OPEN ACCESS

Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) |ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

Graduate Students' Experiences in the "Assessment of Mathematics Education" Course at An-Najah National University: A Qualitative Study

Dr. Yousef Jaber Alawneh[1]*

[1]Faculty of Humanities and Education Sciences, An-Najah National University, Nablus, Palestine

DOI: https://doi.org/10.36348/jaep.2025.v09i08.001 | **Received:** 04.06.2025 | **Accepted:** 05.08.2025 | **Published:** 07.08.2025

***Corresponding author:** Dr. Yousef Jaber Alawneh
Faculty of Humanities and Education Sciences, An-Najah National University, Nablus, Palestine

Abstract

This qualitative study aimed to explore the experiences of graduate students enrolled in the "Assessment of Mathematics Education" course at An-Najah National University. The research sought to understand how these students developed achievement tests as part of the course requirements, the challenges they encountered during this process, and their suggestions for improving future instruction in test construction. An exploratory qualitative methodology was adopted, and data were collected through semi-structured interviews with ten students registered in the course. The data were analyzed using thematic analysis following Braun and Clarke's (2006) model. The findings revealed that the course significantly enhanced students' awareness of test construction principles and assessment competencies. However, certain practical challenges persisted—particularly in crafting higher-order thinking questions and formulating precise test items. Students emphasized the importance of hands-on activities in reinforcing their learning and recommended increased opportunities for practical training and stronger connections between course content and real classroom contexts, The study recommends redesigning assessment-related courses in teacher education programs—particularly for future mathematics teachers—to be more interactive and practice-oriented, with individualized support tailored to students' needs. It also calls for future quantitative studies to examine the impact of such courses on students' assessment competencies during later stages of their professional training.

Keywords: Mathematics education assessment, achievement tests, assessment competencies, higher education, qualitative study, An-Najah National University.

BACKGROUND

Educational assessment is a fundamental pillar of modern teaching and learning. It plays a dual role in measuring learning outcomes and guiding the instructional process toward continuous improvement. Assessment is not merely a tool to judge student performance; it is also diagnostic and formative, revealing both strengths and weaknesses in learning and instruction (Brookhart, 2011). Competence in using assessment tools is therefore considered a core component of teachers' professional capacity, especially in learning environments that aim to meet quality and effectiveness standards (Pastore, 2023).

In light of global shifts toward competency-based and outcome-focused education, the need has grown for more reliable and flexible assessment systems—systems that recognize learner diversity, integrate cognitive, practical, and affective dimensions, and adopt participatory strategies involving learners in the assessment process (Brookhart, 2011). Within this context, assessing mathematics learning has gained particular importance due to the nature of the subject, which blends theoretical abstraction with practical application and requires higher-order thinking skills, including analysis, synthesis, and problem-solving. Success in mathematics is widely regarded as an indicator of readiness for engagement in STEM (science, technology, engineering, and mathematics) fields, which underpin the modern knowledge economy (Çalık & Gündüz, 2022; Güler, 2023).

Numerous studies have emphasized that teachers' competence in constructing achievement tests is critical to ensuring fairness and quality in evaluation.

Poorly designed tests can yield misleading results that do not accurately reflect student learning (Fraenkel *et al*., 2012). Comprehensive reviews of qualitative and quantitative research on pre-service teachers' assessment competencies suggest that many novice educators lack sufficient knowledge of foundational assessment principles and struggle to apply theoretical concepts in practice (Pastore, 2023). This gap is often described as "assessment illiteracy," reflecting a deficiency in skills such as analyzing learning outcomes, interpreting data, and constructing valid and reliable instruments (Zhang & Burry-Stock, 2003; Sahin, 2023).

In Arab contexts, studies from Saudi Arabia, Palestine, and Jordan have reported similar challenges. For instance, Al-Zabon (2013) found that the average practice level of test construction competencies among teachers was only 59.1%. More than 70% of the sample had not received specialized training in educational measurement and assessment, leading them to rely on traditional methods such as multiple-choice and true/false items, while largely neglecting alternative assessment strategies that emphasize performance and critical thinking (Alnahdi, 2014; Al Khawaldeh *et al*., 2022).

These findings underscore the urgent need to strengthen teachers' capacity to design fair and effective achievement tests. Doing so requires equipping them with both theoretical knowledge and technical skills aligned with psychometric standards—content validity, reliability, content coverage, and cognitive level diversity. Bloom's Taxonomy remains a widely accepted educational framework to ensure cognitive variety in test design, organizing learning objectives into six hierarchical levels from recall to creativity (Brookhart, 2011). Studies have shown that in the absence of sufficient training, most teachers disproportionately focus on the first two levels—recall and understanding—while neglecting higher-order thinking tasks like analysis and evaluation (Çalık & Gündüz, 2022).

Content analysis of mathematics tests reveals recurring construction errors by teachers, such as vague wording, imbalanced grading, and over-reliance on copied items from textbooks or online sources. Al-Amir (2011), for example, identified five common flaws in multiple-choice item construction, including ambiguous phrasing, verbatim copying, redundancy in distractors, illogical sequencing, and overly obvious correct answers. The study recommended targeted teacher training to help educators avoid these pitfalls and improve their ability to create items that genuinely assess cognitive skill (Alawneh *et al*., 2023 – ICACITE).

Moreover, program evaluation research indicates that developing teachers' assessment competencies cannot rely on theoretical instruction alone; it demands hands-on involvement in designing, piloting, and analyzing assessments. This supports the call for instructional approaches grounded in project-based learning, practical application, and formative feedback. The literature advocates for the use of authentic tasks as a model for training teachers to build meaningful assessment tools that reflect the realities of classroom teaching (Pastore, 2023).

