Journal of Academia Educational Researches
e-ISSN: 2619-9351
Volume 11 No. 2 (2026)

Fitting objective achievement test items in inferential statistics for first-year social sciences students to the two-parameter logistic model of Item Response Theory.

Hani Dana
Mohamed Khider University of Biskra (Algeria)
hani.dana@univ-biskra.dz
Guenifi Abdelmalek
Mohamed Lamine Debaghine University of Setif 2 (Algeria)
a.guenifi@cu-aflou.edu.dz

Received: 02/08/2026 ; Accepted: 07/11/2026 ; Published: 16/11/2026

Abstract

The present study aimed to examine the extent to which the items of an objective achievement test in Inferential Statistics, designed for first-year Social Sciences students, fit the requirements of Lord's two-parameter model within Item Response Theory (IRT). To this end, a multiple-choice achievement test consisting of four alternatives was developed and administered to a sample of 320 students. The descriptive method was adopted in this study.

The results showed that the assumptions of Item Response Theory were met, as the test satisfied the assumptions of unidimensionality and local independence. The findings also indicated that four test items (16, 22, 31, and 24) had to be removed because they did not meet the requirements of the two-parameter model. The values obtained for these items were statistically significant, and the data for item (24) were complete. Meanwhile, the empirical reliability coefficient of the test reached 0.81.

Keywords: Objective Achievement Test – Inferential Statistics – Lord's Two-Parameter Model – Item Response Theory.

Statement of the Problem

The design of achievement tests is a particularly important stage in the assessment process, since assessment plays a significant role in educational decision-making (Faria & Miranda, 2026). Improving the quality of test items, whether they are essay-type or objective items (multiple-choice, matching, or true/false), requires careful consideration of the successive stages of test construction, beginning with the identification of the test's purpose and extending to the formulation of its items and the empirical verification of their quality through basic statistical indicators. These include difficulty and discrimination indices, reliability, validity, and error rate.

The difficulty index expresses the level of challenge represented by an item. An item is considered easy when many students answer it correctly, and difficult when only a small number of students manage to answer it correctly (Embretson & Reise, 2000). Discrimination, on the other hand, refers to an item's ability to distinguish between students with high and low levels of performance, thereby indicating how effectively the item contributes to measuring performance (Lord, 1980).

Reliability refers to the consistency of test results when the test is administered again, and it is an essential element in ensuring that the results genuinely reflect students' level of knowledge. Verifying test validity, that is, the extent to which a test measures what it is intended to measure, is equally

important. Validity is established through its correspondence with external criteria that reflect a comparable level of performance (Bell & Whedall, 2025). As for the error rate, it serves as an indicator of the degree of accuracy of the results obtained from an item (DeMars, 2010). Furthermore, the different estimation methods used to verify the statistical and substantive homogeneity of test items must also be taken into consideration (Ippel & Magis, 2020). These indicators provide a basis for judging the objectivity of a test and working towards achieving objectivity in measurement, which represents the ultimate goal of the assessment process.

The concept of objectivity in measurement has attracted considerable attention among researchers in the fields of educational assessment and measurement, given its complexity and the lack of clarity surrounding its indicators. Consequently, several theories have emerged in educational and psychological measurement in an attempt to establish and explain the concept of objectivity from different perspectives. Among these is Item Response Theory (IRT), which has emerged as a contemporary approach to test construction. It introduces statistical indicators that explain the relationship between an individual's performance on a test item and several fundamental concepts underlying objectivity in measurement, including unidimensionality, local independence, and the item characteristic curve (Selçuk & Demir, 2024).

According to Brown et al. (2015), unidimensionality represents the fundamental assumption of Item Response Theory, as it refers to a test's ability to measure a single latent trait. Local independence, meanwhile, means that the probability of answering test items correctly is independent across items. The item characteristic curve, in turn, illustrates the response pattern of an item and its relationship with the individual's ability level, in accordance with the model being applied (Jumini & Retnawati, 2022).

