INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

Item Analysis: Analysis of Objective Type Test Items for an Achievement Test

Perera, MAIP[1] , Senarath, S[2]

[1]Department of Educational Psychology, faculty of Education, University of Colombo

[2]Dean, faculty of Education, University of Colombo

DOI: https://doi.org/10.47772/IJRISS.2026.100800215

Received: 13 August 2026; Accepted: 18 August 2026; Published: 31 August 2026

ABSTRACT

This research paper focused on the item analysis of objective type test items constructed for the unit of fractions of the sixth-grade mathematics syllabus in Sri Lankan schools. It was conducted to analyze the difficulty level and discrimination power of each test item. Item analysis permits the selection or omits items from the test, and to develop a good achievement test paper with validated test items for the sixth-grade students. There were 65 objective type test items in the preliminary draft achievement test and was administered to a sample of 123 students of sixth-grade in the junior secondary level, from four state schools in Colombo district using purposive sampling method. The results showed that most of the test items fall within the appropriate range of difficulty index and discrimination index. However, 17 of the test items were eliminated due to very low and very high difficulty level and poor discrimination power and the remaining 48 items were selected for the final achievement test paper, which serves as both the pre-test and the post-test of the selected unit. According the findings of item analysis, relationship between the difficulty index of each item and the corresponding discrimination index was identified using the Pearson correlation coefficient. From the analysis, it was revealed that the discrimination index of selected items increased with the decrease in difficulty index and vice versa. Therefore, a negative relationship was identified between the difficulty index of each selected item and the corresponding discrimination index. On the other hand, the discrimination index of very easy and very difficult test items (rejected items) was decreased or negative. Thus, the positive relationship was identified between the difficulty index of each rejected item and the corresponding discrimination index. This analysis is highly significant for the researcher in organizing test items in which with optimum difficulty and discrimination power to achieve the learning outcomes of the unit of fractions.

Keywords: Item Analysis, Difficulty Index, Discrimination Index, Achievement Test, Test Items.

INTRODUCTION

Assessment is central to education, serving both to evaluate student learning outcomes and to ensure the quality of test instruments. Teaching and learning in classroom can be well assessed by achievement tests and the two most important elements of an achievement test are its reliability and content validity. For a test to be reliable and valid, a systematic selection of items according to the subject content and difficulty level is necessary. Marie and Sreekala (2015), and Ramatu & Kongnyuy (2025) stated that the effectiveness and quality of a test depend upon the each and every individual item. And also, they mentioned that, to determine the quality of an individual item, the item analysis should be done after administering and scoring the preliminary drafted test on a selected sample.

Achievement tests are widely used in education to measure the extent to which learners have mastered specific knowledge, skills, and competencies. Among the various forms of objective-type test items, such as; multiple-choice, true-false, matching, completion items (fill-in-the blanks items) and supplying short answer items are mainly valued because of their reliability, ease of scoring, and objectivity. However, before an achievement test can be used confidently, its items must be tested using a systematic process of standardization. Standardization ensures that test items are valid, reliable, fair, and capable of accurately measuring student achievement (Tavakol and Dennick, 2011; Ramatu & Kongnyuy, 2025).

According to Sireci (2020), educational tests are standardized so that "all examinees are tested on the same material, under the same testing conditions, and with the same scoring protocols." This uniformity is helped to create fairness and comparability among test takers.

An item analysis conducted by the Odukoya, Adekeye, Igbinoba and Afolabi (2017) were emphasized the importance of using item analysis techniques when developing multiple-choice test items. Due to the majority of the test items prepared and administered by them was not met psychometric parameters like difficulty and distractive indices and so, those items were decided to moderate and re-test or reject from the test. From this, it is revealed that prepared all test items by an instructor are not suitable for a test.

The main attention in this research study focused on, preparing a reliable and valid achievement test paper for the unit of fractions of the sixth-grade mathematics syllabus by analyzing constructed objective type test items. According to the item analysis, the main objective of this study is selection of optimal items from the test and omits poor items. Therefore, it is aim to develop a good achievement test paper with validated test items that can be used as the pre-test and the post test of the unit of fractions of the sixth-grade students.

Objectives of the study

(1) Analyzing the item difficulty level and discrimination power of objective type individual test items constructed for the unit of fractions of the sixth-grade mathematics syllabus

(2) Developing a final achievement test paper for the fraction unit of the sixth-grade math syllabus using validated test items

LITERATURE

Theoretical Background

Objective Type Tests Items

For the purpose of formative or summative tests, teachers usually use objective-type questions. A test consists of a number of questions that the test taker must answer, statements that the individuals must select or respond to, or sets of tasks that individuals must complete in order to measure a certain aspect of behaviour.

Objective-type test items are questions for which scoring does not depend on the examiner's judgment. Responses are scored as correct or incorrect according to predetermined criteria.