In this light, specialized university-level courses in educational assessment are essential components of teacher preparation programs, particularly in mathematics. The "Assessment of Mathematics Education" course at An-Najah National University is one such example. This course is designed to provide master's and diploma students in mathematics education with both theoretical grounding and practical experience in test construction. It emphasizes key skills such as developing test blueprints, crafting varied question types, and analyzing the psychometric properties of assessments (An-Najah National University, 2024).

Despite the learning opportunities this course offers, important questions remain regarding its actual effectiveness in enhancing students' practical competencies, and the challenges they encounter while learning to construct assessments. Recent research has increasingly focused on documenting pre-service teachers' lived experiences in such courses to better understand what they truly learn and the obstacles they face in applying that knowledge (Alawneh, Shadid, & Salman, 2024).

Findings from these qualitative studies reveal that while students gain a solid theoretical understanding, they often struggle to implement this knowledge in constructing a comprehensive and valid test. Çalık & Gündüz (2022) argue that overcoming these barriers requires a learning environment that promotes repeated practice, constructive feedback, and critical analysis of real test samples—fostering what is known as an "assessment culture" among future educators.

Hence, examining the lived experiences of students enrolled in the "Assessment of Mathematics Education" course at An-Najah National University—focusing on their practical engagement, encountered difficulties, and perceived value of the content—is a crucial step toward enhancing the course and its outcomes, and ultimately improving teachers' capacity to build meaningful and effective assessments.

Research Problem and Questions

Despite growing global and local attention to pre-service teachers' competencies in achievement test construction, this remains one of the most complex and challenging skills in teacher education—particularly in the domain of mathematics education, which demands high levels of precision and analytical rigor (Brookhart, 2011; Sahin, 2023). Educational research indicates that teaching an assessment course at the university level—

though necessary—is not sufficient to ensure a lasting transfer of theoretical knowledge into sound practice unless accompanied by experiential learning and expert feedback (Pastore, 2023; Alawneh, Shadid, & Salman, 2024).

In the Palestinian context, and more specifically at An-Najah National University, the "Assessment of Mathematics Education" course is a core component of mathematics teacher preparation programs. It aims to equip students with the theoretical foundations and practical skills required to design valid and reliable achievement tests. Although the course covers key competencies such as blueprint design, objective classification, and data analysis, its actual effectiveness in fostering the desired competencies remains uncertain, especially given the scarcity of qualitative research that directly investigates students' own learning experiences.

A review of the literature reveals that many graduates of such courses continue to exhibit signs of "assessment illiteracy," evident in poorly formulated questions, weak data analysis skills, and limited use of diverse assessment techniques (Zhang & Burry-Stock, 2003; Çalık & Gündüz, 2022). Other Arab studies have confirmed that constructing a complete and effective test remains a significant challenge for teachers—even after completing an assessment course (Al Khawaldeh *et al*., 2022). The core issue lies in the disconnect between the theoretical components of the course and their real-world application—especially when students lack proper guidance during the test development process or are assessed based solely on surface-level criteria.

This gap is particularly critical in graduate programs, where students are expected to have achieved a high level of academic maturity. Yet, whether they have truly acquired the competencies required to build valid assessments—or whether they continue to face conceptual and practical barriers—remains an open question. Addressing this question requires a qualitative approach that gives students the space to share their experiences and recommendations in depth.

Accordingly, this study seeks to investigate the lived experiences of graduate students in the "Assessment of Mathematics Education" course at An-Najah National University with regard to test construction. The aim is to understand what they have learned, what obstacles they have encountered, and what suggestions they propose for improvement. This inquiry is vital because these students are future educators and educational leaders. Understanding how they learn such a foundational skill can help improve teacher preparation programs and address significant gaps highlighted in the literature (Pastore, 2023; Alawneh *et al*., 2023).

In this context, the study aims to answer the following key research questions:

1. What educational experiences have students in the "Assessment of Mathematics Education" course undergone in relation to constructing achievement tests?
2. What are the main challenges and difficulties they encountered while learning and practicing test construction within the course?
3. What recommendations do students offer for improving the teaching and evaluation of test construction skills in the course?

Study Objectives

Based on the problem and research questions, this qualitative study aims to:

- Describe and analyze students' learning experiences in the "Assessment of Mathematics Education" course with regard to test construction.
- Identify the key challenges students face while acquiring test construction skills, including difficulties in understanding theoretical content or applying it practically (e.g., blueprint development, question formulation at different cognitive levels, data analysis, etc.).
- Derive recommendations for improving the teaching and evaluation of test construction skills for future teachers—based on students' own feedback and the study's findings—to enhance the effectiveness of the course and broader teacher education programs in the domain of educational assessment.

Significance of the Study

This study addresses a vital aspect of mathematics teacher preparation: the development of assessment competencies—particularly the skill of constructing achievement tests. The study's significance is twofold:

Theoretical significance:

It contributes to the educational literature on mathematics assessment by offering a deeper understanding of pre-service teachers' experiences in learning test construction. While most existing studies have taken a quantitative approach—such as measuring teachers' test construction skills or cataloging common errors—this study adopts a qualitative lens that captures students' perspectives and the challenges they encounter in real time. Thus, it adds a dimension of insight that numbers alone cannot provide. To the researcher's knowledge, this is the first study focused specifically on a single course at An-Najah National University, paving the way for similar research in other institutional settings.

Practical significance:

The study's anticipated findings offer directly actionable insights for improving teacher preparation

programs. By identifying strengths and weaknesses in students' experience with the assessment course, instructors and program designers can make informed adjustments to content and pedagogy. For example, if the study reveals widespread difficulty in designing test blueprints or crafting higher-order questions, it may justify expanding practical training in those areas. Educational supervisors and in-service mathematics teachers may also benefit, as the study highlights areas that may need targeted professional development (e.g., avoiding common item-writing errors or receiving specialized assessment training). Ultimately, enhancing teachers' assessment skills improves the overall quality of education, as well-designed tests guide instruction and motivate students to achieve educational goals more effectively.