Item Response Theory has introduced several probabilistic models for examining the objectivity of measurement in achievement tests, depending on their empirical characteristics and the parameters adopted. Among these models is Lord's two-parameter model, which assumes that test items are governed by two parameters: the difficulty parameter and the discrimination parameter. The difficulty parameter indicates the degree of difficulty of an item and determines its location on the ability scale, whereas the discrimination parameter reveals differences among examinees through the slope of the item characteristic curve (Tolomeo et al., 2026).

The present study adopts this model to examine the extent to which the items of an achievement test in Inferential Statistics, designed for first-year Social Sciences students at the University of Setif 2, meet the requirements of Lord's two-parameter model within Item Response Theory. This is pursued by addressing the following research questions:

1. Do the items of the achievement test in Inferential Statistics for first-year Social Sciences students satisfy the assumption of unidimensionality in Item Response Theory?

2. Do the items of the achievement test in Inferential Statistics for first-year Social Sciences students satisfy the assumption of local independence in Item Response Theory?

3. What is the significance of the reliability coefficient of the achievement test items in Inferential Statistics for first-year Social Sciences students, according to Item Response Theory?

4. To what extent do the items of the achievement test in Inferential Statistics for first-year Social Sciences students meet the requirements of Lord's two-parameter model (difficulty and discrimination)?

Objectives of the Study

- To examine whether the items of the achievement test in Inferential Statistics for first-year Social Sciences students satisfy the assumption of unidimensionality in Item Response Theory.
- To examine whether the items of the achievement test in Inferential Statistics for first-year Social Sciences students satisfy the assumption of local independence in Item Response Theory.
- To verify whether the items of the achievement test in Inferential Statistics for first-year Social Sciences students possess the required psychometric properties according to Item Response Theory.
- To determine the extent to which the items of the achievement test in Inferential Statistics for first-year Social Sciences students fit the requirements of Lord's two-parameter model (difficulty and discrimination).

Significance of the Study

The objectivity of measuring students' academic achievement has become a subject of considerable interest among researchers in the fields of psychology and educational sciences. This is because measurement tools, including achievement tests, serve as a basis for making appropriate educational decisions. A lack of knowledge of the procedures involved in constructing these tests and interpreting their results may lead to bias in educational decision-making. Such bias may arise from ambiguity in test items, failure to take individual differences into account, or the tendency of certain items to favour one sample over another. It may also result from a lack of clarity regarding the assessment framework adopted by these tests, particularly whether they are norm-referenced or criterion-referenced.

Furthermore, the methods employed by Classical Test Theory to improve the quality of test items have become limited, given their dependence on the standardisation sample and the difficulty of generalising results due to differences in the characteristics of other environments. This raises the issue of measurement objectivity in the field of academic achievement.

Accordingly, the significance of the present study lies in its attempt to ensure objectivity in measuring academic achievement in university education through the application of the two-parameter model of Item Response Theory, namely Lord's model, in the construction of an achievement test for the Inferential Statistics course. It also seeks to encourage the adoption of Item Response Theory, with its various models, including Lord's model, in the construction of achievement tests, while taking into consideration the procedures required for their development in accordance with this theory, in order to ensure objectivity in both measurement and results.

Definitions of the Study Terms

Achievement Test:

It is an objective multiple-choice test consisting of four alternatives and comprising 34 items. The test is designed for first-year students enrolled in the Common Core of Social Sciences, in the Inferential Statistics course, during the academic year 2023–2024.

Two-Parameter Model:

It is one of the unidimensional models of Item Response Theory, also known as Lord's model. This model estimates the probability that an individual will answer an item correctly on the basis of two parameters: difficulty and discrimination. Both parameters are treated as variables, while the guessing factor is assumed to be absent and is therefore assigned a value of zero (0), according to the following relationship (Fu et al., 2026).

$$P_{ij} = \frac{\exp(s_i\theta_i + d_i)}{1 + \exp(s_i\theta_i + d_i)}$$

Inferential Statistics:

It is a course taught to first-year Social Sciences students during the second semester. It belongs to the Teaching and Methodology Unit and covers several topics, including sampling, hypotheses, parametric and non-parametric methods, and statistical significance.

