



Examining Students' Perceptions of Assessment Quality in the College of Education, Derj

Saleh Alesawi Alnshuoi
Al- Zintan University
saleh.alesawi@uoz.edu.ly

تاريخ الاستلام: 2025/12/8 - تاريخ المراجعة: 2025/12/12 - تاريخ القبول: 2025/12/19 - تاريخ النشر: 2026 /1/17

Abstract

This study was conducted with students from the College of Education in Derj to determine their perceptions regarding tests and the criteria for an effective test. The study focused on students' perspectives due to their diverse experiences and ideas about examinations.

A questionnaire was administered to undergraduate students, and SPSS was used to analyze the responses received. The results indicated that most students prefer questions that are simple and easy to comprehend. Students found it easier to take tests when the questions were clear and concise. Students also emphasized the need for fairness and believed that tests should be based on the material that was actually taught in class.

Some students also stated that quizzes and homework were more important than just taking final examinations. Additionally, some participants reported that exam-related stress negatively affected their performance. In general, students preferred tests that helped them learn better rather than tests that focused only on achieving grades.

المخلص

دراسة تصورات الطلاب حول جودة التقويم في كلية التربية بدرج

أُجريت هذه الدراسة على طلاب كلية التربية بدرج للتعرف على تصوراتهم حول الاختبارات والمعايير التي يرونها مهمة في الاختبار الجيد. وتركز الدراسة على وجهة نظر الطلاب نظراً لاختلاف تجاربهم وأفكارهم المتعلقة بالامتحانات.

تم توزيع استبانة على طلاب مرحلة البكالوريوس، ثم جرى تحليل الإجابات باستخدام برنامج SPSS. وأظهرت النتائج أن معظم الطلاب يفضلون الأسئلة البسيطة وسهلة الفهم. فعندما تكون الأسئلة واضحة ومباشرة، يجد الطلاب أن أداء الاختبارات يكون أسهل. كما أكد الطلاب على أهمية العدالة، ورأوا أن الاختبارات يجب أن تكون مبنية على ما تم تدريسه فعلياً داخل القاعة الدراسية.

وأشار بعض الطلاب أيضاً إلى أن الاختبارات القصيرة والواجبات تساعد على الدراسة أكثر من الاعتماد على الامتحانات النهائية فقط. كما ذكر بعضهم أن التوتر المصاحب للاختبارات يؤثر سلباً على أدائهم. وبشكل عام، يفضل الطلاب الاختبارات التي تساعد على التعلم بشكل أفضل، وليس تلك التي تركز فقط على منح الدرجات.

Chapter One: Introduction

1.1 Background of the Study

Assessment is one of the main ways through which colleges determine whether students have achieved the intended learning outcomes. (Black & Wiliam, 1998; Nitko & Brookhart, 2014). In a College of Education, tests are not used only to assign marks; they also influence how students study, how motivated they feel, and how confident they become as learners. When students believe that assessment is fair and easy to understand, they are more likely to engage seriously with their learning tasks and prepare in a more organised way. (Harlen & Deakin Crick, 2003). However, when they feel the opposite, tests may increase anxiety and reduce students' trust in the learning process.

In the assessment literature, a “good test” is often described through key quality characteristics, such as validity (measuring what the test is intended to measure), reliability (producing consistent results), fairness (being free from bias), clarity of language, appropriate difficulty, and adequate coverage of the course content. (Messick, 1995; Popham, 2017). Although these principles are widely discussed, there can still be a gap between these standards and how tests are actually designed and used in real classroom settings.

Since students are the direct participants in assessment, their views provide an important perspective on whether testing practices are working as intended. Exploring students’ perspectives at the College of Education in Derj may help identify strengths and areas that require improvement especially in relation to question clarity, feedback practices, grading transparency, and the balance between examinations and continuous assessment.