Study Delimitations

To clarify the scope of the study and define the boundaries for generalizing its findings, the following delimitations are acknowledged:

- **Topical Delimitation (Subject Focus):** This study is limited to the topic of constructing achievement tests within the framework of the "Assessment of Mathematics Education" course. It does not address other domains of educational assessment (such as alternative assessment methods like projects or observations) except insofar as they relate to students' experiences in the course. The primary focus is on how students engaged in designing tests that align with the principles of sound test construction, as taught both theoretically and practically.
- **Human Delimitation (Participants):** The study focuses on graduate students at An-Najah National University enrolled in the "Assessment of Mathematics Education" course. More specifically, the sample consisted of ten students (male and female) who had completed the course during the specific semester in which the research was conducted. These participants were purposefully selected (non-random sampling) to reflect variation in gender and teaching experience.
- **Spatial Delimitation (Location):** The study was conducted at An-Najah National University, located in Nablus, Palestine, within the Faculty of Graduate Studies, specifically the Mathematics Education Program.
- **Temporal Delimitation (Timeframe):** The research was carried out during the first semester of the 2024/2025 academic year.

Operational Definitions of Terms

To ensure clarity and consistency within the context of this study, several key terms are defined operationally as follows:

Assessment of Mathematics Education

- ***Theoretical Definition:*** Mathematics education assessment refers to a systematic and ongoing process of collecting and analyzing data to make educational decisions related to student learning in mathematics. It includes measuring achievement, diagnosing learning difficulties, and evaluating the effectiveness of teaching methods and curricula (NCTM, 2014; Brookhart, 2011).
- ***Operational Definition:*** In this study, the term refers to a graduate-level university course (e.g., MATH 602) offered at An-Najah National University's Faculty of Graduate Studies. The course addresses topics in educational measurement and assessment in mathematics, emphasizing strategies for evaluating student learning, constructing achievement tests, developing test blueprints, analyzing test items, and validating them within an applied academic context.

Achievement Tests

- ***Theoretical Definition:*** Achievement tests are assessment tools designed to measure the extent to which students have acquired knowledge and skills specified in the curriculum. They are typically linked to observable and measurable instructional objectives (Popham, 2017; Nitko & Brookhart, 2014).
- ***Operational Definition:*** In this study, the term refers to the tests that students in the "Assessment of Mathematics Education" course were required to design as part of their coursework. These include various item types (essay questions, multiple-choice, true/false) developed in accordance with the assessment principles they learned, such as comprehensiveness, cognitive level diversity, clarity, content validity, and reliability.

Students' Experiences (in Test Construction)

- ***Theoretical Definition:*** Experience is defined as the individual's interaction with educational or life situations, resulting in emotional, cognitive, and behavioral reflections and insights (Kolb, 1984; Dewey, 1938).
- ***Operational Definition:*** In this study, the term refers to the personal and academic experiences of graduate students during the "Assessment of Mathematics Education" course—particularly in relation to the tasks involving the construction of achievement tests. These experiences include what they learned conceptually and skill-wise, the practical tasks they performed (e.g., drafting questions, building a test, peer review), their emotional responses to these activities, and their self-

evaluation of how beneficial the experience was.

METHODOLOGY

Research Design

This study employed a qualitative exploratory approach, which was deemed appropriate given the research's aim to understand how students perceive and experience a specific phenomenon—namely, learning and practicing test construction. This approach allows for a deep exploration of individual narratives and nuances that are often inaccessible through quantitative methods. Specifically, the study used a qualitative descriptive design that seeks to portray the phenomenon within its natural context without manipulation by the researcher. The focus was on analyzing the language and expressions used by participants to derive meaningful interpretations. This methodology supports giving voice to the students and enabling them to freely articulate their thoughts and perceptions of the course.

Population and Sample

The target population for this study included all graduate students enrolled in the "Assessment of Mathematics Education" course within the master's and diploma programs at the Faculty of Educational Sciences and Teacher Preparation at An-Najah National University. According to academic records, this course is typically offered to students majoring in Mathematics Teaching Methods or Educational Mathematics, with average enrollment ranging from 15 to 20 students per semester.

The sample was selected using purposive sampling. Ten students who had completed the course during the designated semester were included. In selecting participants, diversity was intentionally considered to enrich the data: the sample comprised both male and female students, as well as recent graduates and those concurrently engaged in teaching practice. This deliberate sampling was designed to capture a wide range of experiences and perspectives.

The researcher contacted all students enrolled in the course and invited them to voluntarily participate in the interviews. Those who agreed (a total of 10 students) were included after signing informed consent forms. It is important to note that the sample size in qualitative research is not intended for statistical generalization but for achieving data saturation. In this study, saturation was observed by the tenth interview, indicating that the chosen sample size was sufficient to capture the necessary information.

Data Collection Instrument

This study employed a semi-structured interview protocol as its primary tool for data collection. The researcher developed an interview guide comprising a set of open-ended questions and main discussion points derived from the research questions and objectives. The interview prompts focused on the student's overall experience in the course, what they learned regarding test construction, their evaluation of the benefits gained, the challenges or difficulties they encountered while designing an achievement test as part of the course activities, how they managed these challenges, and their suggestions for improving the course and supporting students in acquiring test construction skills.

To ensure content validity, the interview questions were reviewed by a panel of three experts in mathematics education and assessment. These specialists evaluated the clarity, relevance, and comprehensiveness of the questions in relation to the study's objectives. Based on their feedback, minor revisions were made, such as rephrasing certain questions for clarity and adding a sub-question to address a point that was initially underrepresented. Through this process, both face and content validity of the interview instrument were established.

