Scoring Rubrics Method in Performance Assessment and its effect of Mathematical Achievement

This study aimed to investigate the impact of using scoring rubrics on assessing the performance of students in achievement. The study followed an experimental approach, and the sample consisted of 187 male and female students enrolled in the Calculus course. They were divided into three groups: the first experimental group, whose performance was evaluated using analytical scoring rubrics, the second experimental group, whose performance was evaluated using holistic scoring rubrics, and the control group, whose performance was evaluated using the traditional method. Additionally, a mathematics achievement test was developed, and two of scoring rubrics, one analytical and the other holistic, were prepared to evaluate students' performance. The results of the study favored the use of analytical scoring rubrics over holistic correction rules, as they considered all details, procedures, and levels of understanding and perception. The students expressed satisfaction with the use of analytical and holistic performance scoring rubrics in evaluating their performance. The study recommended the need for students to pay attention to interpreting their procedures when performing mathematical tasks. It also encouraged teachers to use scoring rubrics to evaluate students' performance and called upon curriculum authors to make the necessary modifications and additions to increase students' opportunities for justifying their procedures. Moreover, conducting in-depth studies that allow students to justify their procedures was suggested.

Keywords: *Performance-based Assessment, Scoring Rubrics, Performance, Mathematical Achievement, Composition and Inverse functions*.

Introduction

The past few years have witnessed significant development in the field of assessment and its growing importance among education experts for the purpose of educational reform and improvement. Thomas Kalvin, the president of the International Organization for Measurement and Evaluation, addressed the issue of educational reform and development through various assessment processes during the 27th conference of the organization. He emphasized that most past reform efforts focused on educational inputs, while recent trends have emerged that prioritize educational outcomes and the extent to which students acquire knowledge, skills, behaviors, and attitudes.

According to the National Council of Teachers of Mathematics (NCTM, 2000), assessment is the process of gathering evidence about students' mathematical knowledge, their ability to use mathematical knowledge, and their attitudes towards mathematics. It involves extracting judgments from this evidence for various purposes. Assessment has become a primary source of evidence on which teachers base their inferences about what students know or need to learn. Furthermore, student assessment should not be conducted solely for the purpose of

evaluation but also for guiding and supporting student learning and understanding patterns of knowledge.

Education experts and researchers (Jarrah et al., 2020; Alotaibi et al., 2021; Tami & Roger, 2000) believe that assessment is a systematic process that requires collecting objective and authentic data from multiple sources using various tools, in line with specific objectives, to reach quantitative estimates and descriptive evidence. These estimates and evidence are used to make appropriate judgments or decisions. Undoubtedly, these decisions have a significant impact on learners' performance and their ability to carry out specific tasks or assignments. On the other hand, traditional assessment approaches adopt an educational philosophy that emphasizes highlighting individual differences and encourages competition for achieving a superior relative position among peers, without considering the individual's possession of functional skills, ethics, and constructive positive behaviors.

This narrow perspective focuses on the accumulation of specific information in the learner's mind, which no longer suits the current and future requirements of education and its changing needs in this era characterized by cognitive explosion, information revolution, and communication. In light of this broader perspective, assessment is no longer an end in itself for determining the success or failure of learners and their progression to higher grades or educational stages. Instead, it has become an integral part of the learning process, guiding, enhancing, and correcting its trajectory. This requires a shift from prevailing traditional testing methods, numerical grades, and assessments that focus solely on comparing learners' performance to the adoption of methods and systems that foster the learner's integrated and balanced personality, their possession of functional skills, and deep understanding of the curriculum content acquired through self-learning and curiosity, enabling them to interact with and enrich their environment (Assessment Strategies and Tools Manual, 2004).

Tashtoush & Rasheed (2023) argue that to improve and enhance the assessment process in mathematics, there must be standards or criteria that determine how to deal with assessment mechanisms. This is essential for the development of assessment tools used by teachers. Assessment is defined as the process of making judgments about the value of objects, individuals, or subjects, and in this sense, it requires the use of standards or criteria to estimate this value. It also encompasses the meaning of improvement, modification, and development that relies on these judgments (Tashtoush & Rasheed, 2023b). This guide considers the learning outcomes that should be the basis for planning the assessment process, selecting appropriate tools, distributing grades across assessment tools, and allowing teachers the opportunity to choose the suitable tool that aligns with the educational situation and the learning goal being assessed.

Many educational experts and researchers (Abdallah & Wardat, 2021; Tashtoush et al., 2023 c; Moscal, 2003; Sahin & Baki, 2010; Abo Obaid, 2011) agree that the assessment process utilizes several tools, including:

- **Continuous Assessment:** This is an organized assessment conducted throughout the teaching process, aiming to diagnose strengths and

weaknesses in learners' performance, identify difficulties they face during learning, apply appropriate remedial methods, assist learners in recognizing their abilities and potentials, suggest ways to enhance them to the maximum extent, and consequently issue a realistic judgment determining learners' performance level. This type of assessment contributes to detecting weaknesses and strengths in the educational program, with the purpose of reviewing, modifying, and developing its components.

- **Formative Assessment:** This assessment accompanies daily teaching activities and aims to continuously provide teachers and learners with performance results to improve the educational process. It is used to assess the achievement of objectives and utilize feedback to modify the course and enhance the learning process.
- **Summative Assessment:** It refers to making judgments about the extent of learners' acquisition of learning outcomes with the aim of making decisions such as promoting learners to a higher level or graduation.
- **Self-Assessment:** It involves students' participation in determining levels and criteria of performance and applying them to their own work, issuing judgments related to their achievement of these criteria and levels. Self-assessment serves as a tool for reflection, self-learning, and self-monitoring of performance.
- **Peer Assessment:** It is the active collaboration of a small group of learners to assess the work completed by one of their members or another group, aiming to achieve one or multiple defined objectives within the framework of acquiring cognitive or social knowledge that benefits them through the teacher's evaluation.

