach grounded in advanced statistical methods aimed at examining the correlations between students' performance in peer assessment tasks and their academic achievements in both interim and final mathematics evaluations within the preparatory program. The statistical analysis aimed to examine whether students' performance in peer assessment tasks is meaningfully associated with their academic outcomes in both interim and final mathematics evaluations. This guiding assumption framed the correlation analyses that follow and shaped the interpretation of their results.

Correlation analysis

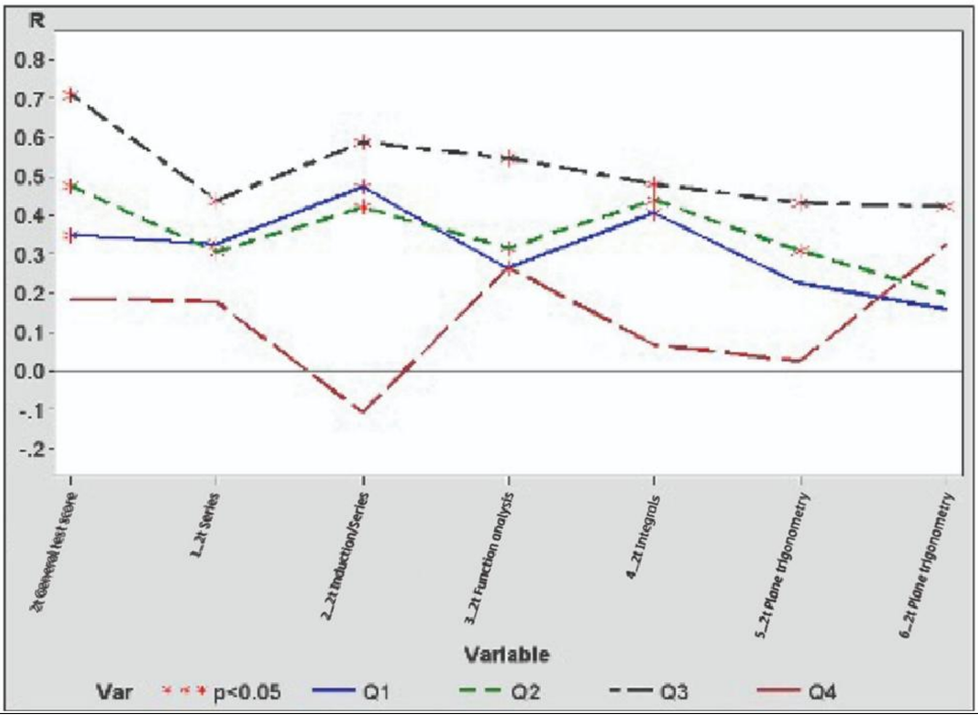
To assess the strength and direction of associations between variables, Spearman's rank-order correlation coefficients (ρ) were calculated. This non-parametric method was selected based on its suitability for ordinal data and its robustness in the presence of non-normally distributed variables. Correlation coefficients were computed for each pair of variables: peer assessment scores (Q1-Q4) against individual exam question scores, teacher evaluations (algebra and geometry), and final mathematics grades. Statistical significance was determined at a conventional threshold of $p < 0.05$; significant correlations are marked with an asterisk (*) in the graphical representations below.

Content and temporal alignment

Subsequent analyses examined whether significant correlations aligned topically and temporally. Specifically, we investigated whether peer assessment tasks that addressed particular mathematical domains (e.g., trigonometry and calculus) correlated most strongly with test items covering the same content areas and studied during the same trimester. This alignment allowed identifying content-specific predictive validity within the peer assessments.

Sample Size Considerations

Attention was given to the sample size (N) associated with each correlation coefficient, as test structures varied–some items were compulsory across all students



Peer Evaluation Questions (Q)	2nd trimester test questions (2t)	
Q1 Extreme value problems in spatial geometry	2t General test score	4_2t Integrals
Q2 Analysis of functions + integrals	1_2t Series	5_2t Plane trigonometry + equations
Q3 Plane trigonometry	2_2t Induction/Series	6_2t Plane trigonometry
Q4 Spherical trigonometry	3_2t Function analysis	

Figure 2. Correlation between students' grades on peer assessment tasks and questions in the second trimester test (Source: Authors' own elaboration)

while others were elective or class-specific. This variability influenced the number of valid data points available for each comparison. Thus, the value of N is reported alongside each correlation to contextualize its stability and generalizability.