As for the reliability of the tool (i.e., consistency of findings upon repeated use), it was ensured through standardized interview procedures and verification of data analysis consistency. The researcher recorded all interviews—after obtaining informed consent from participants—and transcribed them verbatim. To enhance the reliability of qualitative interpretation, an external qualitative analyst (a peer researcher in the same field) independently reviewed and coded the full set of ten transcripts using the same analytical framework. The researchers then compared their codes and interpretations, and inter-rater reliability was calculated using Holsti's formula, a commonly used method in content analysis. The resulting agreement rate was approximately 85%, which is considered an acceptable threshold for reliability in qualitative research (as literature generally considers a rate of 0.80 or higher sufficient). This process helped ensure that the results were robust and impartial, minimizing potential researcher bias.

Procedures for Data Collection and Analysis

Once the final version of the interview guide was approved, the researcher proceeded with field data collection. Individual appointments were scheduled with each participant, and interviews were conducted either in a quiet and suitable space on campus (such as a meeting room or unused classroom) or via a secure online platform in cases where physical presence was not feasible. Each interview lasted approximately 30 to 45 minutes.

At the beginning of each session, the researcher engaged in a brief introductory conversation to build rapport, clarify the purpose of the interview, and emphasize the value of honesty. Participants were reassured that there were no "right" or "wrong" answers, and that their responses would be kept confidential and used solely for research purposes.

Following data collection, the interviews were analyzed through the following systematic steps:

1. **Transcription and Data Organization:** The researcher transcribed each recorded interview verbatim and created a separate document for each. These transcripts were carefully reviewed for accuracy, with any transcription errors or ambiguities corrected.
2. **Immersive Reading and Initial Notes:** The researcher conducted multiple readings of the interview texts to develop a deep understanding of the content. Initial impressions and preliminary notes were recorded for each transcript, facilitating early identification of recurring ideas and emergent themes.
3. **Coding:** Using thematic analysis, the researcher systematically coded the data. This involved identifying phrases or statements with meaningful relevance to the research questions and assigning concise labels or codes to them. For instance, segments indicating a gained skill were labeled as "Skill Gained," while segments describing difficulties were coded as "Challenge/Difficulty X" based on their type.
4. **Theme Development:** Once all transcripts were coded, the researcher grouped related codes into broader thematic categories. For example, codes concerning difficulties with question formulation, test blueprinting, or time management were consolidated under a single theme titled "Challenges in Test Construction." Similarly, codes related to new conceptual understandings or enhanced skills were clustered under "Learning Gains from the Course."
5. **Analysis and Alignment with Research Questions:** The themes were reviewed to ensure clarity and relevance, then organized in a way that directly addressed the study's core questions. Given the structure of the research questions—which focused on experiences and benefits, challenges, and recommendations—the results section was structured accordingly around these three pillars. The analysis also incorporated simple frequency counts (e.g., the percentage of participants reporting a specific view) and included direct quotations from participants to vividly illustrate key findings.
6. **Reliability and Validity of Conclusions:** After formulating the preliminary findings, the researcher revisited all transcripts to verify that the results accurately reflected the data without distortion or selective bias. Furthermore, the findings were discussed with the external analyst who had participated in the coding process, allowing for a collaborative review of the interpretations. Consensus was reached on most points, and where discrepancies arose, the original transcripts were consulted to resolve them based on the actual data.

By following these rigorous procedures, the study ensured that its findings were grounded in a methodical and thorough analysis of the qualitative data, thereby reinforcing the credibility of the conclusions and their alignment with participants' authentic perspectives.

STUDY RESULTS

This section presents the core findings derived from the analysis of interviews conducted with the ten participating students. As previously stated, the results are organized according to the three main research questions: (1) students' experiences and perceived learning benefits; (2) challenges encountered during test construction; and (3) their recommendations for improving the course. Each subsection includes approximate percentages representing the prevalence of specific responses across the sample (out of 10 participants), supported by anonymized direct quotations (e.g., "Student 5") to preserve privacy. The discussion also integrates relevant literature to assess the alignment or divergence of these findings with previous studies.

Findings Related to Students' Positive Experiences and Perceived Learning Gains

Research Question 1: What positive educational experiences and benefits did students gain from the course in relation to achievement test construction?

The interviews revealed that the majority of participants believed the course substantially enhanced their theoretical understanding of educational assessment and their practical skills in test construction. Eight out of ten participants (80%) explicitly stated that the course provided them with new insights or corrected misconceptions they had previously held about test development. As Student 3 explained:

> *"Before taking this assessment course, I used to think that preparing a test just meant gathering questions from the textbook and printing them out. Now I've learned that it's much deeper than that—I've learned how to build a table of specifications to ensure test balance and how to make sure the questions address different cognitive levels, not just definitions."*

Seventy percent of the students reported improved ability to write questions targeting higher-order thinking skills, aided in part by their use of Bloom's Taxonomy. Student 7 shared:

> *"I've realized how important it is for an exam to include questions that ask for analysis, synthesis, and application—not just rote recall. In my course project, I tried to include an application-based question and one that asked students to analyze a math scenario. That's something I never would have done before, because most exams I've seen focus on routine problems."*

This aligns with the broader trend found in studies on teacher assessment competency, which suggest that targeted training—through coursework or

workshops—improves educators' ability to develop high-quality assessments. For example, Al-Zabon (2013) found a statistically significant improvement in test construction competencies among teachers who had taken an assessment-focused course compared to those who had not. Similarly, our participants reported feeling more confident about applying quality criteria such as content validity, reliability, and comprehensive coverage. As Student 1 expressed:

"Now I have a mental checklist for evaluating test quality. I can actually pinpoint why a test is poorly constructed—like if it lacks a test blueprint or has an unfair grade distribution."

One practical learning experience mentioned by nearly all participants was the requirement to design an actual achievement test as part of the final project. Students unanimously agreed that this task was a pivotal learning moment. Student 5 reflected:

"Designing the final test took a lot of time and effort, but I learned so much. For the first time, I imagined myself as a real teacher with a classroom, responsible for creating a fair and comprehensive exam. It wasn't easy, but it was rewarding and motivating."