Previous Studies

Jumini & Heri Retnawati's Study (2022), entitled "Estimating Item Parameters and Student Abilities: An IRT 2PL Analysis of Mathematics Examination"

This study aimed to examine the quality of the items of a mathematics achievement test using the two-parameter model of Item Response Theory. The researcher adopted the descriptive method and used a multiple-choice achievement test consisting of 30 items, which was administered to a sample of 373 tenth-grade students.

The results revealed that two items did not fit the requirements of the two-parameter model. Students' ability estimates ranged from -4.69 to 4.90 logits, with a mean of 0.05 logits.

Alimat, Mahdi's Study (2021), entitled "Constructing a Mathematics Achievement Test According to Item Response Theory (Two-Parameter Model)"

This study aimed to construct an achievement test in mathematics, specifically the Integration Unit, for second-year secondary school students in the scientific stream, based on the two-parameter model of Item Response Theory.

The test consisted of 40 multiple-choice items, each with four alternatives. It was administered to the entire study sample of 679 male and female students enrolled in the second year of the scientific stream at public secondary schools affiliated with the Directorate of Education in the Mafraq District, during the second semester of the 2017/2018 academic year. The BILLOG-MG3 programme was used to analyse the test results.

The findings revealed that five items did not fit the two-parameter model, reducing the final version of the test to 35 items. The empirical reliability coefficient of the test reached 0.82.

The assumptions of Item Response Theory were also examined and verified. Unidimensionality was assessed using Takane's index and the root mean square of the residuals, while local independence was examined using the ZQS index through the LDID programme.

Jabbar Saleh, Nawal's Study (2020), entitled "The Effect of Selecting an Item Response Theory Model on the Accuracy of Estimating the Parameters of Dichotomous Items"

The present study aimed to examine the effect of selecting different dichotomous Item Response Theory models (one-parameter, two-parameter, and three-parameter models) on the estimation of item parameters and the accuracy of these estimates.

To achieve the study's objectives, a Digit Reversal Test designed to assess cognitive abilities was administered. The test consisted of 30 dichotomous items and was applied to a sample of 640 sixth-year scientific-stream secondary school students from schools in Baghdad Governorate.

The assumptions of Item Response Theory were examined, including unidimensionality, local independence, the item characteristic curve, and the absence of a speed factor. All these assumptions

were found to be satisfied. Furthermore, the items were examined for their fit to the three Item Response Theory models, and they were found to fit these models.

The results revealed the following:

- There were no statistically significant differences in the estimation of the item difficulty parameter attributable to the model factor.
- There were no statistically significant differences in the estimation of the item discrimination parameter attributable to the model factor.
- There were no statistically significant differences in the estimation of individuals' ability parameters attributable to the model factor.
- The Digit Reversal Test, designed to assess cognitive abilities, was found to fit the one-parameter model.

Abdel Aal, Sabri Mohamed Ismail, and Al-Anzi, Mohammed Owaid Raja's Study (2019), entitled "The Effect of Multiple-Choice and True/False Item Formats on the Psychometric Properties of a Computer Science Test for First-Year Secondary School Students in Tabuk City According to Item Response Theory"

This study aimed to identify the effect of multiple-choice and true/false item formats on item parameters and the psychometric properties of a test according to the two-parameter model of Item Response Theory.

To achieve this objective, two forms of a Computer Science test were constructed: one consisting of multiple-choice items and the other consisting of true/false items. Each form contained 24 items in its final version and was administered to a sample of 421 first-year secondary school students in Tabuk City.

The findings revealed no statistically significant differences in item difficulty between the two test formats, nor were there any significant differences in the empirical reliability coefficient. However, statistically significant differences were found in the item discrimination parameter, in favour of the true/false test.