Pedagogical Interpretation and Comparative Perspective

Beyond statistical association, the findings were interpreted based on the qualitative characteristics of each peer assessment task, particularly the student-defined evaluation criteria and the level of assessor autonomy. The strength and significance of correlations were discussed in relation to the pedagogical rigor and evaluative coherence of the student-generated rubrics. Findings were further contextualized using relevant literature, particularly the work of van Zundert et al. (2010), which emphasizes the role of training and repeated practice in improving the reliability and validity of peer assessment.

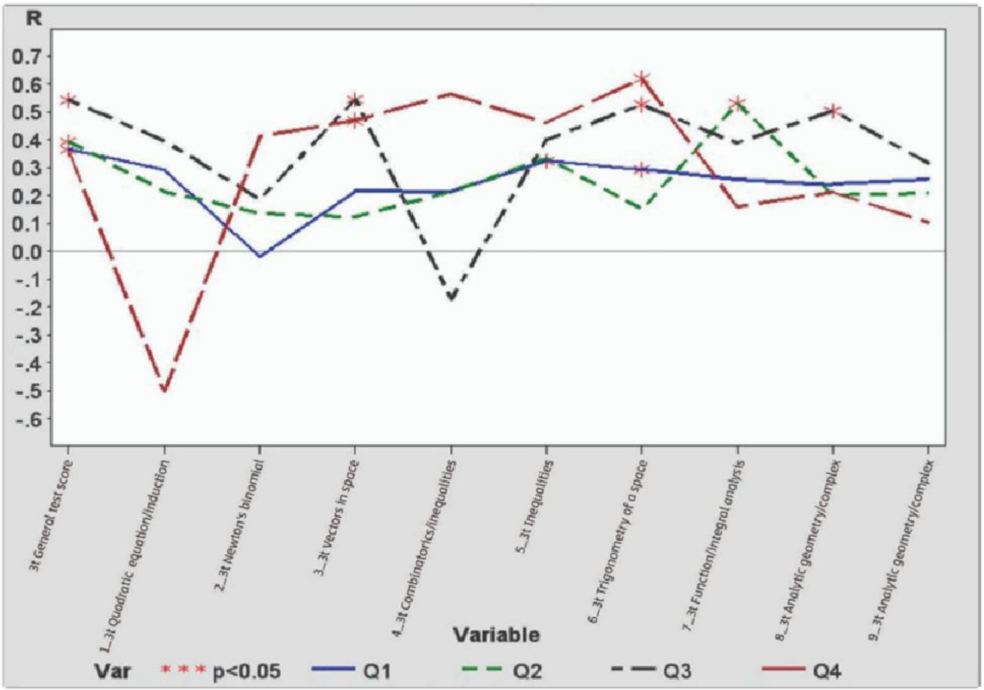
RESULTS

Correlation Between Students' Grades in the Peer Assessment Tasks and Their Success During and at the End of Their Studies in the Preparatory School

Graphs were plotted to illustrate the correlation between students' grades in the actual preparatory-school tests and peer-assigned grades. **Figure 2** plots the correlation coefficient (R, vertical axis) for each of the second trimester test questions (horizonal axis). A significant correlation ($p < 0.05$) between two variables is indicated in the graph by an asterisk (*).

Figure 2 demonstrates that there is a clear correlation between the scores of peer assessment questions Q1, Q2, Q3, and the students' overall grades in the second trimester test. It also shows that questions Q2 and Q3 have strong correlations ($R > 0.4$) with the final grade. The topics of these three questions are studied during the second trimester. A significant, strong correlation is also noticeable between the students' grades in the peer assessment tasks and the grades in the final exam questions on the same subjects. An example of this type of connection can be seen in 4_2t alongside Q2, both of which concern finding a tangent equation with an integral of a trigonometric function. Another example can be seen in questions 5_2t and 6_2t alongside Q3, both of which deal with plane trigonometry. Strong–and even significant–connections were also found between questions not on the same topic, such as question 3_2t, which deals with the analysis of a rational function, paired with question Q3 (plane trigonometry).