Six students noted that receiving feedback from the instructor and classmates on their test drafts helped clarify their understanding of assessment standards. After each student submitted a test, the entire class reviewed and discussed it collectively, allowing them to identify strengths and areas for improvement—an exercise that reinforced practical application of assessment criteria.

Overall, all ten participants (100%) agreed that the course enriched their theoretical knowledge of assessment and provided hands-on skills in test design. This positive impact reflects recommendations in contemporary educational literature, which stress the importance of including substantial practical training in teacher preparation programs. For example, a recent systematic review by Pastore (2023) emphasized the growing commitment among teacher education programs to develop a strong "assessment culture" among pre-service teachers—including training on how to use assessment results to support student learning. In this regard, the experience of our participants can be seen as a promising case of enhanced assessment competency. Four students explicitly mentioned increased self-confidence. As Student 2 noted:

"I used to feel anxious at the thought of designing exam questions on my own. But now I'm actually excited to try it during my teaching practicum. I think I can create a well-balanced exam for my students in the future."

Findings Related to the Challenges Faced by Students

Research Question 2: What were the main challenges and difficulties students encountered while learning and practicing test construction within the course?

Despite the positive experiences described above, the analysis revealed a number of challenges that students faced during their learning journey. These challenges fell into two general categories: cognitive/technical difficulties related to the design process itself, and emotional/organizational barriers linked to students' confidence, time management, and workload. The most prominent difficulties, with the proportion of students who mentioned them, are detailed below:

- **Difficulty in Constructing the Table of Specifications:** This was one of the most commonly reported challenges, cited by 6 out of 10 participants (60%). These students found the process of aligning test items with content and objectives, as well as balancing cognitive levels, to be complex. As Student 8 stated:

  "Even though I understood the concept in theory, when I tried to build a table of specifications for a geometry unit, I struggled to decide how many questions to allocate per objective and how to distribute them across Bloom's levels. I needed help from the instructor to get it right."

  This finding is consistent with previous research that highlights difficulties even among in-service teachers in constructing balanced and comprehensive test blueprints. Mastery of this skill requires time and guided practice.
- **Struggles with Writing Higher-Order Thinking Questions:** Five participants indicated they found it difficult to formulate analytical or application-based questions, even though they understood their importance. As Student 6 put it:

"I know what analysis or problem-solving questions look like in theory, but when I tried to write one, I either made it too hard or too vague. It was just easier to stick with traditional problems."

This mirrors challenges frequently reported by novice teachers, who often default to lower-order questions due to the complexity of writing higher-order items—and fears that students will not be able to answer them. Student 4 commented:

"I was worried that if I asked a deep-thinking question, the students would all get it wrong. That's probably why we tend to write direct questions—to avoid confusing them."

These observations point to a psychological barrier that may require a supportive culture that encourages experimentation and iterative learning.

- **Difficulty with Linguistic Clarity and Precise Wording:** Four students mentioned struggling

to write clear, grammatically sound questions in either Arabic or English. One issue was avoiding verbatim copying from textbooks. Student 1 noted:

> *"I realized that question writing is both a science and an art. Sometimes I'd write something and later notice it sounded exactly like an example from the book. I also worried whether my wording was clear or could be misinterpreted by students."*

- **Estimating Test Length and Timing:** Three students reported difficulty deciding how many questions to include and how much time students would need to complete them. Student 9 shared:

> *"I wasn't sure if the test should take one or two class periods, and that made it hard to decide how many questions to include. I didn't know how long each type of question might take."*

Such judgment typically develops with classroom experience, which most students lacked.

- **Time Constraints Due to Other Coursework:** Approximately five students expressed feeling overwhelmed by the demands of the test project coinciding with other assignments. Student 10 remarked:

> *"We had other major projects at the same time. I wanted to create a high-quality test, but time was limited. Maybe with more time, I could have included more variety in my questions."*

- **Lack of Prior Teaching Experience:** Only two participants had teaching experience during their studies. These individuals reported fewer difficulties in test design, as they could envision real student responses. As one shared:

> *"Having taught actual students for two years, I could imagine how they might react to my questions. That helped me write more realistic items."*

The rest relied solely on theoretical assumptions, which may have limited their ability to assess difficulty levels. This is consistent with findings from Sahin (2023), who observed minimal differences in theoretical knowledge between pre-service and in-service teachers, though experience gave a slight edge in application.

In summary, the challenges reported by students in constructing achievement tests mirror those documented among practicing educators—such as difficulties with blueprinting, writing higher-order questions, and avoiding vague phrasing. While the course included hands-on practice, it may not have been sufficient to internalize the skill. Additionally, psychological apprehension and workload stress emerged as non-trivial factors. Nevertheless, students showed clear awareness of these challenges and a genuine desire to overcome them. As Student 7 put it:

> *"Yes, I faced some technical challenges, but I know it's just a matter of practice. I hope we'll have more opportunities to train—and maybe some extra materials to guide us. I'm sure we can get through these difficulties."*

Findings Related to Students' Suggestions for Improving Test Construction Training

Research Question 3: What recommendations do students propose for enhancing the teaching of test construction and improving learning experiences in this area?

At the end of each interview, participants were invited to share any suggestions they believed would improve the course or better support students in mastering the skill of test construction. Their proposals were notably consistent and demonstrated thoughtful awareness of their actual learning needs. The key suggestions are summarized below, along with the number of students (out of 10) who proposed each:

- **Increase Practical, Hands-On Activities in the Course:** This was the most frequently mentioned recommendation, cited by 8 of the 10 participants. Students emphasized the importance of allocating more class time to practical training in writing test items and designing assessments. Instead of relying solely on a single final project, they suggested diversifying practical tasks throughout the semester—such as regularly creating short questions, analyzing previous test samples, and peer-reviewing classmates' items. Student 5 remarked:

> *"I think we need more in-class practice. Maybe we could dedicate part of each lecture to individual or group exercises where we write or critique questions. That would help us feel more prepared before the final project."*

This recommendation aligns with international studies emphasizing the need to move theoretical instruction in teacher preparation toward more practice-oriented and authentic representations of real classroom tasks.