1.2 Statement of the Problem

Assessment is crucial in teaching and learning. However, there is not enough clear information about how students at the College of Education in Derj perceive their exams. From the students’ point of view, it is sometimes unclear whether exams accurately reflect what they studied in class. It is also not always clear whether exams are fair for all students or if they yield the same results. This can negatively impact students’ learning and motivation (Brown & Hirschfeld, 2008).

Many students encounter problems during exams. Some exam questions are difficult to comprehend. Sometimes students do not receive feedback after exams. Additionally, many students feel extremely stressed during tests. This stress can lower students’ performance and confidence (Harlen & Deakin Crick, 2003). If students’ opinions are not considered, it will be challenging for the College to improve its assessment system.

1.3 Purpose of the Study

The primary purpose of this study is to understand how students at the College of Education in Derj think about exams and assessment. Students are the ones who take the exams, so their opinions are essential to know if assessment is functioning effectively (Struyven et al., 2005). This study aims to:

- Determine how students describe a good exam,
- Identify the main problems students face with assessment,
- Suggest simple ways to improve assessment in the College.

1.4 Research Questions

This study attempts to answer the following questions:

- How do students at the College of Education in Derj view the characteristics of a good test?
- What problems do students face with the current testing and assessment system?
- What improvements can be made to make testing and assessment more fair and effective in the College?

1.5 Significance of the Study

This study is important because it gives students a chance to share their opinions about exams and assessment. It helps the College of Education in Derj understand what students see as fair or unfair in testing. The results may help instructors improve exam questions, be clearer in grading, provide better feedback, and create an assessment environment that supports learning instead of causing fear and stress.

1.6 Scope and Delimitation

This study focuses only on undergraduate students at the College of Education in Derj. Data are collected using a questionnaire that asks students about their exam experiences and views

on test quality and assessment. The findings are limited to this College and cannot be generalized to other universities or colleges in Libya.

1.7 Definition of Key Terms

Assessment Practices: The strategies used to measure learning outcomes, including exams, quizzes, assignments, and projects.

Validity: The degree to which an assessment measures what it is intended to measure.

Reliability: The consistency of test results over time or across different groups of students.

Continuous Assessment: Ongoing evaluation methods such as quizzes, assignments, and class participation.

Alternative Assessment: Non-traditional evaluation methods such as projects, portfolios, and presentations that capture broader learning outcomes.

1.8 Methodology

This study follows a quantitative, descriptive approach using a questionnaire distributed to undergraduate students in the College of Education, Derj. The questionnaire includes items related to (a) perceptions of good test characteristics, (b) students' experiences with tests, and (c) opinions about assessment practices. After data collection, responses are coded and analyzed using SPSS to produce descriptive statistics (such as means and standard deviations). The instrument is checked for quality through validity procedures (e.g., an extreme-groups comparison) and reliability indicators (e.g., Cronbach's alpha and split-half reliability). Participation is voluntary and responses are treated confidentially.

Chapter two:

Literature Review

This section looks at four previous studies that talked about assessment from the students' point of view. These studies help explain how students see tests, what they think is fair, and what they believe a good test should be like. Since students are the ones who take the exams, their opinions are important and should not be ignored.

Study 1: Struyven, Dochy, and Janssens (2005)

Struyven, Dochy, and Janssens (2005) reviewed many studies about assessment in higher education. They focused on how students feel about exams and evaluation. The authors found that students change the way they study depending on how they see assessment. If students think the test is fair and clear, they try to understand the subject. But if they think the exam is unfair or unclear, they usually just memorize information to pass. This study shows that assessment affects learning, not only grades.

Study 2: Brown and Hirschfeld (2008)

Brown and Hirschfeld (2008) studied how students understand the purpose of assessment. The study showed that students do not all think about assessment in the same way. Some students feel that tests help them learn and improve. Others feel that tests only put pressure on them or judge them. These feelings influence how students prepare for exams. The study explains that good assessment is not only about writing test questions, but also about how assessment is used in teaching.