An interesting finding was that question Q4, about spherical trigonometry, was not found to have any clear connection with any of the test questions. Question Q4 was the first in the order of appearance in the peer



Peer Evaluation Questions (Q) (**Appendix C**)	3rd trimester test questions (3t) (**Appendix D**)	
Q1 Extreme value problems in spatial geometry	3t General test score	5_3t Logarithmic and trigonometric inequalities
Q2 Analysis of functions + integrals	1_3t Quadratic equation analysis + induction	6_3t Trigonometry of a space
Q3 Plane trigonometry	2_3t Newton's binomial	7_3t Function + integral analysis
Q4 Spherical trigonometry	3_3t Vectors in space	8_3t Analytic geometry + composites
	4_3t Combinatorics + trigonometric equations	9_3t Composites

Figure 3. Correlation between students' grades on the peer assessment tasks and on the questions of the third trimester test (Source: Authors' own elaboration)

evaluation. It is possible that the evaluators' lack of experience in evaluating and determining the criteria affected the reliability of their evaluations reflected in this finding. Two additional reasons may be the extreme difficulty of the question (reported by the students while performing the task) and the fact that the topic of this problem is in the third trimester test.

Similar statistical tests were also conducted for the final exam, the results of which are shown in **Figure 3**.

Figure 3, similar to the findings from the second trimester test (**Figure 2**), illustrates that there are significant correlations between the scores of peer assessment tasks Q1, Q2, Q3, and the students' overall scores in the final exam for questions on the same subject. An example of this type of connection can be seen in the question dealing with spatial geometry (6_3t) and questions Q1, Q3, and Q4, which deal with similar topics: spatial geometry (extreme value problem) and trigonometry of a plane and of a volume. Another example can be seen in the question on analysis of a function and calculation of an area using an integral (7_3t) together with question Q2, which also concerns analysis of a trigonometric function and calculation of an area. Strong and even significant correlations were also found between questions that are not on the same topic, such as question 3_3t, dealing with vectors in space, and questions Q3 and Q4, which involve trigonometry within a plane and within a volume.

Note that **Figure 3** shows significant connections between the scores on question Q4 and those of the final exam (as opposed to Q4 and the 2nd trimester results). The reason for this is probably because the topic of this question, spherical trigonometry, was studied during the third trimester.

Correlation Between Algebra and Geometry Teachers' Grades and Peer Evaluation

In addition to examining the correlation between the students' scores in the peer assessment tasks and their grades in the course tests, the correlation of the peer assessment scores with the teacher's grades and the final grade in mathematics were also examined. The study of mathematics in the preparatory school is divided into two subjects: algebra and geometry, with the teacher teaching these subjects to the same class.

Table 1 shows the correlation between the various teachers' grades and the final math grade in the preparatory school as well as the students' grades in each of the peer assessment tasks.

Table 1. Correlation between students' grades on peer assessment tasks and their grades in the teacher's assessment

	Algebra teacher	Geometry teacher	Final mathematics score
	Spearman (Prob > |r| under H0: Rho = 0)	Spearman (Prob > |r| under H0: Rho = 0)	Spearman (Prob > |r| under H0: Rho = 0)
Q1. Extreme value problems in spatial geometry	0.39 (0.002) Number of observations: 59	0.46 (0.0002) Number of observations: 60	0.42 (0.0007) Number of observations: 60
Q2. Analysis of functions + integrals	0.47 (0.0017) Number of observations: 42	0.41 (0.0073) Number of observations: 42	0.50 (0.0008) Number of observations: 42
Q3. Plane trigonometry	0.41 (0.0396) Number of observations: 25	0.61 (0.0011) Number of observations: 25	0.54 (0.0054) Number of observations: 25
Q4. Spherical trigonometry	0.34 (0.118) Number of observations: 322	0.51 (0.0127) Number of observations: 23	0.40 (0.0619) Number of observations: 23

Note. Highlighted in blue and green indicates a significant relationship ($p < 0.05$)

The cells highlighted in blue indicate moderate correlations, while the cells highlighted in green indicate strong correlations between teacher-assigned grades, final grades, and peer assessment scores.