Al-Salami, Abdullah bin Mohammed's Study (2018), entitled "Constructing a Criterion-Referenced Mathematics Test Using Item Response Theory Based on the Two-Parameter Logistic Model for Fifth-Grade Primary School Students"

This study aimed to employ the two-parameter model of Item Response Theory in constructing a criterion-referenced achievement test in mathematics for fifth-grade primary school students in Tabuk Governorate.

An achievement test consisting of 32 items was developed and administered to a sample of 400 students. The descriptive method was adopted in the study.

The results indicated that the assumptions of unidimensionality, local independence, and freedom from the speed factor were satisfied. The findings also demonstrated that the items met the requirements of the two-parameter model.

Second: Research Procedures

1. Scope of the Study

A. Geographical Scope: The study was conducted in the Social Sciences Department at the Faculty of Humanities and Social Sciences, Mohamed Lamine Debaghine University of Setif 2.

B. Time Scope: The second semester of the 2023/2024 academic year.

C. Human Scope: First-year students enrolled in the Common Core of Social Sciences.

2. Research Method

The researcher adopted the descriptive-analytical method, which involves describing the phenomenon or characteristic under investigation and analysing it quantitatively in order to determine the extent to which that characteristic is present among the members of the sample representing the study population.

3. Study Population and Sample

The researcher defined the study population as all first-year students enrolled in the Common Core of Social Sciences at Mohamed Lamine Debaghine University of Setif 2 during the 2023/2024 academic year. Their total number was 1,848 students, including 345 male students, representing 18.67% of the entire population, and 1,501 female students, representing 81.33% of the total population.

Accordingly, the researcher selected a sample of 320 students using simple random sampling. The sample consisted of 60 male students and 260 female students.

4. Study Instrument

In this study, the researcher designed an achievement test for the Inferential Statistics course in the form of a multiple-choice test consisting of four alternatives. The construction of the test followed these procedures:

- Determining the purpose of the test:

The purpose was to measure the academic achievement of first-year Social Sciences students at Mohamed Lamine Debaghine University of Setif 2 in the Inferential Statistics course.

- Analysing the content of the course and determining the relative importance of each topic:

This involved identifying the topics covered in the Inferential Statistics course, which included: (hypotheses, sampling, parametric and non-parametric methods for calculating correlation and differences, and statistical significance).

The relative importance of each topic was then determined as follows:

Table No. (1): Relative Importance of the Topics Covered in the Inferential Statistics Course for First-Year Social Sciences Students.

Topic	Number of Sessions	Relative Importance
Sampling	2	25%
Hypotheses	1	12.5%
Parametric and Non-Parametric Methods for Calculating Correlation and Differences	4	50%
Statistical Significance	1	12.5%
Total	8	100%

- Formulating the Behavioural Objectives of the Test:

After determining the relative importance of the topics, the next stage involved formulating the behavioural objectives. A total of 17 behavioural objectives were established for the test.

- Preparing the Table of Specifications:

At this stage, the topics covered in the course and the number of sessions allocated to each were identified in order to determine the behavioural objectives, the cognitive levels they measure, and the number of items to be included in the test, as illustrated in the following table:

Table No. (2): Table of Specifications for the Achievement Test in the Inferential Statistics Course for First-Year Social Sciences Students.

Topic	No. of Sessions	Relative Importance of Topics	No. of Objectives	Knowledge / Recall	Comprehension	Application	Analysis	Synthesis	Evaluation	No. of Items
Sampling	1	12.5 %	2	2	1	1	1	0	0	4
Hypotheses	1	12.5 %	2	1	1	2	0	0	0	4
Parametric and Non-Parametric Methods for Calculating Correlation and Differences	4	50%	9	3	5	8	2	0	0	18
Statistical Significance	2	25%	4	4	3	1	0	0	0	8
Total	8	100 %	17	10	10	12	2	0	0	34
Percentage	100%	100 %	100 %	29.41 %	29.41 %	35.29%	5.89%			

- Constructing the Test Items:

After establishing the behavioural objectives and the cognitive levels they measure, the researcher proceeded to construct the test items. The final number of items was set at 34 multiple-choice questions, each with four alternatives. In doing so, the researcher observed the criteria for constructing

objective test items, particularly clarity, focusing on a single topic, and avoiding any clues that might suggest the correct answer.

