Journal of Education and e-Learning Research
Vol. 12, No. 1, 115-123, 2025
ISSN(E) 2410-9991 / ISSN(P) 2518-0169
DOI: 10.20448/jeelr.v12i1.6676

Evaluating the effectiveness of a computerized achievement test using learn smart for psychometric assessment under item response theory

Mimi Ismail[1]
Ahmed Al – Badri[2]
Said Al – Senaidi[3]

(Corresponding Author)

[1,2,3]Educational Studies Department, University of Technology and Applied Science, Rustaq College of Education, Oman.
[1]Email: Mimi.ismail@utas.edu.om
[2]Email: ahmed.albadri@utas.edu.om
[3]Email: said.alsenaidi@utas.edu.om
[1]Educational Psychology, Zagazig University, Zagazig, Egypt.

Abstract

This study aimed to reveal the differences in individuals' abilities, their standard errors, and the psychometric properties of the test according to the two methods of applying the test (electronic and paper). The descriptive approach was used to achieve the study's objectives. The study sample consisted of 74 male and female students at the University of Technology and Applied Sciences in Rustaq. An electronic test was built on the Learn Smart platform supported by artificial intelligence in psychological measurement. The results showed no statistically significant differences in the individuals' average ability estimates and their standard errors between the electronic and paper-based tests. Besides, the vocabulary difficulty estimates in the electronic test ranged between -2.562 and 2.007 and the vocabulary difficulty estimates in the paper-based test ranged between -3.483 and 2.194. All of them are within the acceptable range. The chi-square values for the electronic test items are not statistically significant except for items (8 and 10). On the other hand, all chi-square values for the items on the paper test are not statistically significant except the item numbers (3, 6, 8, 10 and 12) which are statistically significant at the 0.05 and 0.01 levels.

Keywords: Effectiveness, Electronic test, Measurement accuracy, Psychometric assessment, Item response theory, Learn smart.

Citation | Ismail, M., Al – Badri, A., & Al – Senaidi, S. (2025). Evaluating the effectiveness of a computerized achievement test using learn smart for psychometric assessment under item response theory. *Journal of Education and E-Learning Research, 12*(1), 115–123. 10.20448/jeelr.v12i1.6676
History:
Received: 18 February 2025
Revised: 25 March 2025
Accepted: 11 April 2025
Published: 9 May 2025

Publisher: Asian Online Journal Publishing Group

Funding: This study received no specific financial support.
Institutional Review Board Statement: The Ethical Committee of the Rustaq College of Education, University of Technology and Applied Science, Oman has granted approval for this study on 15 January 2025 (Ref. No. EDU012025).
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

Contents

Contribution of this paper to the literature

This study is distinctive in the Copilot artificial intelligence program to create an electronic computerized achievement test and apply it to students to reveal its effectiveness in the educational and evaluation process. To the best of the researchers' knowledge, this has not been applied before.

1. Introduction

The quality of university education depends on its ability to graduate individuals with professional knowledge and skills that are consistent with the requirements of the labor market and achieve the objectives of the educational process in various fields of knowledge. Hence, evaluating student performance in university education is the most important pillar of the educational process. Its importance lies in its ability to identify and select those who have achieved the desired university education goals to meet the requirements of the labor market and achieve Oman Vision 2040. The methods of evaluating university students vary according to their specialization and the nature of the courses they study, from writing reports, study projects, oral and written tests, and semester work which ultimately result in grades or a cumulative average and serve as the final evaluation of the university student.

However, the written test or the written achievement test at the end of the semester will always remain the most important element and the basic tool for evaluating the student. Its results are relied upon greatly according to the requirements of the different courses. The course of psychological measurement and educational evaluation always seeks to improve the evaluation process and develop its tools to provide accurate information about the performance of university students (Parkin, Frisby, & Wang, 2020).

Bennett (2002) indicated that the use of electronic tests as one of the electronic assessment methods in schools and universities was absent until the beginning of this century. Then, the United States of America began to use electronic tests steadily. Other countries such as Singapore and Norway have begun to study and explore ways to computerize measurement and evaluation tools for public school and university students.

2. Background

2.1. Item Response Theory

The item response theory has become a standard method for constructing and developing tests. The availability of computer programs has encouraged the application of various models of this theory in the field of tests and measurements. These models link the characteristics of items to one or more parameters (Liu et al., 2022). They provide an alternative to classical testing theory by estimating the parameters of the individual and the item with the least amount of error without resorting to a random sample of test items from the measured range or obtaining a very large sample of items representing this range (Wilson, 2023).

Item response theory attempts to describe the relationship between a test-taker's performance on a test and the trait underlying that performance. This theory postulates that (a) a test-taker's performance on a test can be explained by a set of factors called latent traits, or abilities. (b) The relationship between a test-taker's performance on a test and the set of traits that are assumed to influence test performance can be described by an increasing function called an item characteristic function (Umobong, 2017).