Summary of the Correlation Between the Students' Grades in the Peer Assessment Tasks and Their Success During and at the End of Their Studies in the Preparatory School

The peer evaluation tasks in the proposed study were spread over the second and third trimesters. The subjects covered by tasks Q2 and Q3 were studied during the second trimester, while those of Q1 and Q4 were studied during the third trimester. The findings show clear correlations between the evaluations given in peer evaluation task and the grades of the same semester in which the subject was studied. Likewise, there are correlations–some of them even significant–between the task and items from the test on the same subject.

It is important to note that the current study does not examine the development of the assessors from task to task in terms of their prediction abilities. Indeed, note that in **Figure 2** we observe that task Q4, which was the first of the tasks the students attempted, did not have a significant correlation and even had low correlations (compared to the findings of the other questions) with each of the test items. It can be assumed that the reason for this is that, at that point, the students had not had any experience in choosing and ranking criteria. This finding is also reflected in the findings of van Zundert et al. (2010) who argue that student training and practice in assessment are likely to improve reliability and validity.

The correlations between the evaluations done by the students for the peer-assessment tasks and those given by the teachers for the test and exam questions may be due to a number of factors. First, the students worked in collaboration to rank the criteria for evaluating the tasks, which may have led them to results in keeping with what educational experts would also suggest.

The above findings imply that these relationships require much more study. Nevertheless, the presentation of the above correlations may contribute to understanding the criteria necessary for evaluation, the correct distribution, and students' success in understanding the solution they were required to evaluate. Note that the correlations obtained and presented in this study are similar to those reported in the literature review by van Zundert et al. (2010), which focuses on the reliability and validity of peer evaluation, among other things.

DISCUSSION

The present study contributes to the expanding body of research on the value of peer assessment by providing empirical evidence of its potential both as a pedagogical tool and as an evaluative strategy in mathematics education. While existing literature has long emphasized the formative value of peer assessment (Strijbos & Sluijsmans, 2010; Topping et al., 2020), the current findings reveal a particularly notable dimension: peer assessment activities that are entirely student-driven–where the evaluation criteria are developed by the students themselves–can still produce statistically significant correlations with formal academic outcomes.

This finding is striking for two reasons. First, it challenges a prevailing assumption that reliable assessment must be grounded in externally imposed rubrics or instructor-derived benchmarks (Falchikov, 1995; Saito & Fujita, 2004). Second, and more importantly, it underscores the unique methodological contribution of the current study: students were not only assessors, but also the architects of the criteria by which assessment occurred. Even though each peer group generated its own evaluation framework, the assessments they produced demonstrated reliability and alignment with other performance indicators. This suggests that the act of co-constructing criteria may in itself be a powerful mechanism for enhancing evaluative judgment and conceptual understanding.

The robustness of these findings aligns with Biton's (2025) earlier research, which demonstrated how peer assessment occasions deep mathematical discourse, facilitates metacognitive engagement, and promotes disciplinary language acquisition.

In addition, student reflections collected in the above mentioned (parallel) study (Biton, 2025) revealed that despite initial concerns regarding objectivity and fairness, students overwhelmingly recognized the educational value of peer assessment and even reported increased empathy toward teachers' evaluative roles. This supports the conclusion that, when supported with appropriate training and an environment free from formal grade consequences, peer assessment can enhance both the cognitive and affective dimensions of learning.

Nevertheless, the study's findings also reinforce established concerns about reliability. Consistent with prior meta-analytic reviews (Li et al., 2016), the present results suggest that the absence of assessor training and the influence of interpersonal dynamics (e.g., "friendship marking") remain potential limitations. While statistical correlations were observed, they may not generalize across all forms of peer evaluation, particularly when stakes are high or anonymity is not preserved.

In summary, the uniqueness of this study lies in its methodological innovation: allowing students to formulate their own criteria and demonstrating that this autonomy does not diminish, and may even enhance, the reliability of peer-assigned scores. This challenges assumptions in the literature and provides a basis for reconceptualizing reliability not solely as a matter of standardized rubrics, but as a function of authentic cognitive engagement in the evaluative process.