Study 3: Gerritsen-van Leeuwenkamp et al. (2019)

Gerritsen-van Leeuwenkamp and colleagues (2019) examined how students judge the quality of assessment. They talked about fairness, clarity, and trust in assessment results. The study found that when students trust the exam system and believe it is fair, they study in a deeper way. When they do not trust it, their learning becomes weaker. This study clearly shows that students' opinions about assessment quality matter.

Study 4: Baniyadi et al. (2022)

Baniasadi and colleagues (2022) asked students directly about what fair assessment means to them. Students said that questions should be clear and grading should be consistent. They also said that all students should be treated equally. Feedback was seen as very important because it helps students understand their mistakes. On the other hand, unclear questions and different grading styles were seen as unfair. This study shows that fairness is something students feel in daily classroom assessment.

General Discussion

From these studies, it can be seen that students care about assessment more than just passing exams. They want tests to be clear, fair, and related to what they study in class. Their opinions affect how they learn and how motivated they feel. For this reason, studying students' views in the College of Education in Derj is necessary to improve assessment practices and make them more fair and effective.

Chapter Three: Methodology

3.1- Introduction:-

This part talks about how the study was done. It discusses the people who took part, the questionnaire that was used, and how the data were analyzed. A questionnaire was sent to students at the College of Education in Derj as part of a quantitative study. With SPSS, the questionnaire was checked to ensure it was valid and reliable before it was used. The Extreme Groups Comparison was used to assess for validity, and Cornbrash's Alpha and the Split-Half method were used to measure reliability. The tools worked well for the study, as shown by the data. Once all the answers were gathered, they were put into SPSS so that averages and other statistics could be found that would help answer the study questions.

3.2- First: Validity of the Study Instrument (Questionnaire)

The validity of the questionnaire refers to the extent to which its items measure what they are intended to measure (Messick, 1995). To verify the validity of the questionnaire, two statistical methods were used through the Statistical Package for the Social Sciences (SPSS).

Discriminant Validity (Extreme Groups Comparison), which involves calculating the t-test value to determine the significance of the differences between the mean scores of the lower quartile (the lowest 27% of the scores) and the upper quartile (the highest 27% of the scores) for all study scales. The results for each scale of the study instrument were as follows.

3.3- Questionnaire for Students – College of Education, Derj

Table (1) Cronbach's Alpha Test Results

N	Dimensions	Number of Items	Cronbach's Alpha (Reliability)	Square Root of Cronbach's Alpha (Validity)
1	Perceptions of Good Test Characteristics	7	0.623	78.93%
2	Students' Experiences with Tests	6	0.516	71.83%
3	Students' Opinions on Assessment Practices	5	0.572	75.63%
	All	18	0.638	79.87%

It is evident from Table (1) that the study instrument achieved a level of reliability according to Cronbach's Alpha, as the values ranged between 0.516 and 0.623, which indicates a varying degree of internal consistency among the items of each dimension of the study, the dimension "Perceptions of Good Test Characteristics" recorded a Cronbach's Alpha value of 0.623, while the dimension "Students' Opinions on Assessment Practices" recorded a value of 0.516. In contrast, the reliability coefficient for the dimension "Students' Experiences with Tests" represented the lowest value among all dimensions at 0.572, the overall Cronbach's Alpha value for all dimensions combined reached 0.638, reflecting an acceptable level of overall reliability for the study instrument.

As for the face validity index, which was calculated through the square root of the reliability coefficient multiplied by 100, the validity of the dimension "Perceptions of Good Test Characteristics" reached 78.93%, while the dimension "Students' Experiences with Tests" recorded 71.83%, the dimension "Students' Opinions on Assessment Practices" achieved a validity of 75.63%, whereas the combined validity for all dimensions reached 79.87%, indicating varying levels of validity across the dimensions of the instrument.