2.2. One Parameter Logistic Model (IPL)

It is a probabilistic mathematical model within item response theory achieving good results in the field of achieving objectivity in behavioral measurement. This model is one of the most well- known in the theory as this model assumes that the measured trait is one-dimensional (Robitzsch, 2023).

G.Rasch, the author of this model, believes that the individual's response to an item (i) depends on the individual's ability (θ) and the difficulty of the item (δi) and that the probability of this response to item Pi (θ) is a function of the difference between the individual's ability and the difficulty of the item (δi -θ), i.e., Pi (θ) = f (θ- δi)

The previous function can be converted into a probabilistic mathematical model that achieves the relationship between θ, δi and (θ). The mathematical relationship between the three previous variables can be represented by the following equation:

$$P_i(\theta) = \frac{e^{(\theta-\delta_i)}}{1+e^{(\theta-\delta_i)}}, i = 1,2,\ldots,n$$

Where

(θ) Pi is the probability of the correct response of the individual (θ) to the item (i).

(e) is the logarithmic equivalent.

(θ) is the individual's ability.

(δi) is the difficulty of the item (i).

This model focuses on one parameter which is the difficulty of the item meaning that this model is concerned with determining the position of the test item on the difficulty scale of all test items (difficulty parameter). This model assumes that all test items are equal in distinguishing between the levels of measured ability and that ability is one-dimensional. It also assumes that the individual does not resort to guessing (Reeve, 2024).

The first and most important step when using item response models in constructing tests and analyzing data is to estimate the individual and item parameters in a particular model that is chosen. This depends on appropriate methods and techniques for estimating these parameters (Kreiner & Christensen, 2012; Madison, Wind, Maas, Yamaguchi, & Haab, 2024).

There are several methods used to estimate the parameters of the item (difficulty, discrimination, and guesswork) and the parameters of the individual (the ability of the individual) when both the parameters of the item and the individual are unknown. Among the most famous methods is the joint maximum likelihood where the parameters of the item and the individual are estimated jointly. There are also methods to estimate the parameters of the item and the ability when the parameters of the item are known and the parameters of the ability are unknown, including the method of marginal maximum likelihood and conditional maximum likelihood where the distribution of ability is integrated (Pan, Fang, Pan, & Fang, 2002).

2.3. Rational Gap

Many studies such as the study of Clariana and Wallace (2002) and the study of Prisacari, Holme, and Danielson (2017) have emphasized the importance of computerized electronic achievement tests compared to traditional paper tests. The results showed that 50% of students excelled on the computer-based test while 25% excelled on the paper-based test, and 25% performed similarly on both the paper and electronic tests.

James (2016) conducted a study to identify students' attitudes towards the use of computerized tests in the College of Education, Department of Psychology, University of New England, Australia. The results showed that students face many challenges when taking computerized tests related to technical matters, such as the Internet, its speed, the test system itself, and the technical problems associated with it. However, students agreed on the role of computerized tests in reducing their anxiety levels in addition to the lower material costs when taking these tests.

Many teachers tend to prefer the use study showed that many paper tests are doubting the equivalence of electronic tests with them despite the growing awareness of the importance of electronic tests and their increasing benefit (Dikli, 2003). According to James (2016) the results of the study showed that there are many challenges facing students during computer-based tests related to technical matters such as the existence of challenges related to the Internet, its speed, the testing system itself, and the technical problems associated with it. However, the students agreed on the role of computer-based tests in reducing their level of anxiety in addition to the low financial costs of conducting these tests.

Studying the efficiency of electronic tests is extremely important it is imposed by the gap between the accelerating trend in using computers in learning and teaching processes on the one hand and the extreme slowness in adopting electronic tests on the other hand. Therefore, this study aims to assess the effectiveness of the computerized electronic achievement test in the course of psychological measurement and educational evaluation in achieving measurement accuracy compared to the traditional paper test using the modern theory of measurement (IRT). Therefore, this study aims to

1) Detecting the extent of the difference in the estimates of individuals' abilities and their standard errors according to the method of applying the test (electronic paper).
2) Detecting the extent of the difference in the psychometric properties of the achievement test according to the method of applying the test (electronic paper).
3) Identifying the best model for judging the efficiency of information in the achievement test according to the two methods of applying the electronic and paper tests.

3. Method

3.1. Study Design

The descriptive approach was used, appropriate for the study's nature and its objectives. This approach goes beyond collecting and organizing data to analyzing and interpreting it to reach a set of results that help in understanding reality and then working to improve it.

An achievement test was constructed electronically and on paper according to the steps of preparing a good achievement test for the first unit in the psychological measurement and educational evaluation course for third-year students at the University of Technology and Applied Sciences in Rustaq. The researcher used a method combining procedural analysis and hierarchical structural analysis to analyze the main competencies measured by the test by following the following stages:

3.1.1. Analysis Stage

- Determining the basic competencies in the first and second units of the psychological measurement and educational evaluation course.
- Determining the behavioral scope of the competencies.
- Formulating behavioral objectives and arranging them in a hierarchical order
- Construction stage (writing test items).
- Experimentation stage.