After completing the construction procedures, the test was administered to the study sample.

Analysis and Discussion of the Study Results

Analysis and Discussion of the First Research Question

Do the items of the achievement test in the Inferential Statistics course for first-year Social Sciences students satisfy the assumption of unidimensionality in Item Response Theory?

To answer this question, the Winsteps software was used to verify the unidimensionality assumption of the test. The following results were obtained:

Table No. (3): Results of Testing the Unidimensionality Assumption of the Achievement Test

Component	Eigenvalue	Observed Variance	Expected Variance
Total variance in responses	42.4	100%	100%
Variance explained by the principal factor	8.7	20.3%	20.3%
Variance explained by persons	5.1	12.1%	12%
Variance explained by items	3.5	8.3%	8.3%
Total unexplained variance	34	79.7%	79.7%
Variance explained by the first factor	3.4	7.9%	—
Variance explained by the second factor	1.2	3%	—

Based on the results presented in the table, it can be observed that the eigenvalue of the total variance in responses reached 42.4. In order to establish the assumption of unidimensionality, the obtained results must be compared against the specified criteria:

1. The observed variance explained by the principal factor should range between 20% and 80%, as established by Linacre. This criterion was satisfied, as the observed variance explained by the principal factor reached 20.3%.

2. The eigenvalue of the variance explained by the first factor in the residuals should not exceed 5. This criterion was also satisfied, as the obtained value was 3.4.

3. The percentage of variance explained by the second factor should not exceed 3%. This condition was met, as the percentage of variance explained by the second factor in this test reached 3%.

Accordingly, it can be concluded that the test satisfied the assumption of unidimensionality, based on the observed variance percentage, the eigenvalue of the variance explained by the first factor, and the percentage of variance explained by the second factor. Unidimensionality is primarily determined by the dominant factor underlying the structure of the test, indicating the presence and predominance of a single factor governing an individual's performance on the test (Yiğiter & Boduroğlu, 2024).

This finding is consistent with the studies conducted by Al-Salami, Abdullah bin Mohammed (2018), Sabri et al. (2019), Alimat, Mahdi (2021), and Jabbar Saleh, Nawal (2020), in which the assumption of unidimensionality was verified through factor analysis. However, unlike these studies, the present research relied on residual analysis using Winsteps software to verify the unidimensionality assumption.

Analysis and Discussion of the Second Research Question

Do the items of the achievement test in the Inferential Statistics course for first-year Social Sciences students satisfy the assumption of local independence in Item Response Theory?

To answer this question, the Q3 index developed by Yen (1984) was used through Winsteps software to measure the correlations between items and verify the assumption of local independence. The following results were obtained:

Table No. (4): Results of Testing the Local Independence Assumption Using the Q3 Index

Item Pair	Correlation Coefficient
Item 24 – Item 33	0.27
Item 4 – Item 22	0.24
Item 22 – Item 33	-0.30
Item 22 – Item 24	-0.30
Item 10 – Item 12	-0.28
Item 28 – Item 32	-0.25
Item 4 – Item 32	-0.25
Item 1 – Item 22	-0.25
Item 15 – Item 22	-0.25
Item 9 – Item 10	-0.24

As shown in the table above, the correlation values ranged between -0.30 and 0.27. These values are considered acceptable according to Yen's (1984) criterion, which established 0.30 as the upper limit for considering pairs of items statistically independent of one another.

Therefore, based on the results presented, the items can be considered independent of one another, indicating that the test satisfies the assumption of local independence.