Limitations and Implications

While this study offers meaningful insight into the predictive value of peer assessment in mathematics education, several limitations must be acknowledged. First, the research was conducted within the context of a specific pre-university mathematics preparatory program, with instruction and assessment practices that may differ from those in high school or university-level courses. As such, the findings may not fully generalize to other educational settings or populations. In addition, the study did not examine the longitudinal development of students' evaluative abilities across multiple peer assessment cycles. Notably, the correlation between the first peer task (Q4) and exam performance was limited, possibly due to students' lack of prior experience with assessment criteria and calibration procedures. Finally, the peer assessments were not integrated into formal grading, which may have influenced the level of investment or seriousness students devoted to the task. Future studies could address these constraints by embedding peer assessment in high-stakes environments and tracking students' growth over time.

CONCLUSION

This study provides compelling evidence that peer assessment scores are positively correlated with students' academic performance in mathematics, particularly when the tasks are well-aligned with the content of formal evaluations. The findings demonstrate that when students engage in assessing their peers' work–especially in structured and thematically relevant tasks–they produce evaluations that are not only formative but also predictive. This supports the validity of peer assessment as a pedagogical and diagnostic tool. The results further highlight the importance of experience and guidance: tasks implemented later in the program yielded stronger correlations, suggesting a developmental trajectory in students' capacity to evaluate with accuracy. These insights reinforce the need to design peer assessment activities that incorporate structured training, transparent criteria, and reflection. By doing so, educators can harness peer assessment both to enhance learning and to generate meaningful insights into students' cognitive and disciplinary understanding.

Recommendations For Future Investigations

Building on the findings of this study, several avenues for further research are proposed. First, longitudinal studies could explore how students' evaluative judgment improves across multiple peer assessment cycles and whether this growth enhances their own learning outcomes over time. Second, future research should investigate the implementation of peer assessment in formal grading systems, particularly in high-stakes environments, to determine how this integration influences students' perceptions of fairness, accuracy, and motivation. Third, comparative studies across different mathematical domains–such as algebra, geometry, and calculus–could assess whether certain topics are more conducive to reliable peer evaluation. Additionally, cross-context research involving diverse educational systems, student populations, and institutional cultures would deepen our understanding of transferability and contextual factors. Finally, experimental studies that include assessor training protocols and calibration sessions could yield evidence-based guidelines for improving peer assessment reliability. These research directions would contribute to optimizing peer assessment as an evidence-based component of mathematics instruction.

Ethical statement: The author stated that the study was conducted in accordance with applicable ethical research standards, and that participants provided written informed consent after receiving an explanation of the study's objectives and procedures.

AI statement: The author stated that generative AI tools were used to suggest alternative phrasing when needed.

Declaration of interest: No conflict of interest is declared by the author.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the author.

REFERENCES

Bedford, S., & Legg, S. (2007). Formative peer and self-feedback as a catalyst for change within science teaching. *Chemistry Education Research and Practice, 8*(1), 80-92. https://doi.org/10.1039/B6RP90022D

Biton, Y. (2024). Learning mathematics through peer assessment: "How can we assess something that we ourselves don't know how to solve?" *Eurasia Journal of Mathematics, Science and Technology Education, 21*(1), Article em2557. https://doi.org/10.29333/ejmste/15794

Biton, Y. (2025). Student reflections on peer assessments: Benefits and challenges in a mathematics class. *Educational Process: International Journal, 14*, Article e2025003. https://doi.org/10.22521/edupij.2025.14.3

Black, P., & Wiliam, D. (2006). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (pp. 9-25). SAGE.

Brindley, C., & Scoffield, S. (1998). Peer assessment in undergraduate programs. *Teaching in Higher Education, 3*, 79-90. https://doi.org/10.1080/1356215980030106

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*, 233-239. https://doi.org/10.1080/03075079712331381064

Davidson, G. V. (1990). Matching learning styles with teaching styles: Is it a useful concept in education? *Performance and Instruction, 29*(4), 36-38. https://doi.org/10.1002/pfi.4160290410

de-Armas-González, P., Perdomo-Díaz, J., & Sosa-Martín, D. (2023). Peer assessment processes in a problem-solving activity with future teachers. *Eurasia Journal of Mathematics, Science and Technology Education, 19*(4), Article em2245. https://doi.org/10.29333/ejmste/13057

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*, 481-509. https://doi.org/10.1007/s10648-019-09510-3