3.2. Participants

The study community consisted of 150 male and female students from the College of Education at the University of Technology and Applied Sciences in Rustaq studying the course psychological measurement and educational evaluation . The sample for calculating the reliability and validity of the test (exploratory sample) consisted of 35 male and female students from the third year studying the course of Psychological Measurement and Educational Evaluation at the University of Technology and Applied Sciences in Rustaq while the basic study sample consisted of 74 male and female students in the third year at the University of Technology and Applied Sciences in Rustaq, including 26 male students and 48 female students with an average age of 18.8 and a standard deviation of 1.31. Table 1 shows the characteristics of the participants.

Table 1. Characteristics of the participants.

Response rate and characteristic	No. of respondents (%)
Gender (Overall $N = 74$)	
Male	26/74 (35.14%)
Female	48/74 (64.86%)

3.3. Data Analysis

The study used different data analysis tools and statistical software (SPSS) version 27 and the Stata Analysis Program version 14 was used to analyze the results. The following analyses were performed:

- Cronbach's alpha to verify the reliability of the questionnaire.
- Calculating the estimates of individuals' abilities and their standard errors for the computerized electronic test and the traditional paper test
- A paired sample t-test to compare the means of the groups' responses.
- Information function criterion calculation (Akaike's, Bayesian).

3.4. Statistics Reliability

3.4.1. Validity

The study identified the main competencies measured in the first unit of the psychological measurement and educational evaluation course which the student must master to be considered proficient in the two units. They were analyzed into 6 main competencies according to the method of procedural analysis and hierarchical structural analysis. Behavioral objectives were formulated in light of this analysis in an objective, observable, and measurable manner. The number of objectives formulated was 15. Organizing these competencies with the objectives and vocabulary and preparing a list of them to present to experts in educational psychology, measurement, and evaluation to express their opinions.

Each arbitrator was asked to clarify the extent to which each vocabulary is appropriate for measuring the field it measures and to suggest the minimum level of achievement in the test. The arbitrators' agreement coefficient was calculated on the extent to which each vocabulary is appropriate. The agreement rates between the arbitrators on the test vocabulary (15) items ranged between 90% and 100%. The arbitrators agreed on the optimal minimum for the passing level which is 70%. The arbitrators also agreed on the importance of students' mastery of the five main competencies in the unit being measured. All values of the coefficients of agreement of the vocabulary with the arbitration elements ranged between 0.90 and1.00 which indicates that the test has a very high content validity.

3.4.2. Stability

The test stability was calculated according to the two methods of applying the test in electronic and paper form using the SPSS program by calculating the Cronbach's alpha coefficient (if the item score was deleted). The results showed that the test stability coefficient for the computerized electronic test equals 0.742 and the test stability coefficient for the traditional paper test equals 0.714 which are good stability coefficients that confirm the test stability according to the two methods of applying it. The coefficient results are shown in Table 2.

Table 2. Reliability statistics of the test.

Number of cases	Number of questions	Cronbach's alpha for the electronic test	Cronbach's alpha for the paper test
35	15	0.742	0.714

The value of Cronbach's alpha which is 0.741 for the computerized electronic test, and 0.714 for the traditional paper test indicates that the questions in the test are highly reliable and consistent. It means test is a good tool for measuring what it was designed to measure.

3.5. Verifying the Assumptions of Item Response Theory for Achievement Testing

The assumptions of item response theory for achievement testing were verified: unidimensionality, measurement independence and item characteristic curve. The study carried out the following procedures:

3.5.1. Verification of One-Dimensionality

The study is an exploratory factor analysis using the basic components method before rotation for the items of the paper and electronic test. The results for the paper test show that there is one dominant factor, as the latent root of the first and second factors according to the complete method for estimating grades reached (6.335, and 2.087) respectively, and the percentage of explained variance reached (42.233% and 13.910%) respectively, as dividing the latent root of the first factor by the latent root of the second factor equals (0.353) which is greater than the value 2 which is considered a criterion for one-dimensionality (Lian & Idris, 2006) and the percentage of explained variance for the first factor reached 42.233%. This percentage is considered ideal for judging the paper test as one-dimensional, which means that the first factor explains more than 20% of the total variance. The results also show for the electronic test that there is one dominant factor as the latent root of the first and second factors according to the complete method for estimating the grades (6.419 and 1.693). The percentage of explained variance (42.793% and 11.287%) where the division of the latent root of the first factor by the latent root of the second factor equals 7913. This is greater than the value 2 which is considered a criterion for one-dimensionality. The percentage of explained variance of the first factor reached 42.793%. This percentage is considered ideal for judging the electronic test as one-dimensional which is that the first factor explains more than 20% of the total variance.

3.5.2. Measurement Independence

It indicates that the answer to any item of the paper and electronic test items is not affected by the answer to any other item negatively or positively meaning that the assumption of measurement independence is achieved if the probability of the correct answer to an item of the scale items is not related to the probability of the correct answer to any other item as all correlation coefficients were less than 0.7which is the criterion that if the value of the correlation coefficient exceeds it, the assumption of measurement independence is rejected (Linacre, 2018).