- **Connect Training to Real Classroom Contexts:** Five participants proposed including elements that directly link their learning to the realities of school-based assessment. Suggestions included organizing a field visit or inviting an experienced teacher to discuss how tests are designed in actual school settings, or analyzing real test samples from local mathematics classrooms. Student 2 noted:

"If we could examine real school tests and discuss whether they meet the standards, that would be very helpful. Even better, we could hear from a seasoned teacher about how they build their tests."

This would help narrow the perceived gap between theory and practice, particularly in estimating students' skill levels and managing test duration.

- **Provide Individualized Feedback and Guidance:** Three students highlighted the value of receiving detailed, personalized feedback on their test projects. They suggested scheduling one-on-one sessions with the course instructor or a teaching assistant to review strengths and areas for improvement. Student 4 explained:

  "When I met with the instructor after submitting my project and we reviewed my questions together, I learned very specific things that we wouldn't have noticed in group discussions."

  Such targeted guidance can accelerate skill development and address misunderstandings early on.
- **Supplement with Enrichment Materials and Additional Examples:** Four participants suggested making supplementary learning resources available outside of class. These could include a handbook explaining test construction step-by-step with practical examples and well-constructed sample questions, or links to relevant articles and training videos. Some also recommended creating a question bank or repository of high-quality test examples. As Student 8 put it:

  "If we had access to sample tests that are well-designed, it would help us a lot. We could draw ideas and learn how cognitive levels are distributed in a real test."

  This echoes Al-Amir's (2011) recommendation that teachers be trained using clear checklists and practical examples to avoid common errors and improve test design literacy.
- **Improve the Course Timeline and Scheduling of Assignments:** Two participants made administrative suggestions related to the timing of the final project. They recommended breaking the assignment into phases—for example, submitting a draft of the table of specifications mid-semester for formative feedback, followed by the full test at the end. They also proposed avoiding overlaps between the project and final exams in other courses. Such suggestions can help reduce stress and improve the learning experience. As one student noted:

  "We had so many projects due at the same time. If the workload were spaced out, we could focus more and produce better tests."

Overall, these recommendations reflect a strong sense of self-awareness among the students regarding the types of support and structures that would best help them develop assessment skills. Their suggestions are closely aligned with modern approaches in teacher education, which emphasize authentic practice, feedback-rich environments, and sustained engagement with core teaching competencies rather than one-time assignments.

It is also worth noting that several students, when asked for final remarks, expressed appreciation for the instructor's efforts and teaching style. As Student 1 shared:

"Honestly, the instructor did a great job. His approach was effective. But of course, there's always room to improve the content."

This reflects that the suggestions were not made in a critical spirit, but rather as sincere contributions aimed at deepening learning and enhancing future course offerings.

DISCUSSION OF FINDINGS

Through an in-depth analysis of the interviews conducted with ten students enrolled in the "Assessment of Mathematics Education" course at the Faculty of Graduate Studies at An-Najah National University, five key findings emerged—each carrying significant pedagogical implications and aligning closely with trends identified in contemporary educational literature:

1. Notable Improvement in Assessment Awareness and Competency

Most students reported a qualitative shift in their understanding of test construction. They described moving from a superficial and random conception of testing to a more structured understanding of how to construct a high-quality achievement test—starting from designing a table of specifications, identifying learning objectives and cognitive levels, and ending with writing valid, well-structured items.

This improvement resonates with the findings of Alawneh *et al*. (2024), who demonstrated that well-designed assessment courses significantly enhance teachers' professional preparation, particularly when accompanied by practical application. Brookhart (2011) also emphasizes that training teachers in assessment principles is one of the strongest predictors of their future teaching quality and has a direct impact on student learning outcomes.

In this context, the course appears to have helped bridge a critical knowledge gap. Several participants expressed greater confidence in constructing tests and applying criteria such as content validity, reliability, and comprehensiveness:

> *"I used to think tests were just random questions we wrote. Now I understand that every item should have a purpose, a thinking level, and a place in the test blueprint."* — (Student Quote)

2. Continued Challenges in Complex Aspects of Test Construction

Despite this progress, many students noted ongoing difficulties with more advanced aspects of assessment—particularly writing higher-order thinking questions, achieving cognitive balance in test blueprints, and crafting questions with precise, unambiguous language.

These findings are consistent with Sahin (2023), who reported that even in-service mathematics teachers often lack sufficient competence in designing items that target complex cognitive skills. Similarly, Al-Amir (2011) identified common flaws in multiple-choice item design, such as ambiguity and weak distractors—issues echoed by participants in this study. This supports the broader claim that assessment skills cannot be fully developed in a single semester but require continued practice and ongoing feedback.

> *"I feel I now understand how to start building a test, but when I tried writing an analysis or evaluation question… it was really hard to phrase it properly."* — (Student Quote)

3. Practical Activities Deepened Understanding

Students unanimously agreed that practical tasks—such as designing a complete test, reviewing a peer's assessment, and analyzing questions—contributed more to their learning than theoretical instruction alone.

This finding is backed by the literature on task-based learning. Çalık & Gündüz (2022) found that applying skills in authentic or realistic settings leads to deeper learning and better retention of assessment concepts.

> *"I learned the most from the practical assignment. When I wrote the questions myself, I understood how precise they really need to be."* — (Student Quote)

4. Influence of Context and Prior Background

The interviews revealed noticeable individual differences in students' ability to grasp concepts and apply assessment skills. Those with prior teaching experience appeared more confident in linking theory to real classroom situations, while others struggled to visualize how concepts would function in practice. Additionally, some participants noted that their limited linguistic background made it harder to phrase questions accurately.