The fulfilment of this assumption is consistent with the findings of previous studies by Al-Salami, Abdullah bin Mohammed (2018), Sabri et al. (2019), Jabbar Saleh, Nawal (2020), Alimat, Mahdi (2021), and Jumini and Retnawati (2022). In these studies, local independence was established through its correspondence with the assumption of unidimensionality. However, this approach may affect the validity of the verification process, given that unidimensionality indicates the presence of a dominant dimension rather than necessarily implying the existence of only one dimension.

Furthermore, the emergence of multidimensional models provides justification for adopting other methods to verify the assumption of local independence, such as Yen's Q3 index. This is the method employed by the researcher in the present study, and the results demonstrated that the test satisfied the assumption of local independence.

Analysis and Discussion of the Third Research Question

What is the significance of the reliability coefficient of the achievement test items in the Inferential Statistics course for first-year Social Sciences students, according to Item Response Theory?

The reliability coefficient of the achievement test was calculated using Jmetrik software, and the following results were obtained:

Table No. (5): Reliability Results of the Achievement Test

Measure	Kuder–Richardson 21 (KR-21)	Split-Half Method	Standard Deviation	Standard Error of Measurement
Items	0.81	0.82	6.30	2.60540

The results presented in the table indicate that the empirical reliability coefficient was estimated using the Kuder–Richardson 21 formula and the split-half method. The reliability coefficients obtained using these two methods were 0.81 and 0.82, respectively. These values indicate the consistency of individual results and the reliability of the scores obtained.

As for validity, according to modern test theory, one of the sources of validity evidence is test reliability. This was demonstrated by the findings of the present study; therefore, the test can be considered valid.

Analysis and Discussion of the Fourth Research Question

To what extent do the items of the achievement test in the Inferential Statistics course for first-year Social Sciences students fit the requirements of Lord's Two-Parameter Model (difficulty and discrimination)?

To answer this question, Jmetrik software was used to analyse the data. The following results were obtained:

Table No. (6): Results of the Achievement Test Item Analysis and Fit Statistics

Item	Difficulty (logit)	Standard Error	Discrimination (logit)	Standard Error	Chi-Square Fit Statistic	Significance Level	Decision
1	-2.17	0.90	0.30	0.11	22.2815	0.7229	Not significant
2	-0.10	0.20	0.59	0.12	29.4705	0.3890	Not significant

3	0.20	0.22	0.54	0.12	22.3351	0.7655	Not significant
4	-0.56	0.14	1.01	0.15	35.1044	0.1668	Not significant
5	-0.16	0.13	0.97	0.15	36.5125	0.1300	Not significant
6	0.36	0.12	1.08	0.15	31.5461	0.2934	Not significant
7	-0.34	0.14	0.94	0.14	41.0247	0.0685	Not significant
8	-0.76	0.16	0.95	0.15	27.4084	0.4419	Not significant
9	-0.27	0.23	0.50	0.12	37.1511	0.1156	Not significant
10	0.18	0.11	1.12	0.15	26.3414	0.4997	Not significant
11	0.44	0.18	0.72	0.13	29.9683	0.3156	Not significant
12	-1.38	0.34	0.58	0.13	28.0159	0.4636	Not significant
13	0.27	0.17	0.70	0.13	28.7992	0.3707	Not significant
14	0.14	0.12	1.03	0.15	34.1092	0.1629	Not significant
15	-0.06	0.14	0.85	0.14	21.7211	0.7514	Not significant
16	0.84	0.16	0.99	0.15	71.8903	0.0000	Significant
17	-1.13	0.16	1.23	0.18	25.9393	0.5764	Not significant
18	0.40	0.10	1.34	0.17	26.6585	0.4823	Not significant
19	0.14	0.16	0.75	0.13	20.7040	0.8697	Not significant
20	1.63	0.46	0.47	0.12	19.8099	0.8386	Not significant
21	1.81	0.32	0.82	0.15	14.4038	0.9770	Not significant
22	-0.04	0.11	1.19	0.16	83.4979	0.0000	Significant