3.5.3. Item Characteristic Curve

It is the probability of reaching the correct response to the item as a function of the latent trait or measured ability in light of performance on the items. The graphs of this curve show the probability of reaching the correct response to the test as a function of the latent trait or measured ability in light of performance on the items. To verify this assumption, the test characteristic curve was extracted in its paper and electronic forms for items (1) and (2) using the (STATA) program in light of the Rasch model as shown in Figures 1 and 2.

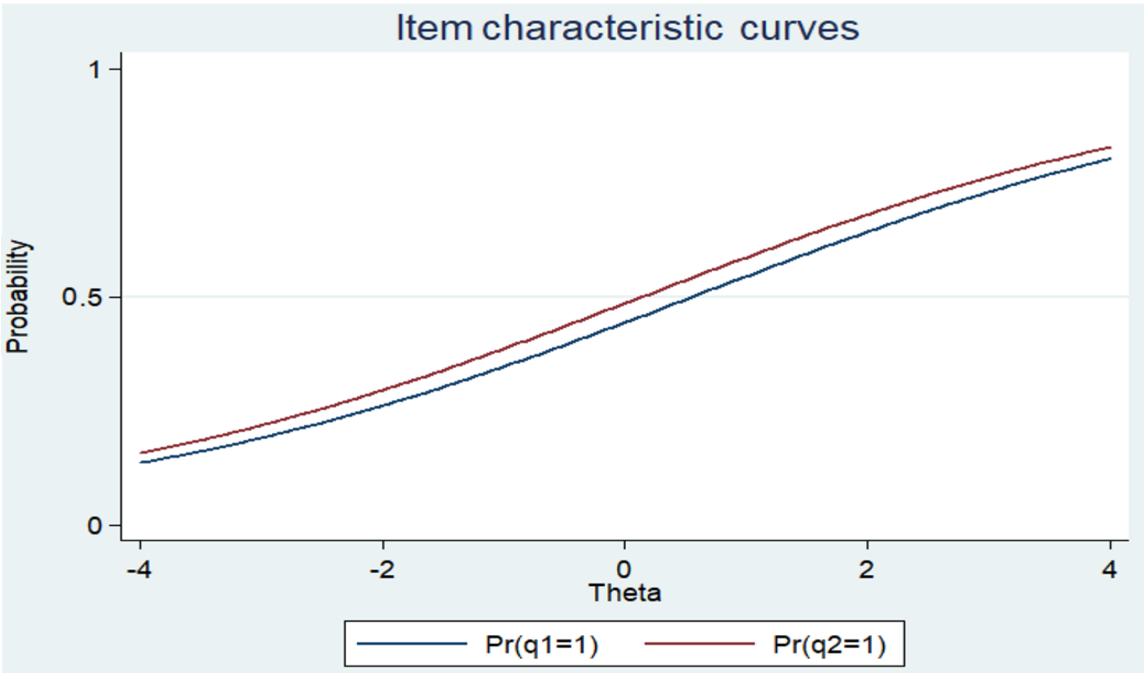

Figure 1. Paper test characteristic curve for items (1) and (2).

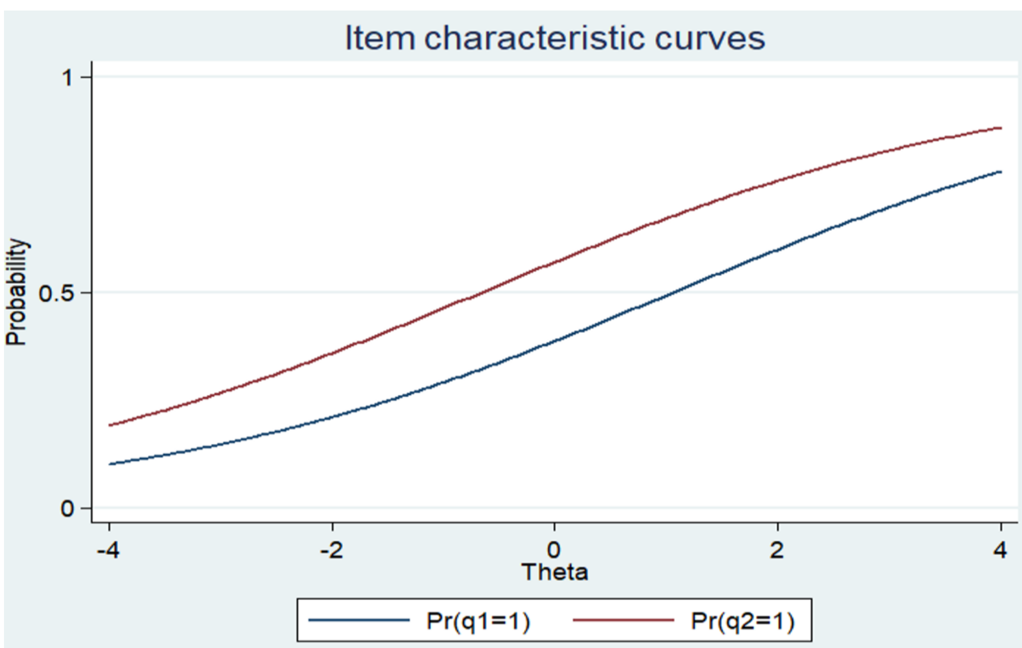

Figure 2. Electronic test characteristic curve for items (1) and (2).