Common forms of objective type test items are:

(1) The polytomous format : Multiple-Choice Questions (MCQs), three or more points/options

(2) The dichotomous format : True-False items, Yes/No items, two points items

(3) Matching items

(4) Completion or fill-in-the-blank items

(5) Supplying short answer items

(Niwlikar, 2024; Marie & Sreekala, 2015; Ramatu & Kongnyuy, 2025)

These items are particularly suitable for testing the achievement because they cover a broad content area and certify objective scoring. Ramatu & Kongnyuy (2025), suggested that a broader range of item formats beyond multiple-choice questions should be integrated to improve the content validity and engagement of the mathematics examinations. This combination of diverse types of items can assess different cognitive skills and provide a more comprehensive measure of student understanding.

Item Analysis

According to Gregory (2015), cited Charles Spearman (1904) laid down the foundation for the Classical Test Theory and he revealed that later the theory was extended and revised by contemporary psychologists like Feldt & Brennan, (1989); Lord & Novick, (1968) & Kline, (1986). So, the existing classical test theory is used to analyze items in a test as a basis for developing items in this paper also.

According to American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014), item analysis is a set of techniques used to evaluate the statistical qualities of individual items in a test or psychological measure. Items (questions) are representative the psychological test that measuring a particular concept, may it be an ability, aptitude, personality, attitude or any psychological trait (Urbina, 2016). Hence, the quality of the test decides on the quality of the items in it. During test construction, test maker creates a large pool of test items and needs to select limited best possible items to form a highly reliable and valid test. Item analysis is the process that helps test maker to evaluate the test items and determine which items should be retained, which revised, and which thrown out (Gregory, 2015; & Niwlikar, 2024).

According to Tavakol and Dennick (2011), item analysis, helps to improve the validity and reliability of objective assessments by identifying weak or problematic questions. The standardization of objective type test items includes item construction, expert validation, pilot testing, item analysis, reliability estimation, and the formation of final test instrument.

Thorndike and Hagen (1964), explained the effectiveness of analyzing and using the results of objective type tests as follows:

“Giving the test, scoring it, and recording a score for each pupil frequently ends the matter as far as the teacher is concerned. However, if the teacher drops the test at this point, he loses much of its value. An analysis of the responses the pupils made to the items can serve two important purposes. In the first place, the test results provide a diagnostic technique for studying the learnings of the class and the failures to learn and for guiding further teaching and study. In the second place, the responses of pupils to the separate items and a review of the items in the light of these responses provide a basis for preparing better tests another year” (Thorndike and Hagen, 1964, e-book in 2015, p. 89).

Thorndike and Hagen (1964), raised following questions to answer for each item:

(1) How hard is the item?

(2) Does it distinguish between the better and poorer students?

(3) Do all the options attract responses, or are there some that are so unattractive that they might as well not be included?

(Thorndike and Hagen, 1964; e-book in 2015, p. 90).

According to the above theoretical review there are two criteria that are used to decide whether the item is good or not:

(1) Item Difficulty Index and (2) Item Discrimination Index (*DI*)

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

Item Difficulty Index

According to Niwlikar (2024), item difficulty indicates the proportion of the total number of test takers who answered the item correctly. The item difficulty (easiness, facility index,) is the percentage of students who answered an item correctly (Sajjad, et. al., 2020; Sugianto, 2020; Sharma, 2021 & Cakir and Ari, 2022). According to them, the difficulty index ranges from 0 (if no one answered a question correctly) to 1 (if everyone answered a question correctly) or 0% to 100%, whereas the higher values indicate the easier the question and the lower values represent the difficulty of hard items. The formula of calculating the item difficulty index is as follows.

Item difficulty index = number of correct responses/ total number of responses

Niwlikar, (2024), shows that the optimal difficulty of an item should be between 0.3 to 0.7. This range avoids ceiling effect and floor effect in the test scores. Floor Effect : Large number of test-takers achieve the lowest possible score - occurs if items are too difficult (figure 2.2) and Ceiling effect : Large number of test-takers achieve the highest possible score - occurs if items are too easy for the target population (figure 2.3).

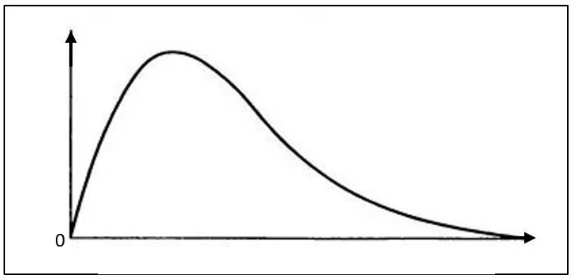

Figure 2.2: Floor Effect - Large number of test-takers achieve the lowest possible score (Positive Skewness)