23	0.39	0.14	0.94	0.14	24.1610	0.6214	Not significant
24	4.50	3.14	0.09	0.13	26.8338	0.5807	Not significant
25	0.08	0.12	1.02	0.15	21.4837	0.8046	Not significant
26	0.06	0.14	0.88	0.14	21.2401	0.8503	Not significant
27	0.47	0.13	1.02	0.15	16.8727	0.9511	Not significant
28	0.85	0.16	1.16	0.16	38.9885	0.1155	Not significant
29	-0.66	0.18	0.80	0.14	36.7929	0.1235	Not significant
30	-1.27	0.18	1.13	0.18	21.3996	0.7672	Not significant
31	-1.13	0.20	0.93	0.16	49.7945	0.0095	Significant
32	0.19	0.16	0.78	0.13	31.1821	0.3091	Not significant
33	0.19	0.29	0.40	0.11	27.7116	0.4798	Not significant
34	-0.43	0.14	0.90	0.14	23.7510	0.6946	Not significant

The results presented in the table show the parameter estimates for the test items according to Lord's Two-Parameter Model (difficulty and discrimination). The difficulty parameter values were generally satisfactory, ranging from -2.17 to 1.81 logits, with the exception of Item 24, which had a very high value of 4.50 logits. This indicates that none of the respondents was able to answer this item correctly.

The discrimination parameter values, meanwhile, ranged from 0.09 to 1.34 logits. Items 1, 20, 24, and 33 had low discrimination parameter values, amounting to 0.30, 0.47, 0.09, and 0.40 logits, respectively.

To determine whether the test items fitted Lord's model, the chi-square goodness-of-fit test was used. An item is considered to misfit the model if its chi-square value is statistically significant at the 0.05 level.

The results revealed that three items did not meet the model's requirements, namely Items 16, 22, and 31. Their chi-square values were 71.8903, 83.4979, and 49.7945, respectively, and were statistically significant. This indicates that the responses to these items deviated from the model's expectations; consequently, they were excluded from the test.

The results also showed that, although Item 24 had a very high difficulty parameter value (4.50 logits) and a very low discrimination parameter value (0.09 logits), it nevertheless fitted Lord's Two-Parameter Model, as the significance level of its chi-square test was 0.5807. However, this item was excluded because its response data were extreme: none of the students was able to answer it correctly.

These findings indicate that most of the test items fitted the Two-Parameter Model. This result is consistent with the study by Debous, Mohammed (2016), which found that all test items fitted the Two-Parameter Model. It is also consistent with the study by Shehata and Ahmed (2026), which found that five items did not meet the requirements of the Two-Parameter Model compared with the Rasch Model. This demonstrates the model's flexibility in handling items with extreme difficulty and discrimination parameter values.

The findings are also consistent with those of Jumini and Retnawati (2022), Jabbar Saleh and Nawal (2020), Sabri et al. (2019), and Al-Salami, Abdullah bin Mohammed (2018). These studies reached findings similar to those of the present study regarding the fulfilment of the requirements of the Two-Parameter Item Response Theory Model. However, they differed in the estimated values of the difficulty and discrimination parameters, which may be attributed to differences in sample size and test format.

Conclusion

Objectivity in measurement is one of the main principles advocated by Item Response Theory. This theory seeks to ensure objectivity in the construction of measurement instruments, particularly achievement tests, as well as objectivity in their analysis through the use of the various models it provides. Accordingly, the Two-Parameter Model of Item Response Theory was adopted in designing an objective achievement test and examining whether its items fitted this model. The study resulted in the removal of four items from the test because their results were inconsistent with the model's expectations. The test also demonstrated an acceptable empirical reliability coefficient.

Based on the findings of the present study, the researchers make the following recommendations:

Conduct further studies on the Three-Parameter and Four-Parameter Models of Item Response Theory.

Compare the various Item Response Theory models to identify the most effective model for constructing objective tests.

Develop and construct objective item banks using the Two-Parameter and Three-Parameter Models of Item Response Theory.