It is clear from Figures 1 and 2 that the greater the individual's ability, the greater his score on the item. It is noted that the difficulty coefficient affects the probability of a correct response. There is no effect of the discrimination coefficient meaning that the Rasch model helps in achieving parallelism of the item characteristics curves.

4. The Results and Discussion

4.1. Answering the First Question: Do the Estimates of Individuals' Abilities and Their Standard Errors Differ Depending on the Method of Applying the Test (Electronic – Paper)?

The estimates of individuals' abilities and their standard errors were calculated for the computerized electronic test and the traditional paper test using the statistical program STATA in light of the Rasch model through the post estimation tool. The SPSS statistical analysis program was used through a paired sample t-test to determine the significance of the differences in the averages of individuals' ability estimates and their standard errors. The obtained results are presented in Table 3.

Table 3. Results of the t-test for the significance of the differences between the means of the estimates of individuals' abilities.

Variables	Test	Mean	Standard deviation	DF	T	Significance level
Individual ability	Electronic	0.874	0.0002	73	0.997	Not significant
	Paper	0.708	0.0001			
Standard error of individual ability	Electronic	0.496	0.016	73	54.25	Significant
	Paper	0.658	0.028			

The results in Table 3 showed that the highest value of the average ability estimate of the study sample individuals was in favor of the computerized electronic achievement test, while the lowest value of the average ability of individuals was in the case of the paper-based achievement test. The results also show no statistically significant differences between the average ability estimate for both the computerized electronic test and the traditional paper test, which means that the computerized electronic test provides the same level of ability estimate as the traditional paper test. The previous results also showed that the average standard error of the computerized electronic ability estimate was significantly lower at a significance level of 0.01 than the average standard error of the ability estimate in the traditional paper test which indicates the accuracy of measuring the computerized electronic test in estimating the standard error of students' abilities compared to the traditional paper test to actively participate in innovative endeavors within their academic field.

4.2. Answering the Second Question: Do the Psychometric Properties of the Achievement Test Differ Depending on the Two Methods of Applying the Test (Electronic and Paper) Using the Rasch Model?

The electronic test items were analyzed using the STATA statistical program to estimate the difficulty parameters and their standard errors, the chi-square values, and their statistical significance for each item according to the two test application methods (electronic paper) using the Rasch model. The reliability was calculated according to the two methods of applying the test (electronic paper) using the STATA program by calculating the Cronbach's alpha coefficient using the multivariate analysis tool. The results obtained are displayed in Table 4.

Table 4. Psychometric properties of items for the achievement test according to the method of applying the test using the Rasch model.

Questions	Electronic test				Paper test			
	Difficulty	Standard error of difficulty	Chi square	Sig	Difficulty	Standard error of difficulty	Chi square	Sig
Q1	1.064	0.617	1.38	0.115	0.555	0.610	0.91	0.363
Q2	-0.656	0.584	1.72	0.085	0.139	0.595	0.23	0.815
Q3	-2.562	0.839	-1.12	0.262	-3.483	1.078	-3.23	0.001**
Q4	-1.376	0.717	-1.05	0.212	-1.267	0.675	-1.88	0.061
Q5	-1.198	0.632	-2.50	0.058	-0.835	0.630	-1.32	0.185
Q6	-1.771	0.707	-1.79	0.072	-1.415	0.694	-2.04	0.042*
Q7	0.928	0.604	1.54	0.125	0.555	0.610	0.91	0.363
Q8	-2.232	0.781	-2.86	0.014*	-2.031	0.787	-2.58	0.010*
Q9	-0.924	0.605	-1.53	0.127	-1.267	0.675	-1.88	0.061
Q10	2.077	0.735	2.76	0.006**	2.194	0.998	2.69	0.007**
Q11	-0.523	0.576	-0.91	0.364	-0.554	0.998	0.91	0.364
Q12	-1.561	0.839	-2.05	0.051	-1.565	0.998	-2.19	0.029*
Q13	-1.338	0.649	-2.06	0.189	-0.835	0.998	-1.32	0.185
Q14	0.395	0.569	0.69	0.488	0.139	0.998	0.23	0.815
Q15	-0.523	0.576	-0.91	0.364	0.001	0.998	0.00	0.999

Note: * Significant at a significance level of 0.05.
**Significant at a significance level of 0.01.

The results in Table 4 show the following:

1) The estimates of vocabulary difficulty ranged between -2.562 and 2.007 in the computerized electronic test using the Rasch model. All of them are within the acceptable range which is between -2.95 and 2.95. The estimates of vocabulary difficulty ranged between -3.483 and 2.194 in the paper test using the Rasch model, and all of them are within the acceptable range which is between -2.95 and 2.95 except for item No. 3.