One of the common methods that has gained significant popularity in recent years in assessment is performance-based assessment, also known as scoring rubrics. Abu Obaid (2011) defines scoring rubrics as plans developed by specialists or teachers to guide them in analyzing students' performance tasks. Moscal (2003) also defines them as methods that can be used to evaluate students' responses to performance-based assessments, based on beliefs that good assessment begins with considering what students should know and how to assess this knowledge. Linda (1999) views scoring rubrics as benchmarks that include rules, principles, and explanations used to assign grades to responses on each criterion and all the criteria that make up the performance task.

In contrast to traditional assessment approaches, which usually involve more objective methods of grading tasks, alternative assessment and its accompanying use of scoring rubrics involve self-judgments that create a greater challenge in fostering trust and agreement in grading performance tasks. Using scoring rubrics extensively contributes significantly to providing high levels of confidence in assigning authentic and reliable grades to students (Cohen, 1994).

Scoring rubrics have garnered significant attention from many education experts in recent times, owing to the increasing emphasis on performance-based assessment. They provide indicators of performance quality in tasks and make

self-judgments more objective and realistic, moving away from biases when evaluating student performance. Several education experts (Hart, 1994; Moscal, 2003; Alarabi & Wardat, 2021; Sahin & Baki, 2010; Abo Obaid, 2011; Tashtoush et al., 2022b) classify scoring rubrics into four main types as follows:

- **Holistic Scoring Rubrics:** These rubrics provide an overall assessment of the student's performance in a comprehensive and holistic manner. It estimates the student's proficiency as a whole, where each rating on the scale represents a general impression. This type of rubric is not suitable for classroom use as it focuses on overall competence and is not designed to align with curriculum or instructional objectives.
- **Analytic Scoring Rubrics:** In these rubrics, the scoring is divided into separate categories or dimensions that represent different aspects of performance. Each dimension is measured separately, and the results of the dimensions are combined to determine an overall score. The multiple dimensions provide teachers with the opportunity to assess various areas that may differ in their overall importance. Analytic scoring rubrics also provide more information to students about their strengths and weaknesses in multiple areas of performance.
- **Single-Trait Scoring Rubrics:** These rubrics involve pre-determining the main criterion for successful task performance that needs to be assessed. The single-point feature is identified by the teacher based on the nature of the task. This involves narrowing down the criteria for judging performance in the task to a single classification or main dimension. It helps teachers and students focus on a single aspect of performance.
- **Multi-Trait Scoring Rubrics:** These rubrics resemble single-point scoring rubrics, but they allow for the assessment of performance across multiple dimensions. While they share similarities with analytic scoring rubrics in measuring multiple areas, multi-trait scoring rubrics differ in the nature of the dimensions or traits that make up the assessment scale.

These different types of scoring rubrics offer educators a range of options to assess and provide feedback on student performance, allowing for a more comprehensive evaluation aligned with specific objectives and providing students with valuable information about their performance in different areas.

Both holistic and analytic scoring rubrics are characterized by their ability to gather data and information about students' performance levels in educational tasks and improve their performance and cognitive skills. After constructing them, teachers use them for evaluating students' performance on one hand and benefiting from them in the teaching process on the other hand (Stanley, 2014). However, holistic scoring rubrics estimate students' performance as a whole and provide an individual score. They are employed when a quick and consistent judgment is needed, especially for assessing complex and internally interrelated skills. This type of scoring rubric is often used in standardized tests (Tashtoush et al., 2023a; Mertler, 2001). On the contrary, analytic scoring rubrics make judgments on each dimension of performance independently, offering a gradation for each dimension

and an overall assessment of all dimensions. They provide more detailed information but take longer to administer compared to holistic scoring rubrics. Diagnostic tests frequently utilize this type of scoring rubric (Nitko, 2001).

Based on the foundation that traditional assessment methods in most educational institutions have proven to be ineffective in measuring students' skills and knowledge, it became necessary to develop and adopt modern assessment methods, diversify student assessment approaches, and focus on performance based on performance criteria. These approaches aim to measure learning outcomes and processes simultaneously (McLellan, 2008). Therefore, this study considers performance assessment rubrics as indicators of performance quality in specific tasks. They have the ability to gather information about students' task performance, improve their performance and cognitive skills, and contribute to making self-judgments on student performance more realistic. Many education experts in this field recognize the use of performance assessment rubrics as providing convincing justifications for the feasibility and effectiveness of this type of assessment.

Problem Statement

Scoring rubrics are educational concepts and a type of assessment based on performance that plays a significant role in evaluating the process of teaching and learning mathematics. Undoubtedly, this type of assessment provides teachers with information about students' understanding of knowledge and skills and their ability to apply them in learning mathematics. It also enables teachers to integrate classroom teaching with performance-based assessment and scoring rubrics, enriching their educational and pedagogical experience.

Since associations are the fundamental building blocks of studying calculus, this study aims to provide a background on some scoring rubrics that fall under the title of performance-based assessment. These rules can benefit teachers and university instructors, as they are the key element in the assessment process in the classroom. Two scoring rubrics were adopted for assessing mathematical tasks: the first is the holistic correction rule, and the second is the analytical correction rule. Each of them has its own aspects. The holistic correction rule provides a basis for comparing two performances, but it does not rely on task analysis or provide diagnostic information about students' task performance. On the other hand, the analytical scoring rubrics provides more detailed grading and its results are described as more accurate, although it may focus more on certain aspects of performance compared to others.