Thus, the validity and reliability of the study instrument have been confirmed, which provides confidence in its accuracy and suitability for analyzing the results and answering the study's hypotheses or research questions.

Table (3) Results of the Split-Half Reliability Test

Scale	Number of Items	Correlation Coefficient	Spearman Reliability Coefficient
Understanding of Good Test Characteristics	7	0.593	0.729
Practices in Test Design and Adjustment	6	0.423	0.588
Assessment Practices	6	0.541	0.678
All	19		

The results in Table (3) indicate that the Understanding of Good Test Characteristics dimension, obtained a correlation coefficient of 0.593, while the Spearman–Brown reliability coefficient reached 0.729, this value demonstrates a good level of internal consistency and adequate reliability for this dimension.

The results in Table (3) show that the Practices in Test Design and Adjustment dimension, recorded a correlation coefficient of 0.423, whereas the Spearman–Brown reliability coefficient was 0.588, this relatively moderate value indicates limited reliability for this dimension compared to the others.

The results in Table (3) reveal that the Assessment Practices dimension, achieved a correlation coefficient of 0.541, while the Spearman–Brown reliability coefficient reached 0.678, these values reflect a good level of internal consistency and reliability for this dimension.

3.4- Section A: Background Information

1–Gender: Table (4) shows the distribution of respondents according to their gender.

Table (4) shows the frequency distribution of respondents according to gender.

Gender	Number	Percentage%
Male	0	0.0%
Female	30	100.0%
Total	30	100%

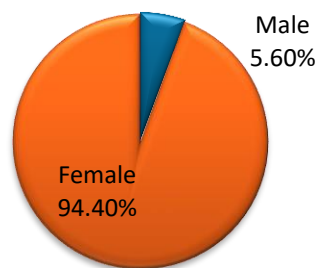


Figure (1) illustrates the percentages of the study population by respondents' gender.

Based on Table (4) and Figure (1), it is evident that the highest percentage of the study sample by gender was within the female category, which reached 100.0%, while the male category recorded 0.0%.

2– Age: Table (5) and Figure (2) show the distribution of the study sample according to age.

Table (5) shows the frequency

Age	Number	Percentage%
Under 20 years	12	40.0%
20–22 years	14	46.7%
23–25 years	4	13.3%
Above 25 years	0	0.0%
Total	30	100%

distribution according to age.

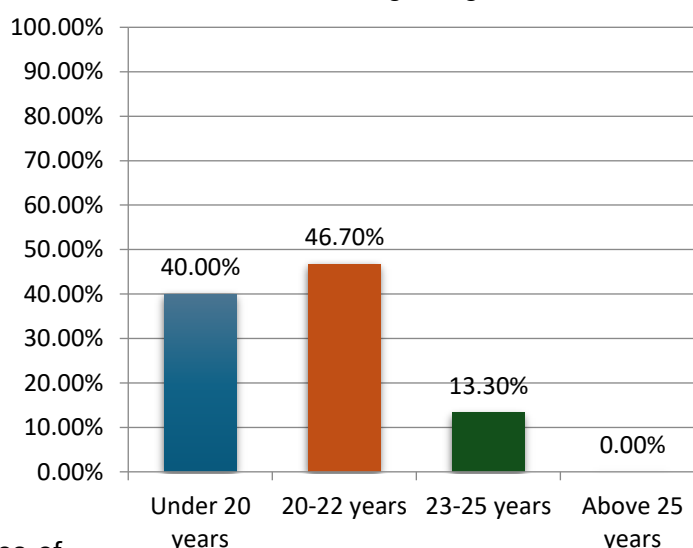
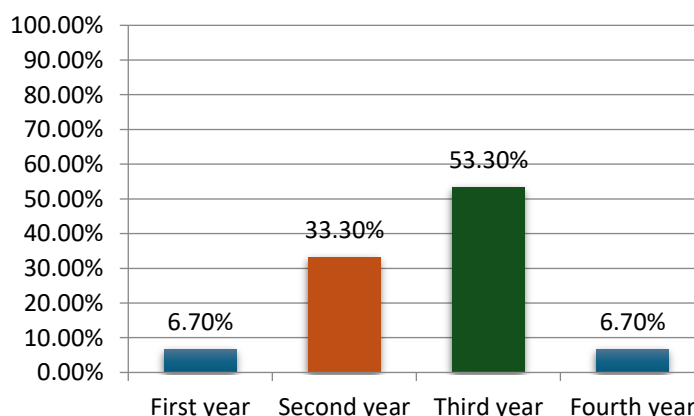