2) All chi-square values for the items of the electronic achievement test are not statistically significant except for item 8 which is statistically significant at a significance level of 0.05 and item 10 which statistically significant at a significance level of 0.01. All chi-square values for the items of the paper achievement test are not statistically significant except for items 3 and 10 which are statistically significant at a significance level of 0.01. The terms 6, 8 and 12 are statistically significant at the 0.05 level.

Table 5. The reliability coefficient of the achievement test according to the test application method (Electronic paper) using the Rasch model.

Test	Electronic	Paper
Test reliability coefficient	0.754	0.735

Table 5 shows that the stability coefficient of the electronic test is equal to 0.754 and the stability coefficient of the paper test is equal to 0.735 which are good stability coefficients that confirm the validity of the test in its electronic and paper forms.

4.3. Answering the Third Question: What is the Best Model for Judging the Efficiency of Information in the Achievement Test According to the Two Methods of Applying the Electronic and Paper Test?

The items of the computerized electronic achievement test and the paper test were analyzed using the statistical program (STATA) through the post estimation tool to calculate the information criterion (Akaike's, Bayesian). Table 6 shows the results.

Table 6. Information function criterion (Akaike's, Bayesian) according to the two methods of applying the test (Electronic paper) using the Rasch model.

Models	Electronic test	Paper test
Akaike's (AIC)	1457.21	1484.023
Bayesian(BIC)	1494.075	1520.888

The results in Table 6 showed that the value of the information function criterion for the computerized electronic test using Akaike's model (AIC) is equal to 1457.21 using the Bayesian model is equal to 1494.075 and the value of the information function criterion for the paper test using Akaike's model (AIC) is equal to 1484.023 and using the Bayesian model is equal to 1520.888. The value of the information criterion (AIC) for the computerized electronic test is less than the value of the information criterion (AIC) for the paper electronic test, and the value of the information criterion (BIC) for the computerized electronic test is less than the value of the information criterion (BIC) for the paper electronic test. This indicates that the best model for judging the efficiency of information is the model that depends on applying the test electronically.

5. Discussion

The result of the first question is that the highest mean ability estimate for the study sample was for the computerized electronic achievement test while the lowest mean ability estimate was for the paper-based achievement test. The results also showed no statistically significant differences between the mean ability estimate for both the computerized electronic test and the traditional paper-based test indicating that the computerized electronic test provides the same level of ability estimate as the traditional paper-based test. Previous results also showed that the mean standard error of the ability estimate for the computerized electronic test was significantly lower than the mean standard error of the ability estimate for the traditional paper-based test, demonstrating the accuracy of the computerized electronic test in estimating the standard error of students' abilities compared to the traditional paper-based test.

This result is consistent with the results of the study. Akdemir and Oguz (2008) indicated no statistically significant differences between students' performance in the computerized and paper-based tests. Most students also agreed that computerized tests are easy to use and more acceptable to them than traditional objective tests.

The results of the second question show a good fit of the data to the assumptions of the Rasch model in light of the two methods of applying the test (electronic and paper). This procedure is necessary before relying on any of the item response models. This is consistent with the results of the study by AlQuraan and Kuwaiti (2017) which indicated a good fit of the data with the model used by calculating the psychometric properties of the items and that the Rasch model is one of the most effective item response theory models in providing good information about estimating the parameters of the items.

The results indicate differences in the estimates of both vocabulary difficulty and chi-square values for the electronic and paper tests and that all vocabulary difficulty estimates fall within the acceptable range for the computerized electronic test. As for the paper test, all vocabulary difficulty estimates fall within the acceptable range except for item 3. All chi-square values for the computerized electronic test items are not statistically significant except for items 8 and 10. All chi-square values for the paper test items are not statistically significant except for items numbered (3, 6, 8, 10 and 12). The test reliability coefficient value was psychometrically acceptable according to the two test application methods (electronic paper) which indicate that the data are consistent with the Rasch model in the electronic test compared to the paper test.

The results of the third question show that the best model for assessing information efficiency is one based on the computer-based application of the Rasch model compared to the traditional paper-based test. Typically, when selecting a model, one accepts that models only approximate reality. Given the dataset, the goal is to determine which candidate models best approximate the data ensuring that information loss is minimized.

The Akaike and Bayesian information criteria are criteria for choosing between nested statistical models, and are essentially a measure of the quality of each of the available statistical models in terms of their relevance to each other for a given set of data, making them an ideal method for selecting models. They also help the researcher estimate the missing information if a particular model is used to represent the process that generated the data. This result is consistent with the results of the study by Berg and Lu (2014) which showed positive attitudes among students towards using computerized tests that provide more information compared to traditional paper tests. The researcher believes that this result is due to the fact that the information function curve for the computerized electronic test is better at providing information and has a lower standard error than the information function curve for the paper test using the Rasch model as shown in Figures 3 and 4.