Due to significant shortcomings in the methods and techniques of assessment adopted in our various educational institutions, and based on field observations by researchers who are mathematics teachers at different educational levels, as well as their supervision of pre-service mathematics teachers in the educational field, and the participation of others in international assessment tests such as TIMSS, PISA, and STEM, this study aims to respond to the importance of having modern assessment methods and techniques. It aligns with the global, Arab, and local

movements in the Jordan to develop assessment methods and techniques based on student performance. The idea of this study also stems from the researchers' struggle to understand students' thought processes when they attempt to perform tasks of different levels in operations on associations, especially in tasks related to composing associations and finding inverse associations, and interpreting the low level of performance in these tasks. Specifically, this study aims to answer the following main question: ***What is the impact of using scoring rubrics for performance assessment on the achievement of students in Calculus course?***

Study Importance

The importance of the study stems from the significance of the topic it addresses. The theoretical importance lies in the fact that it is one of the studies that call for the examination of associations in general, and specifically the operations of composing associations and finding inverse associations. It involves using methods and techniques to assess students' performance in Calculus, based on alternative evaluation that requires searching for multiple sources of evidence, building conclusions, and judging what students know to achieve realistic assessment. Researchers hope that this study will enrich the theoretical and research literature in mathematics and fill a gap in this area of research. The practical importance of this study lies in its potential to benefit students, faculty members, and specialists in developing the assessment process. It enables them to become familiar with performance-based assessment tools and the use of correction rules.

Limitations

- **Time Boundaries:** This study was conducted during the second semester of the academic year 2022/2023.
- **Spatial Boundaries:** The study was conducted at AL-Huson College University at AL-Balqa Applied University.
- **Human Boundaries:** The study was conducted on students enrolled in the Calculus course.
- **Subject Boundaries:** This study addressed the overall and analytical correction rules, their results using data collection tools, procedures, the nature of the community and the sample, and the operations of composing associations and finding inverse associations.
- The study is defined by its psychometric properties, including acceptable validity and reliability, for the purposes of scientific research to achieve the study's objectives.

Procedural Definitions

- **Performance-based Assessment:** It is a type of assessment designed to measure a student's ability to perform specific tasks and judge their achievement using assessment tools that estimate their level.
- **Scoring Rubrics:** A method of evaluating a student's performance based on meaningful criteria and judging their performance level in a single task or multiple tasks.
- **Holistic Scoring Rubrics:** A scale that provides a general overview and estimation of a student's performance in a specific task.
- **Analytical correction Rubrics:** A scale that categorizes a student's performance in a specific task into multiple levels, where each level is measured separately, and then an overall judgment is made based on all levels.

Literature Review

Through reviewing the theoretical and educational literature that addressed the importance of using scouring rubrics in teaching and learning mathematics, this study discussed some previous studies related to the subject of the current study, which could be useful to mention in the current study. McBride and Carifio (1995) conducted a study aiming to investigate the effectiveness of using analytical scouring rubrics to assess students' performance in geometric proof by employing cognitive behavior theories of geometric knowledge. A group of evaluators assessed the performance of 241 students in geometric proof using a correction rule that included multiple criteria through various tests measuring students' ability in geometric proof. The study results showed that the use of analytical scouring rubrics yielded better results than traditional assessment methods.

Similarly, Lumely and Yan (2001) examined the impact of broad-spectrum assessment on teaching methods and instructional strategies followed by teachers in Pennsylvania. The study identified factors influencing teachers' beliefs and usage of broad-spectrum assessment and used a questionnaire administered to 168 teachers from 20 different schools. The study revealed that while teachers recognized the importance and value of the scouring rubrics presented in the assessment guide, they did not use them. They attributed this to developing their own correction rules. Furthermore, teachers adhered to traditional methods and did not adopt advanced scouring rubrics in assessment.

Al-Absi (2007) conducted a study aiming to investigate the effect of using scouring rubrics to assess performance in achievement and attitudes of tenth-grade students in mathematics in Jordan. Two correction rules, holistic and analytical, were developed, and an achievement test and a questionnaire to measure students' attitudes were administered to 128 students divided into three groups: the first experimental group was assessed based on an analytical correction rule, the second experimental group was assessed based on a holistic correction rule, and the

control group was assessed using traditional methods. The study results showed the effectiveness of both holistic and analytical assessment methods on students' performance. The study recommended incorporating scouring rubrics as a means of assessing students' performance and achievement in mathematics and training teachers to use them.

In the same context, Sahin and Baki (2010) conducted a study aimed at investigating the possibility of using scouring rubrics as a multidimensional assessment approach to evaluate mathematical power. The study included 62 students from three eighth-grade classes in Turkey and employed a case study methodology. The study results showed that the use of scouring rubrics effectively contributed to the students' growth in problem-solving skills, decision-making abilities, communication skills, as well as the evaluation of practical and conceptual knowledge. This demonstrated the potential for assessing learning outcomes and processes.

Furthermore, Balawneh (2010) aimed to examine the effectiveness of performance-based assessment in developing mathematical thinking and problem-solving abilities among secondary school students. The study included a sample of 74 female students from the eleventh grade in a school in Jordan. An experimental group was evaluated using performance-based assessment, while a control group was assessed using traditional methods. The researcher utilized tests for mathematical thinking and problem-solving to collect data. The study results indicated statistically significant differences favoring the experimental group in terms of mean scores in mathematical thinking and problem-solving tests. The researcher recommended further studies on alternative assessment methods, the development of specialized assessment guidelines, and the diversification of classroom exercises and homework to include performance tasks that stimulate logical thinking.