Figure No. (2) illustrates the percentages of the study population according to age.

It is evident from Table (5) and Figure (2) that the highest percentage of the study sample by age category was within the group 20–22 years, which accounted for 46.7%, followed by the group under 20 years with 40.0%, then the group 23–25 years with 13.3%, while the category above 25 years recorded 0.0%.

3–Year of study: Table (6) and Figure (3) show the distribution of the study sample according to Year of study.

Table (6) shows the frequency distribution according to Year of study.



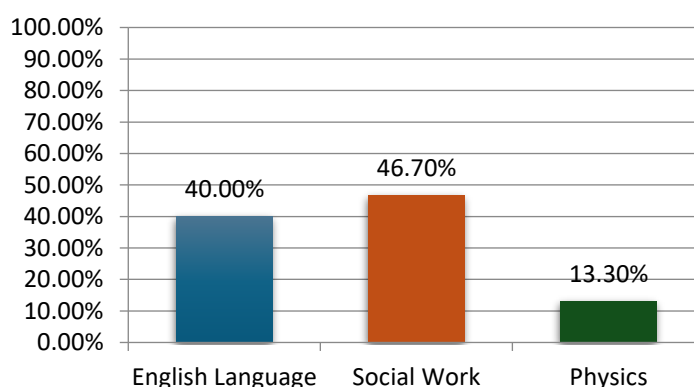
Year of study	Number	Percentage%
First year	2	6.7%
Second year	10	33.3%
Third year	16	53.3%
Fourth year	2	6.7%
Total	30	100%

Figure No. (3) illustrates the percentages of the study population according to Year of study.

It is evident from Table (6) and Figure (3) that the highest percentage of the study sample according to the year of study was within the Third Year category, which reached 53.3%. This was followed by the Second Year category with 33.3%, then the categories First Year and Fourth Year with 6.7% each.

4-Department: Table (7) and Figure (4) show the distribution of the study sample according to Department.

Table (7) shows the frequency distribution according to Department.



Year of study	Number	Percentage%
English Language	12	40.0%
Social Work	14	46.7%
Physics	4	13.3%
Total	30	100%

Figure No. (4) illustrates the percentages of the study population according to Department.

It is evident from Table (7) and Figure (4) that the highest percentage of the study sample according to the field of study was within the Social Work category, which reached 46.7%. This was followed by the English Language category with 40.0%, then the Physics category with 13.3%.

3.5– Section B : Responses to the Study Questions

To achieve the objectives of the study and analyze the data that were collected, several appropriate statistical methods were used through the Statistical Package for Social Sciences (SPSS). After collecting the distributed questionnaire forms, a numerical coding method was applied to code the responses related to the five–point and three–point Likert scales as shown in Table (8).

Table (8) Distribution of scores for the responses related to the scale items

Response	Not at all important	Slightly important	Moderately important	Important	Very important
Score	1	2	3	4	5

The arithmetic mean (weighted mean) is then calculated to determine the weights of the statements according to the weighted mean values obtained as a result of analyzing the responses, as shown in Table (8). This is done after coding and entering the data into the computer. To determine the length of the cells of the five–point scale (lower and upper limits) used in the study dimensions, the range was divided by the number of scale cells to obtain the range length ($5-1 = 4$), and then the range was divided into five levels ($4 \div 5 = 0.80$), and this number represents the length of a single category or level. The weights are shown in the following weighted mean table.