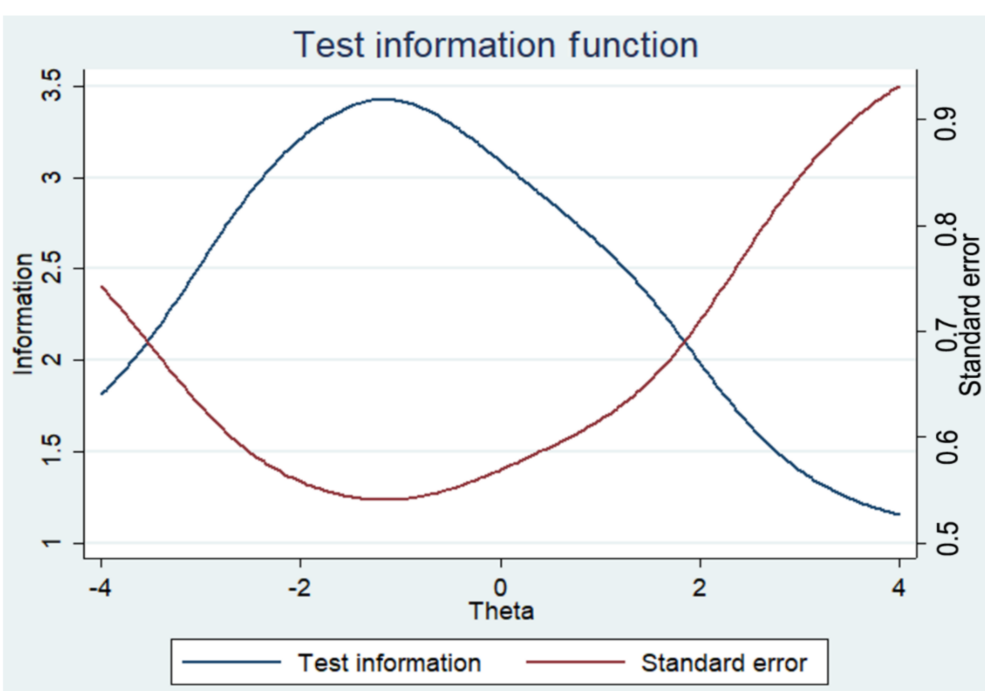

Figure 3. Information function of the computerized electronic test using the Rasch model.

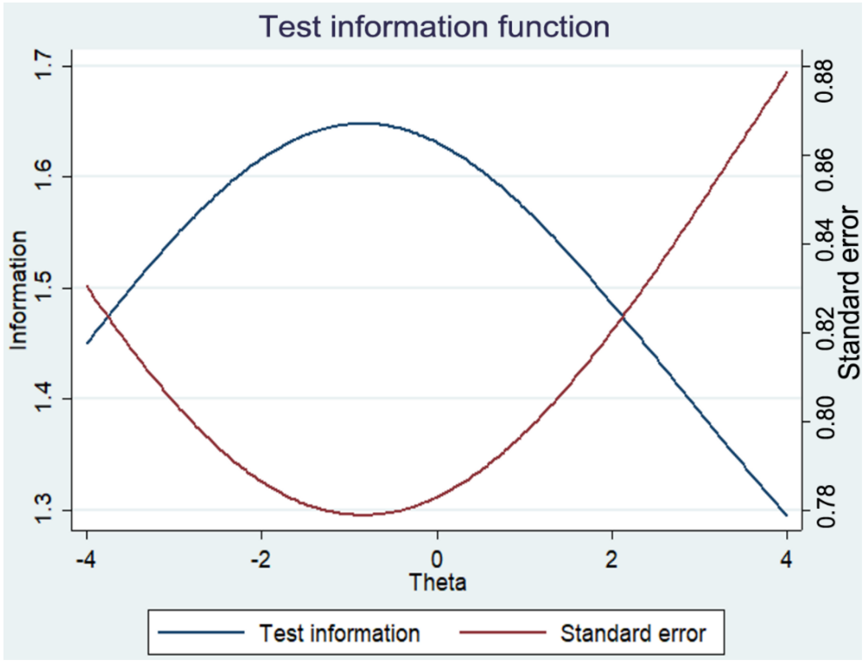

Figure 4. Paper test information function using the Rasch model.

The study explains that the computerized electronic test helps in varying the grades that the student deserves more than the paper test. It also provides a high degree of accuracy in measurement by reducing measurement errors that usually accompany the process of estimating the scientific level of the tester, meaning that it is more stable and reliable than its counterpart, traditional paper tests, and is characterized by a high level of security protection compared to tests in general. Electronic tests also take into account the cognitive style of the test takers by adopting the determination of the type of questions at the beginning of the test, the number of its questions, and their difficulty, unlike traditional paper tests based on the unit of type and size of the test. Therefore, the method of applying the test electronically provides the greatest possible amount of clear information compared to the method of applying the test on paper.

6. Conclusion

The study aimed to reveal the effectiveness of a computerized electronic achievement test using the Moodle platform in measuring accuracy in light of the item response theory in the course of psychological measurement and educational evaluation among university students.