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*, 175-187. https://doi.org/10.1080/1355800950320212

Johnson, R. T., & Johnson, D. W. (1985). Student-student interaction: Ignored but powerful. *Journal of Teacher Education, 36*(4), 22-26. https://doi.org/10.1177/002248718503600406

Kerr, N., & Tindale, R. (2004). Group performance and decision making. *Annual Review of Psychology, 55*, 623-655. https://doi.org/10.1146/annurev.psych.55.090902.142009

Kirkpatrick, D., & Fuller, R. (1995). The challenge of peer assessment. In L. Summers (Ed.), *Proceedings of the 4th Annual Teaching Learning Forum* (pp. 146-149). Edith Cowan University.

Kollar, I., & Fischer, F. (2010). Commentary: Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*, 344-348. https://doi.org/10.1016/j.learninstruc.2009.08.005

Kruger, S. J. (2025). The reliability of peer assessment in a final-year information systems course. *South African Journal of Higher Education, 39*(1), 85-104. https://doi.org/10.20853/39-1-6740

Leikin, R., & Zaslavsky, O. (1999). Connecting research with practice: Cooperative learning in mathematics. *Mathematics Teacher, 92*, 240-246. https://doi.org/10.5951/MT.92.3.0240

Li, L., Liu, X., & Steckelberg, A. L. (2016). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology, 41*(3), 525-536. https://doi.org/10.1111/j.1467-8535.2009.00968.x

Li, L., Xiong, Y., Zang, X., Kornhaber, M., Lyu, Y., Chung, K. S., & Suen, H. K. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193-211. https://doi.org/10.1080/02602938.2019.1620679

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*, 23-38. https://doi.org/10.1080/02602930050025006

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. *Assessment & Evaluation in Higher Education, 41*(2), 302-318.

Radford, L. (2011). Classroom interaction: Why is it good, really? *Educational Studies in Mathematics, 76*, 101-115. https://doi.org/10.1007/s10649-010-9271-4

Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching, 8*, 31-54. https://doi.org/10.1191/1362168804lr133oa

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*, 71-81. https://doi.org/10.1080/13558000110102904

Strachan, I. B., & Wilcox, S. (1996). Peer and self-assessment of group work: Developing an effective response to increased enrollment in a third-year course in microclimatology. *Journal of Geography in Higher Education, 20*, 343-353. https://doi.org/10.1080/03098269608709377

Strijbos, J. W., & Sluijsmans, D. M. A. (2010). Guest editorial. Unraveling peer assessment: Methodological, functional, and conceptual developments. *Learning and Instruction, 20*, 265-269. https://doi.org/10.1016/j.learninstruc.2009.08.002

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276. https://doi.org/10.3102/00346543068003249

Topping, K. J., Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193-211. https://doi.org/10.1080/02602938.2019.1620679

van Zundert, M., Sluijsmans, D. M. A., & Van Merrienboer, J. J. G. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction, 20*, 270-279. https://doi.org/10.1016/j.learninstruc.2009.08.004

Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International, 40*, 281-290. https://doi.org/10.1080/1470329032000103816

Webb, N. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389. https://doi.org/10.5951/jresematheduc.22.5.0366

Wen, M. L., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*, 27-44. https://doi.org/10.1007/s10734-004-6375-8

Xiang, X., Yuan, R., & Yu, B. (2021). Implementing assessment as learning in the L2 writing classroom: A Chinese case. *Assessment & Evaluation in Higher Education, 47*(5), 727-741. https://doi.org/10.1080/02602938.2021.1965539

Yin, C., Chen, X., & Chang, Y. (2022). Peer assessment as a means to promote language proficiency: Evidence from classroom practices. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.912568

Zariski, A. (1996, February). Student peer assessment in tertiary education: Promise, perils, and practice. In J. Abbott, & L. Willcoxson (Eds.), *Teaching and learning within and across disciplines* (pp. 189-200). *Proceedings of the 5th Annual Teaching Learning Forum*.

Zevenbergen, R. (2001). Peer assessment of student constructed posters: Assessment alternatives in preservice mathematics education. *Journal of Mathematics Teacher Education, 4*, 95-113. https://doi.org/10.1023/A:1011401532410