Regarding studies targeting secondary school students, Abu Obeid (2011) aimed to investigate the impact of using scouring rubrics to assess performance on students' achievement and attitudes towards mathematics. Two correction rules, holistic and analytical, were developed for this purpose. The study included a sample of 128 students divided into three groups: the first experimental group was assessed using analytical correction rules, the second experimental group was assessed using holistic correction rules, and the control group was assessed using traditional methods. Achievement tests and an attitude scale towards mathematics were administered after the study implementation. The study results revealed statistically significant differences attributed to the application of the assessment method in favor of the first and second experimental groups compared to the control group, while no statistically significant differences were found between the two experimental groups.

In a related context, Al-Maliki (2011) conducted a study aimed at investigating the impact of using analytical performance evaluation criteria on the academic achievement of third-grade students. The study included a sample of 46 students from schools in KSA, divided into two groups: experimental and control. A test was administered to assess academic achievement. The results showed the effectiveness of analytical performance evaluation criteria in assessing students'

performance, and the study recommended the use of these criteria in teaching mathematics to elementary stage students.

Al-Ruwaili (2016) conducted a study to examine the effect of scouring rubrics on performance evaluation and its impact on the academic achievement and attitudes of eleventh-grade female students towards mathematics. The purposive sample consisted of two groups: experimental (29 students) and control (28 students). An achievement test of multiple-choice type and an attitudes scale were prepared. The results of the study indicated the effectiveness of using scouring rubrics in evaluating students' performance, and the study recommended the utilization of scouring rubrics in assessing students' performance.

Sarhani (2016) aimed to investigate the effectiveness of using analytical performance evaluation criteria in solving mathematical problems and enhancing the academic achievement of seventh-grade students in KSA. The researcher used an experimental approach with two groups: control and experimental. The study included a sample of 46 students. The researcher developed a student activity guide and a teacher guide for teaching ratio and proportion lessons based on analytical performance evaluation criteria, along with a achievement test. The results showed the effectiveness of the experimental treatment in improving students' achievement in the achievement test with a significant effect size. The study suggested conducting further research on analytical performance evaluation criteria to enhance learning in other mathematical subjects and different educational stages, as well as comparing them with holistic performance evaluation criteria.

Tashtoush and Rasheed study (2023) aimed to evaluate the performance of calculus students on mathematical tasks and the procedures they follow in the processes of composite functions and finding inverse functions, as well as interpreting these procedures. The study used a mixed-methods approach and included three students enrolled in the Calculus course at Sohar University. The students were subjected to a short test, and their performance was evaluated using both holistic and analytical correction rules. Individual interviews were conducted and analyzed. The results of the study favored the use of analytical scouring rubrics over holistic correction rules, as they considered all details, procedures, and levels of understanding and perception. The students expressed more satisfaction with the analytical correction rules, which considered their responses to the test tasks. The study recommended that students pay attention to explaining their procedures when performing mathematical tasks and encouraged teachers to use scouring rubrics to assess students' performance.

While previous studies mentioned focused on students in different educational stages, covering various topics such as algebraic equations, operations on numbers, ratio and proportion, fractions, and geometry, the current study focused on university-level students in advanced topics related to composite functions and finding inverse functions. Additionally, the current study followed an experimental methodology, whereas some previous studies followed mixed-methods, descriptive, or case study approaches. However, this study aligns with previous studies in utilizing tests as a data collection tool. This study contributes to

supporting previous studies that emphasized the importance of using scouring rubrics in evaluating students' performance in mathematics.

Methodology

The current study followed an experimental methodology to investigate the impact of using scouring rubrics for evaluating performance in the achievement of students taking the Calculus course. The study was conducted based on three groups: the first experimental group, the second experimental group, and the control group.

Participants

The population of the study consisted of regularly enrolled students at AL-Huson College University at AL-Balqa Applied University during the second semester of the academic year 2022/2023. The study sample was purposefully selected from students enrolled in the Calculus course and distributed across five sections. Three sections were randomly selected to represent the three study groups: the first experimental group, whose performance was evaluated using analytical correction rules; the second experimental group, whose performance was evaluated using holistic correction rules, and the control group, whose performance was evaluated using the traditional method. Table (1) illustrates the distribution of study participants.

Table 1. Study Sample

Group	**No**
First Experimental Group	62
Second Experimental Group	64
Control Group	61
Total	**187**

Instruments

Mathematical Achievement Test: After reviewing the theoretical and research literature (Tashtoush, 2009; Al-Ruwaili, 2016; Sarhani, 2016; Tashtoush & Rasheed, 2023b), a Mathematical Achievement Test was developed, the test consisted of three mathematical tasks that covered the process of combinatorial permutations, the nature of combinatorial permutations (whether they are switchable in general or not), finding the inverse permutation and its relationship to the process of combining it with itself, each task was assigned five points, and the maximum score for the Mathematical Achievement Test was 15, the test tasks were structured as follows: **Direct Task:** Students answered this task by identifying the given data and applying the rule directly in two steps. **Reverse Task:** Students used the given data and performed the steps of the previous task in

reverse order, linking the results. **Comprehensive Task for both Concepts:** Students analyzed the given data, found the inverse permutation, and linked it to the concept of combining it with itself when verifying the correctness of the answer.

To ensure the validity of the instrument, it was presented to a group of experts in pure mathematics, mathematics curricula, and teaching methods. Their opinions and suggestions regarding the tasks, language accuracy, question nature, difficulty level, and alignment with the study objectives were taken into consideration. Based on their feedback, the test items were revised until the test reached its final form. The opinions of the students were also obtained through a survey sample of 25 students who were selected from outside the study sample. Their feedback was used to modify the wording of the third task, which students had difficulty understanding regarding the method of verifying the solution's accuracy.