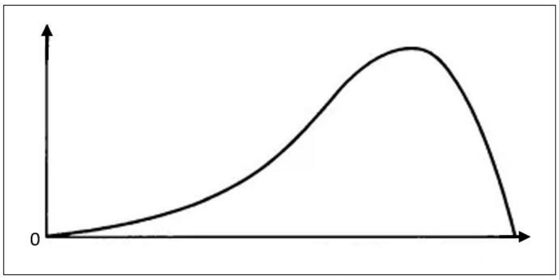

Figure 2.3: Ceiling effect - large number of test-takers achieve the highest possible score (Negative Skewness)

The majority of items should be moderately difficult. When a test has many items that are too difficult or too easy, the test scores are clustered at only one end of the curve. This prevents the recognition of differences in the behaviors or abilities of various individuals.

When reviewing literature, various ranges of difficulty indices and various interpretations presented for them could be found. Niwlikar, (2024); Suek, (2021) and Sugianto, (2020) have interpreted that the difficulty index of an item is above 0.70, it is a very easy item. However, Cakir and Ari (2022); Sharma, (2021); Marie and Sreekala (2015) & Tavakol and Dennick, (2011) have revealed that an item difficulty index is above 0.80, it is too easy to answer. On the other hand, Niwlikar, (2024); Cakir and Ari (2022); Suek, (2021); Sugianto, (2020) & Tavakol and Dennick, (2011) have presented that if the difficulty index of an item is below 0.30, it is a very difficult item hence it should be removed from the test whereas, Marie and Sreekala (2015) and Sabri (2013) have revealed that the item difficulty index is below 0.20, it is too difficult.

The above description presents different criteria used by researchers to interpret the difficulty index of test items. Most authors identified a middle range, as approximately 0.30 - 0.70, as representing items with moderate or optimal difficulty that are generally suitable for inclusion in an achievement test. Values below approximately 0.20 or 0.30 are commonly considered difficult or very difficult, whereas values above 0.70 or 0.80 are generally regarded as easy or very easy. The literature suggests that items with moderate difficulty levels are preferable because they provide better differentiation among learners and contribute to the quality of an achievement test.

Item Discrimination Index (*DI*)

The ability of an item to distinguish between high and low performers is demonstrated by the Item Discrimination Index. According to Sugianto (2020) and Gergory (2015), item discrimination is a statistic that shows an ability of an item to separate students who performed well from those who did not on a test overall. According to Sharma (2021), the index of discrimination is showed that the ability of an item to differentiate students between the superior (a group of high achievers) and inferior (a group of lower achievers). It ranges from -1 to +1. According to Gajjar et al. (2014), item discrimination is the power of an item has to separate students who have higher and lower skill levels.

To compute this index, after scoring the test items, test scores are arranged in descending order (from highest to lowest). The discrimination power of an item is calculated by categorizing the test takers into upper 27% and lower 27% according to their total test scores. The difference between the upper and lower groups is divided by the number of the test takers in the upper group or the lower group or by half of the total number of the test takers or by the total number of test takers (Rezigalla, 2022; Sharma, 2021; Sugianto, 2020). The top 27% and the bottom 27% of test takers are selected in order to conduct the item analysis, as this proportion is identified to maximize score differences in a normal distribution while maintaining a sufficient sample size. Middle 46% of the test scores were excluded as they behave in a similar pattern contributing insignificantly to discriminate the performance of students.

The formular of calculating the item discrimination power is as follows. The difference between the number of students who responded correctly in the upper group and the number of students who responded correctly in the lower group is divided by the Number of students in the upper group or lower group.

Item Discrimination Index (DI) = Ru - Rl / Nu (or) Nl

Where, ***Ru*** = the number of students in the upper group who responded correctly
Rl = the number of students in the lower group who responded correctly
Nu = Number of students in the upper group
Nl = Number of students in the lower group

After reviewing literature, the researcher could find various ranges of discrimination indices that can be used to select the appropriate items for a test and various interpretations given for them. Sharma, (2021); Suek, (2021); Sugianto, (2020); Marie and Sreekala (2015) and Sabri, (2013) have revealed that if the discrimination index of an item is 0.40 or above, it is a very good and excellent item for separating students who performed well in the test from those who did not. They also showed that, if the discrimination index of an item is in between 0.20 to 0.39, it is a marginal or reasonably good item and below 0.20 or negative, it is a worst and defective item that should be excluded from the test. In contrast, Ramatu & Kongnyuy (2025) showed that, if the discrimination index of an item is 0.30 or above it is a satisfied item and it is below 0.30 it is not a satisfied item. Some researchers, such as Niwlikar (2024), recommend 0.50 or above as the optimum level.