Table (9) The Weighted Mean Obtained from the Analysis of Responses

Level	Very Low	Low	Moderate	High	Very High
Weighted Mean	From 1 to less than 1.80	From 1.80 to less than 2.60	From 2.60 to less than 3.40	From 3.40 to less than 4.20	From 4.20 to 5

3.6–Section 1: Perceptions of Good Test Characteristics

To answer the research question, the arithmetic mean, standard deviation, relative weight, and ranking of the variables of the level of psychological well-being in descending order were used, and the results were presented in the following table:

Table of Arithmetic Means, Relative Weight, and Ranking of Dimensions

No.	Dimensions	Arithmetic Mean	Standard Deviation	Relative Weight	Rank	Degree of Agreement
1	Validity (measures what it should)	4.50	0.6823	90.00%	3	Very High
2	Reliability (consistent results)	3.70	1.2077	74.00%	7	High
3	Fairness (free from bias)	4.63	0.7184	92.67%	2	Very High
4	Clarity of language	4.70	0.9154	94.00%	1	Very High
5	Appropriate difficulty (not too easy, not too hard)	3.97	1.0662	79.33%	6	High
6	Coverage of syllabus topics	4.23	0.9353	84.67%	5	Very High
7	Usefulness for feedback	4.37	1.0981	87.33%	4	Very High
	Overall Mean	4.30	0.6293	86.00%		Very High

It is evident from Table (10) that the degree of agreement on the statements as a whole was Very High, as the overall arithmetic mean reached 4.30, with a standard deviation of 0.6293, and a relative weight of 86.00%. The statements were ranked in descending order in terms of the degree of agreement from the perspective of the study sample as follows:

1. Statement No. (4): "Clarity of language" ranked first in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, as the arithmetic mean reached 4.70, the standard deviation 0.9154, and the relative weight 94.00%.
2. Statement No. (3): "Fairness (free from bias)" ranked second with a Very High degree of agreement, with an arithmetic mean of 4.63, a standard deviation of 0.7184, and a relative weight of 92.67%.
3. Statement No. (1): "Validity (measures what it should)" ranked third in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, where the arithmetic mean reached 4.50, the standard deviation 0.6823, and the relative weight 90.00%.
4. Statement No. (7): "Usefulness for feedback" ranked fourth with a Very High degree of agreement, in terms of the degree of agreement from the perspective of the study sample, with an arithmetic mean of 4.37, a standard deviation of 1.0981, and a relative weight of 87.33%.
5. Statement No. (6): "Coverage of syllabus topics" ranked fifth in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, where the arithmetic mean reached 4.23, the standard deviation 0.9353, and the relative weight 84.67%.
6. Statement No. (5): "Appropriate difficulty (not too easy, not too hard)" ranked sixth with a High degree of agreement, in terms of the degree of agreement from the perspective of the study sample, with an arithmetic mean of 3.97, a standard deviation of 1.0662, and a relative weight of 79.33%.
7. Statement No. (2): "Reliability (consistent results)" ranked seventh in terms of the degree of agreement from the perspective of the study sample, with a High degree of agreement, where the arithmetic mean reached 3.70, the standard deviation 1.2077, and the relative weight 74.00%.