This supports Pastore's (2023) call for university programs to accommodate individual differences and offer targeted support for students who need it. Brookhart (2011) likewise suggests that building assessment competencies requires a differentiated, not one-size-fits-all, approach.

5. Alignment of Student Suggestions with Best Practices

Students proposed several thoughtful suggestions to improve the course, including more opportunities for hands-on practice, access to guided examples, increased use of feedback, and stronger links between theory and real-world application.

These recommendations align with global research findings. For example, a 2023 report in *Frontiers in Education* emphasized that assessment skills should be developed through practical tasks that include structured feedback and reflection. Similarly, Al Khawaldeh *et al*. (2022) highlighted a lack of training among Jordanian and Palestinian teachers in evaluating the quality of their test items, supporting the need for applied programs like the one suggested by the participants.

Ultimately, this study confirms that assessment literacy is a complex skill that cannot be acquired through theory alone. It must be developed through practice, reflection, and meaningful interaction with feedback. Importantly, this qualitative research offers an authentic platform for students' voices—contributing to curriculum development from within the learning process.

While the sample size was limited to ten participants, the findings serve as a valuable starting point for future research. Quantitative studies could further examine the course's impact by measuring students' competencies before and after the course or evaluating the psychometric quality of tests they construct.

Recommendations

Based on the findings and discussion, the researcher offers the following practical recommendations for improving the teaching of test construction within the "Assessment of Mathematics Education" course and similar teacher training programs, as well as suggestions for future research:

- **Increase Practical Activities Throughout the Semester:** The course should include ongoing, small-scale assignments where students build test items or short question sets regularly—not just a single final project. In addition, the course content could integrate a critical analysis activity where students evaluate real-world tests (e.g., from national exams or schools) using assessment criteria. This would enhance

students' authentic learning experience and contextual awareness.

- **Organize Individual or Small Group Coaching Sessions:** To support students more effectively, instructors (or teaching assistants) could provide two individual feedback sessions per student—one after the draft blueprint and another after drafting the test questions. These sessions can help identify subtle design issues that might not surface in group discussions.
- **Develop a User-Friendly Assessment Handbook or Checklist:** A practical guide outlining the steps and standards of good test design—along with common mistakes to avoid—should be distributed to all students. This could be accompanied by well-constructed sample tests for reference, as well as flawed examples annotated with explanations, for training purposes.
- **Incorporate Realistic Assessment Simulations:** Whenever possible, students should be given opportunities to test their assessments in semi-authentic settings—for example, administering a quiz they designed to classmates or to real students during a supervised field visit. Where field testing is not feasible, simulated case studies can be used to replicate classroom contexts and generate discussion.
- **Improve Course Scheduling and Assessment Load Distribution:** Course coordinators should avoid clustering major project deadlines across different courses within the same timeframe. For the assessment course specifically, tasks should be distributed throughout the semester. Moreover, if students are engaged in field training, the test construction project could be partially integrated into that practicum, allowing them to design tests for actual learners under joint supervision.
- **Offer Continuous Professional Development for In-Service Teachers:** Faculties of education and educational authorities should regularly organize training workshops for practicing teachers focused on assessment skills. Test design training should not be limited to university coursework. Given the ongoing curriculum reforms and the shift toward higher-order thinking assessments, continuous professional development in this area is essential.
- **Support Teachers with Evaluation Tools:** Teachers should be provided with self-assessment tools such as quick-check review lists to help them evaluate the quality of their own tests. These tools might include prompts related to content coverage, cognitive diversity, language clarity, and grade distribution.

CONCLUSION

This study reinforces the importance of test construction as a core competency for any effective teacher. As such, it deserves focused attention in both pre-service teacher education and in-service professional development. While the experiences of students in the "Assessment of Mathematics Education" course at An-Najah University revealed encouraging progress in assessment literacy, they also highlighted areas in need of further support and refinement.

Implementing the above recommendations—across curriculum design, instructional strategy, and institutional policy—can contribute to producing teachers who are more competent in fairly and effectively evaluating student learning. A teacher who masters assessment is one who can accurately diagnose student understanding and guide them toward meaningful growth—and that is the ultimate goal of education.

REFERENCES

- Alawneh, Y., Al-Momani, T., Salman, F., Alkhwaldeh, A., Al-Dlalah, M., Kaddumi, T. (2023). The state of musically gifted students in Palestine: a case study, *Res Militaris*,13(2). 2058-2069.
- Zhang, Y., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323–342. https://doi.org/10.1207/S15324818AME1604_4
- Alawneh, Y., Sleem, H., Al-Momani, T., Salman, F,. Al-Dlalah, M., Kaddumi, T., Kharashqah, w.(2023) Strategic Pioneering And Its Connection To Faculty Members' Administrative Creativity At Palestinian And Jordanian Universities, *Journal of Namibian Studies*,34(Special Issue 1),808-828.
- Alawneh, Y. (2022). Role of Kindergarten Curriculum in Instilling Ethical Values among Children in Governorates of Northern West Bank, Palestine, *Dirasat: Educational Sciences*,49(3),360-375.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills required in teacher assessment literacy standards. *Educational Measurement: Issues and Practice, 30*(1), 3–12. https://doi.org/10.1111/j.1745-3992.2010.00195.x
- Çalık, T., & Gündüz, M. (2022). Pre-service mathematics teachers' competence in assessment and evaluation: A qualitative study. *Participatory Educational Research, 9*(4), 281–301. https://doi.org/10.17275/per.22.89.9.4
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.
- Güler, G. (2023). Investigating pre-service mathematics teachers' assessment literacy and perspectives on alternative assessments.