To establish the reliability of the instrument, the method of inter-rater agreement was followed by calculating the correlation coefficient between the evaluations of the two researchers for the survey sample. The coefficient was found to be 0.958. This coefficient was chosen because it is difficult to conduct a test-retest reliability coefficient for the students' responses. It is not preferable to reapply the test to the same group with a time interval between the applications, as it may be influenced by the students' recall factor of the test tasks, leading to an increased reliability coefficient. Alternatively, students might become familiar with the test, resulting in higher scores in the second application and reducing the reliability coefficient.

Scoring Rubrics: This study relies on evaluating students' performance as an attempt to develop the assessment system for university instructors in general and specifically for instructors at AL-Balqa Applied University. Two scoring rubrics were applied to the three tasks given to students: the holistic and analytical. A set of descriptors was developed to describe different performance levels on the tasks using the holistic scoring rubric, which includes the following categories: (Perfect, Median, Novice, Weak). The following tables provide a description of each level along with the corresponding score for each performance level on the three tasks.

Table 2. Student performance using the holistic scoring rubrics on the first task

Performance Level	**Performance Details**	**Scoring**
Perfect	Utilizing rules and algorithms correctly and in a proper sequence to find the matching $(fog)(x)$	**2**
Median	Improper utilization of rules and algorithms to find the structure, or an incorrect sequence in using the rules in opposite side $g(f(x))$ not $f(g(x))$	**1**
Weak	Erroneous utilization of rules and algorithms without considering any sequencing to find the structure, or no solution exists	**0**

Table 3. Student performance using the holistic scoring rubrics on the second task

Performance Level	Performance Details	Scoring
Perfect	Proper utilization of rules and algorithms involves finding $(gof)(x)$ and establishing the relationship between $(fog)(x)$, followed by performing a comparison	**2**
Median	Incorrect utilization of rules and algorithms, failing to establish any connection between direct and inverse usage, thus neglecting to mention the relationship between them	**1**
Weak	Incorrect application of the inverse procedure without addressing any relationship between it and the original procedure, or no solution exists	**0**

Table 4. Student Performance using the holistic scoring rubrics on the third task

Performance Level	Performance Details	Scoring
Perfect	Proper utilization of rules and algorithms with a correct sequence and validating the correctness of the solution using two valid methods	**3**
Median	Proper utilization of rules and algorithms with a correct sequence and validating the correctness of the solution using one valid method	**2**
Novice	Proper utilization of rules and algorithms with a correct sequence, but without verifying the accuracy of the solution, or errors in the operations resulting in an incorrect final answer	**1**
Weak	Incorrect utilization of rules and algorithms with an incorrect sequence (resulting in an incorrect solution), without verifying the accuracy of the solution, or no solution exists	**0**

In the analytical scoring rubrics, estimates have been provided to describe the different levels of performance for students in three tasks, which include: (Task Understanding, Solution Planning, Solution Execution, and Solution Verification). The following tables illustrate a description of each performance level for the three tasks, along with the corresponding Scoring.

Table 5. Student performance using the analytic scoring rubrics on the first task

Performance Level	Performance Details	Scoring
Task Understanding	Clear and accurate understanding is demonstrated through writing the rules for both $(f)(x)$ and $(g)(x)$ and identifying the task's objective, which is to find $(fog)(x)$	**2**
	Clear identification of the given information without specifying the objective $(fog)(x)$, or specifying the objective without writing the given data	**1**
	Failure to specify the given data or the objective (indicating a lack of understanding of the problem)	**0**
Solution Planning	The required strategy is complete: $(fog)(x)$, followed by a correct sequence of steps	**2**
	The strategy only addresses a portion of g(x), or the strategy does not lead to solving both $(g)(x)$and $(f)(x)$separately.	**1**
	There is a lack of strategy.	**0**

Solution Execution	correct manner: Applying f to g.	**2**
	Incorrect solution execution: Applying g to f.	**1**
	Failure to execute the solution	**0**
Solution Verification	Complete verification: Repeating the solution to ensure its accuracy.	**2**
	Partial verification: Incorrectly repeating the solution	**1**
	Failure to verify.	**0**

Table 6. Student performance using the analytic scoring rubrics on the first task

Performance Level	Performance Details	Scoring
Task Understanding	Clear and accurate understanding is demonstrated through writing the rules for the two mappings, $(f)(x)$ and $(g)(x)$, and specifying the objective of the task, which is to find $(gof)(x)$. The understanding is directly related to the first task.	**2**
	Specifying the given data and identifying the first objective $(gof)(x)$ without understanding its relationship to the first task, or specifying the given data and the first objective $(gof)(x)$ with a misunderstanding of the second objective (its relationship to the first task	**1**
	No given data, no objective specified, and no understanding of the relationship between the objective and the first task.	**0**
Solution Planning	Complete plan: Sequential steps to find $(gof)(x)$ and compare it to $(fog)(x)$.	**2**
	A plan to find $(gof)(x)$ without connecting it to the first task.	**1**
	No plan provided.	**0**
Solution Execution	Correct solution execution: Finding $(gof)(x)$, comparing it to $(fog)(x)$, and writing the complete commentary	**2**
	Correct solution execution without linking it to the first task, or incorrect solution execution with an incorrect link to the first task.	**1**
	Failure to execute the solution.	**0**
Solution Verification	Complete verification: Repeating the process of finding $(gof)(x)$ and ensuring its relationship to the first task.	**2**
	Incorrect repetition of the solution.	**1**
	Failure to verify the solution.	**0**

Table 7. Student performance using the analytic scoring rubrics on the first task