6.7-Section 2: Students' Experiences with Tests

To answer the research question, the arithmetic mean, standard deviation, relative weight, and ranking of the variables of the level of psychological well-being in descending order were used, and the results were presented in the following table:

Table of Arithmetic Means, Relative Weight, and Ranking of Dimensions

No.	Dimensions	Arithmetic Mean	Standard Deviation	Relative Weight	Rank	Degree of Agreement
1	Test questions are clear and easy to understand	3.63	0.9643	72.67	2	High
2	Tests reflect what was taught in class	2.77	1.3047	55.33	6	Moderate
3	Test difficulty matches the level of study	3.07	1.2299	61.33	5	Moderate
4	Different types of questions are used (MCQ, essay, short answer)	3.37	1.2452	67.33	3	High
5	Teachers explain how grades are given	3.67	1.0613	73.33	1	Very High
6	Students receive feedback on strengths and weaknesses	3.30	1.1788	66.00	4	High
	Overall Mean	3.30	0.5314	66.00		High

It is evident from Table (11) that the degree of agreement on the statements as a whole was High, as the overall arithmetic mean reached 3.30, with a standard deviation of 0.5314, and a relative weight of 66.00%, The statements were ranked in descending order in terms of the degree of agreement from the perspective of the study sample as follows:

1. Statement No. (5): "Teachers explain how grades are given" ranked first in terms of the degree of agreement from the perspective of the study sample with a Very High degree of agreement, as the arithmetic mean reached 3.67, the standard deviation 1.0613, and the relative weight 73.33%.

2. Statement No. (1): "Test questions are clear and easy to understand" ranked second with a High degree of agreement, in terms of the degree of agreement from the perspective of the study sample with an arithmetic mean of 3.63, a standard deviation of 0.9643, and a relative weight of 72.67%.
3. Statement No. (4): "Different types of questions are used (MCQ, essay, short answer)" ranked third in terms of the degree of agreement from the perspective of the study sample with a High degree of agreement, as the arithmetic mean reached 3.37, the standard deviation 1.2452, and the relative weight 67.33%.
4. Statement No. (6): "Students receive feedback on strengths and weaknesses" ranked fourth in terms of the degree of agreement from the perspective of the study sample with a High degree of agreement, with an arithmetic mean of 3.30, a standard deviation of 1.1788, and a relative weight of 66.00%.
5. Statement No. (3): "Test difficulty matches the level of study" ranked fifth in terms of the degree of agreement from the perspective of the study sample with a Moderate degree of agreement, as the arithmetic mean reached 3.07, the standard deviation 1.2299, and the relative weight 61.33%.
6. Statement No. (2): "Tests reflect what was taught in class" ranked sixth with a Moderate degree of agreement, in terms of the degree of agreement from the perspective of the study sample with an arithmetic mean of 2.77, a standard deviation of 1.3047, and the relative weight 55.33%.

3.8–Section 3: Students' Opinions on Assessment Practices

To answer the research question, the arithmetic mean, standard deviation, relative weight, and ranking of the variables of the level of psychological well-being in descending order were used, and the results were presented in the following table:

Table of Arithmetic Means, Relative Weight, and Ranking of Dimensions

No.	Dimensions	Mean	Standard Deviation	Relative Weight	Rank	Degree of Agreement
1	Exams are the best way to measure student achievement	3.03	1.3767	60.67	6	Very High
2	Continuous assessment (quizzes, assignments) helps me learn better	4.03	1.1290	80.67	2	Very High
3	Teachers provide fair grades	3.63	1.2452	72.67	5	Very High
4	Alternative assessments (projects, presentations) are useful	3.77	1.3309	75.33	3	Very High
5	I feel anxious during exams, which affects my performance	4.10	1.0619	82.00	1	High
6	Exams are the best way to measure student achievement	3.71	0.7714	74.27	4	High
	Overall Mean	3.03	1.3767	60.67		Very High

It is evident from Table (12) that the degree of agreement on the statements as a whole was Very High, as the overall arithmetic mean reached 3.03, with a standard deviation of 1.3767, and a relative weight of 60.67%. The statements were ranked in descending order in terms of the degree of agreement from the perspective of the study sample as follows:

Statement No. (5): “I feel anxious during exams, which affects my performance” came in the first place in terms of the degree of agreement from the perspective of the study sample, with a High degree of agreement, as the arithmetic mean reached 4.10, the standard deviation 1.0619, and the relative weight 82.00%.