International Journal of Curriculum and Instruction, 15(1), 303–316. https://ijci.wcci-international.org/index.php/IJCI/article/view/867

- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Addison-Wesley.
- Pastore, S. (2023). Preservice teachers' assessment literacy: A systematic literature review (2012–2022). *Assessment in Education: Principles, Policy & Practice, 30*(4), 462–481. https://doi.org/10.1080/0969594X.2023.2177384
- Sahin, I. (2023). Assessment literacy of mathematics teachers: A comparative analysis of pre-service and in-service teachers. *International Online Journal of Education and Teaching (IOJET), 10*(2), 590–608. https://iojet.org/index.php/IOJET/article/view/1976
- Alawneh,Y. Al-Momani,T., Salman,F,. Al-Ahmad,S,. Kaddumi,T., Al-Dlalah,M. (2023). The Extent of the Prevalence of Pronunciation Problems among Students of the First Primary Stage in the Point of View of their Teachers and Treatment Methods, *Educational Administration: Theory and Practice*,29(3),19-33.
- Alawneh,Y., Al-Momani,T., Salman,F., Kaddumi,T.,Al-Dlalah,M. (2023). A Detailed Study Analysis of Artificial Intelligence Implementation in Social Media Applications.*2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* DOI: 10.1109/ICACITE57410.2023 12-13 May 2023.
- Alawneh,Y. Al-Momani,T., Salman,F,. Al-Ahmad,S,. Kaddumi,T., Al-Dlalah,M. (2023). The Extent of the Prevalence of Pronunciation Problems among Students of the First Primary Stage in the Point of View of their Teachers and Treatment Methods, Educational Administration: Theory and Practice,29(3),19-33.
- Alawneh,Y. Shadid,R. Salman,F(2024)T he Effectiveness Of The Self-Questioning Strategy In Developing Scientific Research Skills Among Students Of The Faculty Of Education At Israa University,Educational Administration: Theory and Practice,30(5),810-825-33.
- Alawneh,Y., Al-Momani,T., Salman,F., Alkhwaldeh,A., Al-Dlalah,M., Kaddumi,T. (2023).The state of musically gifted students in Palestine: a case study, *Res Militaris*,13(2). 2058-2069.
- Abu Shkheedim,S., Alawneh,Y., Khuwayra,O.,Salman,F., khayyat,T.(2022). The Level Of Satisfaction Of Parents Of Students With Learning Difficulties Towards Distance Learning After The Corona Pandemic, *NeuroQuantology*,20(19),1299-1311.
- Alawneh,Y., Ashamali,M., Abdel-Hassan,R., Al-khawaldeh,S., Engestroom,y.(2022) Degree Of Use Of E-Learning Science Teachers In Public High Schools In During The Corona-Covid 19 Pandemic, *Journal of Positive School Psychology*,6(2), 1060-1070.
- ALrashidi,N,. Sahib,R,. Alawneh,Y,.Alawneh,A. (2023). Post-Pandemic Higher Education: Arabic Universities, *Elementary Education Online*,22(2),1-11.
- Al-Ahmad,S., Al-Dlalah,M., Al-Momani,T., Barakat,S., Kaddumi,T., Alawneh,Y,. Al Zboun,M.(2023) effectiveness of e-learning in Palestinian and Jordanian universities from the viewpoint of faculty members Perspective, *Journal of Southwest Jiaotong University* ,58(1),463-472.
- Alawneh,y., abu shokhedim,s., al-khazalah.f.(2022). Trends of teachers with handicapped towards the e-learning program at basic education in schools during covid-19, *journal of positive school psychology*,7(6)1876-1886.
- Abu Shkheedim,S., Alawneh,Y., Khuwayra,O.,Salman,F., khayyat,T.(2022). The Level Of Satisfaction Of Parents Of Students With Learning Difficulties Towards Distance Learning After The Corona Pandemic, *NeuroQuantology*,20(19),1299-1311.
- Alawneh,Y.(2022). Role of Kindergarten Curriculum in Instilling Ethical Values among Children in Governorates of Northern West Bank, Palestine, *Dirasat: Educational Sciences*,49(3),360-375.
- Alawneh,Y., Al-Shara'h,N. (2022) Evaluation of the e-learning experience in Palestinian universities during the Corona pandemic "in light of some quality standards of the Jordanian Higher Education, *Journal of the College of Education (Assiut)*,38(2.2) 181-204
- Alawneh,Y. Al-Momani,T., Salman,F,. Al-Ahmad,S,. Kaddumi,T., Al-Dlalah,M. (2023). The Extent of the Prevalence of Pronunciation Problems among Students of the First Primary Stage in the Point of View of their Teachers and Treatment Methods, *Educational Administration: Theory and Practice*,29(3),19-33.
- Al Khawaldeh,S., Alawneh,Y., Alzboun,M.(2022)., the availability of quality standards for the construction of science achievement tests from the point of view of the examination committees, *Journal of Hunan University (Natural Sciences)*,49(9),1233-1247.
- Alawneh,Y., Ashamali,M., Abdel-Hassan,R., Al-khawaldeh,S., Engestroom,y.(2022) Degree Of Use Of E-Learning Science Teachers In Public High Schools In During The Corona-Covid 19 Pandemic, *Journal of Positive School Psychology*,6(2), 1060-1070.
- Alawneh,Y., Al-Shara'h,N. (2022) Evaluation of the e-learning experience in Palestinian universities during the Corona pandemic "in light of some quality standards of the Jordanian Higher Education, *Journal of the College of Education (Assiut)*,38(2.2) 181-204

- Alawneh,Y., Abualrub,D., Jbara,L.,(2021)Behavioral Phenomena Common Among Kindergarten Students In Nablus Governorate From The Point Of View Of Principals And Teachers, *Turkish Journal of Physiotherapy and Rehabilitation*, 32, (3)231-247.
- Alawneh,Y., Al-Shara'h,N. (2022) Evaluation of the e-learning experience in Palestinian universities during the Corona pandemic "in light of some quality standards of the Jordanian Higher Education, *Journal of the College of Education (Assiut)*,38(2.2) 181-204