Performance Level	Performance Details	Scoring
Task Understanding	Clear, accurate, and comprehensive understanding is demonstrated through writing the rule for f(x), specifying the objective as finding $f^{-1}(x)$, and mentioning the validation rules for confirming the solution	**2**
	Partial understanding: Providing the given data without specifying the objective.	**1**
	No understanding demonstrated.	**0**
Solution Planning	The strategy is complete: Symbol substitution, followed by crucial steps, and then an overview for verification	**2**
	The strategy is incomplete: Steps are provided without symbol substitution, and there is no plan for verification.	**1**
	No strategy or verification plan provided.	**0**
Solution Execution	Correctly: Applying the steps in a proper sequence.	**2**
	Completely incorrect: Making errors in executing the steps or in the symbols used.	**1**
	Failure to execute the solution	**0**
Solution	Complete verification: Checking in at least one direction or both	**2**

Verification	directions.	
	Partial verification: Not reaching a conclusive result.	**1**
	No verification conducted.	**0**

To ensure the validity of the scoring rubrics, they were presented to a group of experts in mathematics curriculum and teaching methods. Their opinions and suggestions regarding the criteria on which the rubrics were built, the nature of the assessment, and the extent to which they achieve the study's objectives were taken into consideration. Based on their feedback, some criteria in the scoring rubrics were modified until they reached their final form.

To establish the reliability of the instrument, the researchers evaluated the responses of the survey sample twice, with a one-week interval between evaluations. The results of each evaluator were compared independently, and the agreement coefficient was measured using Holsti's equation (Holsti, 1969), which yielded a value of 0.938. Additionally, the Pearson correlation coefficient was calculated for the evaluations of student performances according to the holistic and analytical scoring rubrics, resulting in a coefficient of 0.965.

Procedures

The study sample was determined, and the mathematics achievement test and the scoring rubrics scale for holistic and analytical scoring were prepared to be implemented during the experimental period. Validity and reliability checks were conducted for both instruments. The students were exposed to an objective explanation of the concept of permutations and finding the inverse permutation two weeks prior to the study. The survey sample of students was given a brief test containing the three tasks of the study during one of the lectures. The students' opinions about the test were taken into consideration, and based on their feedback, the test duration was increased from 15 minutes to 20 minutes. Furthermore, the third task was modified after the students expressed their lack of direct understanding of the intended purpose of verifying the correctness of the answer. Afterwards, the test was administered to the study sample under natural conditions, resembling a short classroom test. The students had already been acquainted with the details of the holistic and analytical scoring rubrics before taking the test. Each student's performance on the test tasks was evaluated according to the holistic and analytical scoring rubrics. A comparison was made between the evaluation results, and the students' performance assessment results were entered into the SPSS program for data analysis, addressing the research question, discussing the findings, and presenting recommendations and suggestions.

Data Analysis

For the statistical analysis procedures, the data was entered into the computer memory and analyzed using the SPSS software for statistical data processing. The

mean scores and standard deviations were calculated for the three groups of students on the achievement test to detect apparent differences in the mean scores. Additionally, one-way analysis of variance (ANOVA) was used to determine the presence of differences among the mean scores of the three groups of students based on the evaluation method for the mathematics achievement test.

Results

The main objective of the research question was to determine the effect of using scoring rubrics for performance assessment on the achievement of students in Calculus. To detect statistically significant differences at a significance level (α=0.05) among the students' performances on the mathematics achievement test, the mean scores and standard deviations were calculated for each of the three groups: the first experimental group, the second experimental group, and the control group. Table (8) illustrates this.

Table 8. The Mean Scores and Standard Deviations

Group	No	Test	Mean	SD
First Experimental Group	62	Pre-test	9.88	2.64
		Post-test	12.44	3.28
Second Experimental Group	64	Pre-test	9.72	2.61
		Post-test	12.28	3.21
Control Group	61	Pre-test	9.71	2.63
		Post-test	9.88	2.69
Maximum Score of Achievement Test is (15)				

Table (8) demonstrates the apparent variability in the arithmetic means and standard deviations of students' performance on the mathematics achievement test, according to the group variable (First Experimental Group, Second Experimental Group, and Control Group). It aims to determine whether the observed differences in arithmetic means are statistically significant at a significance level of (α=0.05). One-way analysis of variance (ANOVA) was utilized, and Table (9) illustrates the results.

Table 9. ANOVA Test for the Math. Achievement Test

Source of Variance	Sum of Squares	df	Mean of Squares	F	Sig.	Effect Size (η^2)
Between Group	208.65	2	104.33	5.01	0.000*	0.72
Within Group	3825.36	184	20.79			
Total	4034.01	186				
* Sig. Level (α=0.05)						

From Table (9), it is evident that there are statistically significant differences (α = 0.05) attributed to the effect of the group, with the differences favoring the first experimental group. The results also indicate the effectiveness of using scouring rubrics and their significant impact on evaluating students' performance in the achievement test. The effect size, measured using eta squared (η^2), was 0.72,

meaning that 72% of the variance in students' performance can be attributed to the use of correction rules, while 28% can be attributed to uncontrolled factors.

Discussion

The study results showed statistically significant differences in the means of students' grades among the three groups in the achievement test, attributed to the evaluation method. This was in favor of the first experimental group, who were evaluated using analytical correction rules, and the second experimental group, who were evaluated using holistic correction rules, compared to the control group, who were evaluated using the traditional method. This result aligns with the modern educational system's aim to integrate assessment as an integral part of the educational process. Assessment greatly contributes to students' learning of mathematics, improves their achievement, and provides opportunities for obtaining information about their performance, which positively impacts their learning in Calculus.