Statement No. (2): "Continuous assessment (quizzes, assignments) helps me learn better" came in the second place in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, with an arithmetic mean of 4.03, a standard deviation of 1.1290, and a relative weight of 80.67%.

Statement No. (4): "Alternative assessments (projects, presentations) are useful" came in the third place in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, with an arithmetic mean of 3.77, a standard deviation of 1.3309, and a relative weight of 75.33%.

Statement No. (6): "Exams are the best way to measure student achievement" came in the fourth place in terms of the degree of agreement from the perspective of the study sample, with a High degree of agreement, where the arithmetic mean reached 3.71, the standard deviation 0.7714, and the relative weight 74.27%.

Statement No. (3): "Teachers provide fair grades" came in the fifth place in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, where the arithmetic mean reached 3.63, the standard deviation 1.2452, and the relative weight 72.67%.

Statement No. (1): "Exams are the best way to measure student achievement" came in the sixth place in terms of the degree of agreement from the perspective of the study sample, with a Very High degree of agreement, as the arithmetic mean reached 3.03, the standard deviation 1.3767, and the relative weight 60.67%.

Chapter Four: Findings and Recommendations

4.1 Findings:–

This study looked at students in the College of Education in Derj and asked them about their views on exams and assessment. The questionnaire answers were read carefully. From these answers, several clear ideas appeared.

Many students talked about the language of exam questions. They said that questions should be clear and easy to read. When the wording is simple, students feel more relaxed during the exam. They understand what is required from them. On the other hand, unclear questions cause problems. Some students said they became confused even though they studied the subject well. This issue appeared many times in the responses.

Fairness was also mentioned by many students. They want exams to be fair for everyone. Students believe that questions should come from what was actually explained in lectures. Some students felt that a few exams included questions that were not discussed clearly in class. Because of this, they felt the exam was not fair.

Students did not all have the same experience with tests. Some students said that it helps when teachers explain how marks are given. When grading is clear, students feel more comfortable. Other students said that some exams do not show their real level. They felt that their answers did not always reflect what they had learned. A number of students also said that some exams were harder than expected.

Students also shared their opinions about different assessment methods. Many students preferred quizzes and homework. They said these help them study little by little. Projects and presentations were also seen as useful. These methods were described as less stressful than final exams. In contrast, many students said that they feel nervous during exams. Stress was a common problem. Some students explained that stress makes them forget answers during the test.

In general, students want exams to be clear, fair, and related to the course content. They also want assessment to help them learn, not only to give marks. These results are similar to what has been found in earlier studies about good assessment and formative practices (Black & Wiliam, 1998; Brookhart, 2013; Torrance, 2007).

5.2 Recommendations

Some straightforward recommendations can be derived from the students' responses.

First, exam questions should be clear and easy to understand. Teachers should avoid using complex vocabulary.

Second, exams should be based on the material covered hat in class. This helps students feel the exam is fair.

Third, teachers should incorporate more quizzes and homework into their teaching. This helps students feel that the opportunities for exam is fair. It also reduces exam stress.

Fourth, there should be more projects and presentations. Exams should not be limited to written tests. Clear marking helps students feel more at ease.

Finally, students need feedback after exams. Even brief feedback can help students identify their mistakes and improve their performance in the future.

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
<https://doi.org/10.1080/0969595980050102>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3–17.
<https://doi.org/10.1080/09695940701876003>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Van der Vleuten, C. P. M. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Studies in Educational Evaluation*, 63, 72–82.
<https://doi.org/10.1016/j.stueduc.2019.07.005>
- Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169–207.
<https://doi.org/10.1080/0969594032000121270>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson Education.
- Popham, W. J. (2017). *Classroom assessment: What teachers need to know* (8th ed.). Pearson Education.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325–341. <https://doi.org/10.1080/02602930500099102>
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education undermines learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281–294.
<https://doi.org/10.1080/09695940701591801>