The results of this study help in developing educational systems and their outcomes and achieving objectivity in educational measurement and evaluation using modern theory of measurement. The use of multiple-response item response models in estimating the features of vocabulary and individuals is positively reflected in the quality of estimates, which in turn develops future skills among teachers.

The results of this study also help in developing electronic tests that are compatible with evaluating different types of students with different learning styles and keeping pace with regional and global variables. They help enhance competitiveness in the educational system to achieve Oman Vision 2040, and prepare awareness programs for the educational community about the importance of employing computerized electronic tests in evaluating the learning process, and their role in raising the levels of professional performance of teachers and students, and the necessity of including teacher preparation programs in programs on designing and employing electronic tests in educational evaluation according to the modern theory of measurement that ensures a lifelong learning system.

References

Akdemir, O., & Oguz, A. (2008). Computer-based testing: An alternative for the assessment of Turkish undergraduate students. *Computers & Education, 51*(3), 1198-1204. https://doi.org/10.1016/j.compedu.2007.11.007

AlQuraan, M., & Kuwaiti, A. A. (2017). Differential item functioning in students rating of teaching effectiveness surveys in higher education according to academic disciplines: Data from a Saudi University. *Journal of Educational and Psychological Studies, 11*(4), 770-780. https://doi.org/10.53543/jeps.vol11iss4pp770-780

Bennett, R. E. (2002). Using computer technology to assess student learning. *The Journal of Educational Research, 95*(6), 328–336.

Berg, R. D., & Lu, Y. (2014). *Student attitudes towards using Moodle as a course management system.* Paper presented at the International Conference on Recreation and Leisure Industry & Language Application, 327-336.

Clariana, R., & Wallace, P. (2002). Paper–based versus computer–based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology, 33*(5), 593-602. https://doi.org/10.1111/1467-8535.00294

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *Turkish Online Journal of Educational Technology, 2*(3), 13-19. https://eric.ed.gov/?id=EJ1101956

James, R. (2016). Tertiary student attitudes to invigilated, online summative examinations. *International Journal of Educational Technology in Higher Education, 13*, 1-13. https://link.springer.com/article/10.1186/s41239-016-0015-0

Kreiner, S., & Christensen, K. B. (2012). Person parameter estimation and measurement in Rasch models. *Rasch Models in Health*, 63-78. https://doi.org/10.1002/9781118574454.ch4

Lian, L. H., & Idris, N. (2006). Assessing algebraic solving ability of form four students. *International Electronic Journal of Mathematics Education, 1*(1), 55-76. https://doi.org/10.29333/iejme/171

Linacre, J. M. (2018). *Facets Rasch model computer program. Chicago, IL. Winsteps. com.* Retrieved from www.winsteps.com

Liu, D. T., Phillips, K. M., Speth, M. M., Besser, G., Mueller, C. A., & Sedaghat, A. R. (2022). Item response theory for psychometric properties of the SNOT-22 (22-Item Sinonasal Outcome Test). *Otolaryngology–Head and Neck Surgery, 166*(3), 580-588. https://doi.org/10.1177/01945998211018383

Madison, M. J., Wind, S. A., Maas, L., Yamaguchi, K., & Haab, S. (2024). A one-parameter diagnostic classification model with familiar measurement properties. *Journal of Educational Measurement, 61*(3), 408-431. https://doi.org/10.1111/jedm.12390

Pan, J.-X., Fang, K.-T., Pan, J.-X., & Fang, K.-T. (2002). Maximum likelihood estimation. *Growth Curve Models and Statistical Diagnostics*, 77-158. https://doi.org/10.1007/978-0-387-21812-0_3

Parkin, J. R., Frisby, C. L., & Wang, Z. (2020). Operationalizing the simple view of writing with the Wechsler individual achievement test. *Contemporary School Psychology, 24*(1), 68-79. https://doi.org/10.1007/s40688-019-00246-z

Prisacari, A. A., Holme, T. A., & Danielson, J. (2017). Comparing student performance using computer and paper-based tests: Results from two studies in general chemistry. *Journal of Chemical Education, 94*(12), 1822-1830. https://doi.org/10.1021/acs.jchemed.7b00274

Reeve, B. (2024). Item response theory [IRT]. In Encyclopedia of Quality of Life and Well-Being Research. In (pp. 3689-3697). Cham: Springer International Publishing.

Robitzsch, A. (2023). Relating the one-parameter logistic diagnostic classification model to the Rasch model and one-parameter logistic mixed, partial, and probabilistic membership diagnostic classification models. *Foundations, 3*(3), 621-633. https://doi.org/10.3390/foundations3030037

Umobong, M. E. (2017). The one-parameter logistic model (1PLM) and its application in test development. *Advances in Social Sciences Research Journal, 4*(24). https://doi.org/10.14738/assrj.424.3938

Wilson, M. (2023). *Constructing measures: An item response modeling approach*. New York: Routledge. https://doi.org/10.4324/9781003286929.