The reason for the superiority of the first and second experimental groups, evaluated using analytical and holistic correction rules, may be attributed to the fact that performance-based learning is more student-centered. Performance tasks help students understand their strengths to enhance them and identify their weaknesses for improvement. They also evaluate learning processes and outcomes, providing an accurate assessment of students' actual performance. Furthermore, the use of scouring rubrics aims to identify the individual student's performance level. Students need to feel that to achieve excellence, they must be diligent and perseverant, which positively affects their achievement. The use of scouring rubrics integrates learning and assessment processes, resulting in students acquiring knowledge, understanding, and skills using various teaching styles and strategies (Tashtoush et al., 2022a; Fannakhosrow et al., 2021; Rasheed et al., 2021; McLellan, 2008; Rasheed & Tashtoush, 2023) to achieve that knowledge through a multidimensional assessment of their performance. It is natural for students to have more confidence in their actions when they are aware of the details of the scouring rubrics they will encounter. However, they need to become accustomed to such procedures that will be applied to their performance before they face specific tasks. This increases the value of their actions in response to these tasks and generates the conviction that if they fail in some aspects, they can still succeed in other areas.

The results also indicate a preference for analytical scouring rubrics over holistic correction rules. The reason for this may be that the analytical correction method considers all the details of the procedures, as well as the levels of understanding, perception, problem-solving, and verification for students. This generates a sense of assurance among students that their rights will not be completely lost if they fail in some aspects. They trust some aspects and strive in others, even if they are unsure of their ability to accomplish them. This type of assessment gives attention to details that students would not have cared about without being exposed to their evaluation method, such as writing a solution plan

and verifying its correctness. This result aligns with the modern assessment system's goal of highlighting the mathematics that students must know and perform, providing each student with the opportunity to demonstrate their mathematical ability according to their capabilities. This positively affects students' performance and achievement. Additionally, introducing students to the assessment method and criteria used to judge their performance can have a positive impact on alleviating their anxiety, as indicated by the results of some studies (Tashtoush et al., 2020; Sarhani, 2016; Wardat et al., 2023; Tashtoush & Rasheed, 2023 b; Carifio & McBride, 1995).

Recommendations

Based on the positive results obtained from this study, the researchers recommend the following:

- Increase teachers' awareness and understanding of the value of encouraging students to explain their procedures while performing assigned tasks. Create an environment that facilitates this in the classroom and focus on verifying the correctness of solutions to increase students' confidence in themselves and their procedures.
- Encourage teachers to push students to deepen their understanding and awareness of the steps involved in their procedures, as this has an impact on their learning and progress.
- Encourage mathematics teachers at AL-Balqa Applied University and various Jordanian universities to prepare and use different types of performance scouring rubrics to assess their students' performance based on performance-based assessment criteria.
- Call upon curriculum and textbook authors, especially those responsible for the Calculus course and other mathematics courses in general, to make necessary additions and modifications based on the results of this study. This will increase opportunities for students to justify their procedures through tasks that require it or incorporate it to verify the solution.
- Encourage planners and developers of the Calculus course, as well as other mathematics courses, to focus on performance scouring rubrics by enriching the curriculum with performance-based tasks that are built on assessment criteria.
- Conduct similar studies on other tasks that require students to justify their procedures and demonstrate logic and sequencing.

References

Abdallah, R., & Wardat, Y. (2021). Teachers' perceptions on the effectiveness of professional development programs in improving the curriculum implementation at Jordanian schools. ***Elementary Education Online***, *20(5),* 4438- 4449.

Abu Obeid, A. (2011). The Scoring Rubrics for performance assessment and its impact on the achievement and attitudes of the 11th grade students towards mathematics, ***Psychological and Educational Studies***, 7, 25-57.

Abu Saleh, M. and Awad, A. (1997). ***Introduction to Statistics***, Jordan Book Center, Amman: Jordan.

Al-Absi, M. (2007). The Scoring Rubrics for performance assessment and its impact on the achievement and attitudes of tenth grade students towards mathematics, ***Journal of Educational Sciences***, *12*(12), 133-157.

Alarabi, K., & Wardat, Y. (2021). UAE-based teachers' hindsight judgments on physics education during the COVID-19 pandemic. ***Psychology and Education Journal***, *58(3),* 2497-2511.

Al-Maliki, A. (2011). The effect of using analytical rubric on the academic achievement of third graders of primary school, ***Journal of Educational and Psychological Sciences***, *15*(3), 285-298.

Alotaibi, A., Khalil, I., & Wardat, Y. (2021). Teaching practices of the mathematics male and female teachers according to the PISA framework and its relation to their beliefs towards their students. ***Elementary Education Online***, *20*(1), 1247-

Al-Ruwaili, E. (2016). The Scoring Rubrics for performance assessment and its impact on the achievement and attitudes of the 11th grade students towards mathematics, ***Studies: Educational Sciences,*** *43*(5), 1903-1914.

Assessment Strategies and Tools Manual (Theoretical Framework). (2004), Prepared by the National Assessment Team, Examinations Directorate, Ministry of Education, Jordan.

Balawneh, F. (2010). The effect of the performance-based assessment on developing mathematical thinking and problem-solving ability among secondary school students, ***An-Najah University Journal for the Humanities***, *24*(8), 2227-2270.

Cohen, A. (1994). ***Assessing Language Ability in the Classroom***, Second Edition. Boston, MA: Heinle and Heinle.

Fannakhosrow, M., Nourabadi, S., Huy, D., Trung, N., Tashtoush, M. (2022). A Comparative Study of Information and Communication Technology (ICT)-Based and Conventional Methods of Instruction on Learners' Academic Enthusiasm for L2 Learning, ***Education Research International***, 2022, Article ID 5478088.