Conduct comparative studies of the Two-Parameter Model in objective tests with dichotomous scoring and other tests with polytomous scoring, in order to determine the forms of testing for which the model is most accurate and effective.

References

Bell, N., & Whedall, K. (2025). A validation study of the Martin and Pratt nonword reading test. Australian Journal of Education, 69(1), 44–57.

Embretson, S. E., & Reise, S. P. (2000). Item Response Theory for Psychologists. Lawrence Erlbaum Associates.

Faria, A., & Miranda, G. L. (2026). Operationalising CTT and IRT in Spreadsheets: A Methodological Demonstration for Classroom Assessment. Analytics, 5(1), 1–20.

Fu, J., Service, T., Ho, T., & Tan, X. (2026). Item Parameter Estimation of the 2PL IRT Model With Fixed Ability Estimates: Choices of Ability Estimation Methods and Priors on Slopes. Practical Assessment, Research and Education, 30(2), 1–31.

Ippel, L., & Magis, D. (2020). Efficient Standard Errors in Item Response Theory Models for Short Tests. Educational and Psychological Measurement, 80(3), 461–475.

Jumini, & Retnawati, H. (2022). Estimating Item Parameters and Student Abilities: An IRT 2PL Analysis of Mathematics Examination. Al-Ishlah: Jurnal Pendidikan, 14(1), 385–398.

Lord, F. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale: Erlbaum.

Selçuk, E., & Demir, E. (2024). Comparison of item response theory ability and item parameters according to classical and Bayesian estimation methods. International Journal of Assessment Tools in Education, 11(2), 213–248.

Tolomeo, E., Ceraudo, L., Kolb, R., Dalton, P. H., Liuzza, M. T., & Parma, V. (2026). Improving olfactory assessment: An item response theory analysis of the American English version of the Sniffin' Sticks identification subtest. Frontiers in Psychology, 17(1), 1–13.

Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. Applied Psychological Measurement, 8(2), 125–145. http://www.copyright.com/

Yiğiter, M. S., & Boduroğlu, E. (2024). Item Response Theory Assumptions: A Comprehensive Review of Studies with Document Analysis. International Journal of Educational Studies and Policy, 5(2), 118–138. https://doi.org/10.5281/zenodo.14016086

Jabbar Saleh, Nawal. (2020). The effect of selecting an Item Response Theory model on the accuracy of estimating the parameters of dichotomously scored items. Lark Journal of Philosophy, Linguistics and Social Sciences, 3(38), 135–160.

Debous, Mohammed. (2016). Using Item Response Theory to construct a criterion-referenced mathematics test according to the Two-Parameter Logistic Model, with dichotomous and polytomous scoring. An-Najah University Journal for Research, 30(7), 1453–1480.

Al-Salami, Abdullah bin Mohammed. (2018). Constructing a criterion-referenced mathematics test using Item Response Theory according to the Two-Parameter Logistic Model for fifth-grade primary school students. Studies in Psychological and Educational Counseling, 1(1), 1–27.

Shehata, Ghada Ahmed Mohammed, & Ahmed, Ibrahim Mohammed Ali. (2026). Constructing and calibrating a criterion-referenced achievement test for the Thinking Skills course among students of the Faculty of Education at Zagazig University in light of a comparison between the One- and Two-Parameter Models of Item Response Theory. The Arab Journal of Measurement and Evaluation, 7(14), 1–142.

Sabri, Abdel Aal Mohammed Ismail, & Al-Anzi, Mohammed Owaid Raja. (2019). The effect of multiple-choice and true/false item formats on the psychometric properties of a Computer Science test for first-year secondary school students in Tabuk, according to Item Response Theory. Rifad Journal for Studies and Research, 6(1), 1–17.

Alimat, Mahdi. (2021). Constructing a mathematics achievement test according to Item Response Theory (Two-Parameter Model). An-Najah University Journal for Research, 35(7), 1079–1104.