Hart, D. (1994). ***Authentic Assessment: A Handbook for Educators***, Reading, MA: Addison Wesley Publishing Company.

Holsti, R. (1969). ***Content Analysis for Social Sciences and the Humanities Addison***, Welay Publishing Company.

Jarrah, A. M., Khasawneh, O. M., & Wardat, Y. (2020). Implementing pragmatism and John Dewey's educational philosophy in Emirati elementary schools: case of mathematics and science teachers. ***International Journal of Education Economics and Development***, *11*(1), 58.

Linda, M. (1999). Writing to the Rubric: Lingering Effects of Traditional Standardized Testing on Direct Writing Assessment, ***Phi Delta Kappan***, *80*(9), 673-679.

Lumely, D. & Yan, W. (2001). ***The Impact of state Mandated, Large – Scale Writing Assessment in Pennsylvania***, ERIC Document Reproduction No. ED 453220.

Mathematics Assessment Document for Students for Grades (5-10). (2012), Prepared by the General Directorate of Educational Assessment, Ministry of Education, Oman.

McBride, B. & Carifio, J. (1995). ***Empirical Results of using an Analytic versus Holistic Scoring Method to Score Geometric Proofs***, ERIC Document Reproduction No. ED 401307.

McLellan, S. (2008). When Students Teach: Performance Based Assessment, ***Transformative Dialogues: Teaching & Learning Journal***, *2*(2), 1-12.

Mertler, C. (2001). Designing Scoring Rubrics for your Classroom, ***Practical Assessment, Research, and Evaluation***, *7*(25), 1-8.

Moscal, B. (2003). Recommendations for developing classroom performance assessments and scoring rubrics, ***Practical Assessment, Research and Evaluation***, *8*(14), 1-9.

NCTM. (2000). ***Principles and Standards of School Mathematics***, The National Council of Teachers of Mathematics, Inc.

Nitko, A. (2001). ***Educational Assessment of Students***, Third Edition, Upper Saddle River, NJ: Merrill.

Rasheed, N. & Tashtoush, M. (2021). The Fertility and its Relation with Some Demographic, Economic and Social Variables in Jordan. *Turkish Journal of Computer and Mathematics Education*, *12*(11), 5088-5095. https://www.turcomat.org/index.php/turkbilmat/article/view/6710

Rasheed, N., Tashtoush, M. (2023). The Impact of Cognitive Training Program for Children (CTPC) to Development the Mathematical Conceptual and Achievement. *Journal of Higher Education Theory and Practice*, *23*(10). 218-234. https://doi.org/10.33423/jhetp.v23i10.6196

Sahin, S., Baki, A. (2010). A new model to assessment mathematical power, ***Procedia- Social and Behavioral Sciences***, 9, 1368–1372

Sarhani, M. (2016). The Effectiveness of Using Analytical rubric to Solve Mathematical Problems in Developing Academic Achievement for 7th Grade Students in KSA, ***Journal of the College of Education***, *171*(1), 665-684.

Stanley, T. (2014). ***Performance-Based Assessment for 21st-Century Skills***, 1st edition, Taylor and Francis Group.

Tami S. & Roger D. (2000). Performance-Based Assessment of Secondary Mathematics Student Teachers, ***Action in Teacher Education***, *22*(3), 86-95.

Tashtoush, M. & Rasheed, N. (2023b). The Assessment of the Performance of Calculus Students in Composition Function and Finding an Inverse Function.

6th Sohar University Teaching and Learning Conference (Innovations and Applications in Teaching and Learning), 2 March, 2023, Sohar University, OMAN.

Tashtoush, M. (2009). *Calculus 1 With Examples*, 1st Edition, Text book in Calculus, Dar AL–Amal for Publishing and Distributing, Jordan.

Tashtoush, M., Alali, R., Wardat, Y., AL-Shraifin, N., Toubat, H. (2023c). The Impact of Information and Communication Technologies (ICT)-Based Education on the Mathematics Academic Enthusiasm. ***Jounal of Educational and Social Research***, *13*(3), 287-296.

Tashtoush, M., Alshunaq, M., & Albarakat, A. (2020). The Effectiveness of Self-Regulated Learning (SRL) in CreativeThinking for CALCULUS Students, ***PalArch's Journal of Archaeology of Egypt/ Egyptology***, *17*(7), 6630-6652.

Tashtoush, M., Wardat, Y., Aloufi, F., Taani, O. (2022a). The Effectiveness of Teaching Method Based on the Components of Concept-Rich Instruction Approach in Students Achievement on Linear Algebra Course and Their Attitudes Towards, ***Journal of Higher Education Theory and Practice***, *22*(7), 41-57, USA.

Tashtoush, M., Wardat, Y., Aloufi, F., Taani, O. (2022b). The Effect of a Training Program Based on (TIMSS) to Developing the Levels of Habits of Mind and Mathematical Reasoning Skills among Pre-service Mathematics Teachers, ***EURASIA Journal of Mathematics, Science and Technology Education***, *18*(11), Article No em2182.

Tashtoush, M., Wardat, Y., Elsayed, A. (2023a). Mathematics Distance Learning and Learning Loss During COVID-19 Pandemic: Teachers' Perspectives, ***Journal of Higher Education Theory and Practice***, *23*(5), 162-174.

Wardat, Y., Tashtoush, M., Alali, R., Jarrah, A. (2023). ChatGPT: A Revolutionary Tool for Teaching and Learning Mathematics. *EURASIA Journal of Mathematics, Science and Technology Education*, *19*(7), 1-18, Article No: em2286. https://doi.org/10.29333/ejmste/13272