Empirical Studies

Boopathiraj and Chellamani (2013) investigated the quality of test items developed to assess the Research in Education subject among Master of Education (M.Ed.) student-teachers. A multiple-choice test was administered to a randomly selected sample of 200 student-teachers from different Colleges of Education. The researchers examined the difficulty index and discrimination index of the items. The findings indicated that most items fell within an acceptable range of difficulty, with items having difficulty indices between 20% and 80% being represented in the test. The study also identified items with satisfactory discrimination power, while some items were rejected because of poor discrimination. This study demonstrates the practical value of item analysis in identifying items that appropriately differentiate between higher and lower performing examinees.

Similarly, Ramatu and Kongnyuy (2025) examined the psychometric characteristics of multiple-choice items in the 2023 and 2024 General Certificate of Education Ordinary Level (GCE O/L) Mathematics examinations in Cameroon. The study involved 248 students selected from 17 public and private English-speaking secondary schools. The researchers analyzed two examination papers, each containing 50 standardized multiple-choice items. The item difficulty index was calculated based on the proportion of students who answered each item correctly, while the discrimination index was determined using the upper and lower 27% of examinees. The findings revealed that a considerable proportion of the items met acceptable psychometric standards. The study therefore provides evidence that systematic item analysis can contribute to the validity and reliability of standardized examinations and increase confidence in the interpretation of examination results.

Zulaiha et al. (2020) extended the investigation of item characteristics by comparing multiple-choice and essay-type questions in mathematics. The researchers used 22 multiple-choice items and 22 essay items administered to 452 students from six schools in Jakarta. The essay questions were derived from the multiple-choice questions by eliminating the options. The researchers employed Classical Test Theory (CTT), the Rasch model, SPSS, and Microsoft Excel to examine item difficulty and discrimination. According to CTT, multiple-choice questions were, on average, easier than essay questions, whereas the Rasch analysis indicated relatively similar levels of difficulty between the two item formats. Importantly, the study found that the average discrimination power of multiple-choice items was higher than that of essay items under both CTT and the Rasch model. These findings suggest that item format can influence psychometric characteristics and that multiple-choice items can provide effective discrimination when they are appropriately constructed.

Cakir and Ari (2022) focused specifically on the development and standardization of a multiple-choice achievement test to measure students' conceptual understanding of the cell unit in biology. Their test-development process incorporated item planning, item writing, expert review, item analysis, and item selection. An initial pool of 29 multiple-choice items was reviewed by biology subject experts, biology education specialists, and biology teachers to establish content validity. The revised test was piloted with 158 students. Following item analysis, two items with discrimination indices below 0.19, three items with difficulty indices above 0.80, and two items with difficulty indices below 0.30 were removed. Consequently, 22 items were retained for the final achievement test. The final test was administered to 123 university students, and its reliability was assessed using the KR-20 coefficient, which yielded a value of 0.89. This study provides particularly strong support for the argument that expert validation followed by empirical item analysis is an effective procedure for constructing a valid and reliable achievement test.

Marie and Sreekala (2015) examined the difficulty level and discrimination power of test items in an examination paper for B.Ed. Physical Science students. The sample consisted of 100 B.Ed. students selected using purposive sampling. An achievement test containing 100 objective-type items was administered, comprising fill-in-the-blank, true/false, one-word-answer, and multiple-choice questions. The test was developed with the assistance of subject experts and was based on selected chapters from the Teaching of Physical Science course. The researchers also prepared a test blueprint that gave appropriate weight to instructional objectives, subject content, and item formats, with Bloom's revised taxonomy being used to frame the items. For item analysis, the upper 27% and lower 27% of examinees were selected after arranging total test scores in descending order. Most items were found to have acceptable difficulty and discrimination indices. However, 17 items were eliminated because of excessively high or low difficulty indices and inadequate discrimination. The study further reported a moderate negative relationship between difficulty index and discrimination power, indicating that discrimination tends to increase as items become more difficult up to an optimum point, after which further increases in difficulty may reduce discrimination.

After reviewing the above literature: the theoretical background and the empirical studies, the various forms of objective type test items that can be included in a test paper was identified (Niwlikar, 2024; Marie & Sreekala, 2015; Ramatu & Kongnyuy 2025). Moreover, the procedure of standardization of test items and the importance of item analysis to select the optimal test items to an achievement test using the classical test theory was revealed (Tavakol and Dennick 2011; Gregory, 2015 & Niwlikar, 2024). Methods selecting a proper sample, items scoring methods, data analysis techniques and the optimal ranges of indices that can be used to select items were found and a new knowledge was obtained to develop the methodology of current study due to the literature review of empirical studies.

METHODOLOGY

Population and the Sample of the Study

All students of Sinhala medium in sixth grades of state schools in Sri Lanka, is the population of this study. A sample of 123 students in sixth grades were adopted using the purposive sampling method from four schools comprising two 1AB schools and two 1C schools in Colombo District for collecting data.

Data Collection Instrument

An Achievement test consist of 65 objective type test items was used as the data collection instrument of the study. First of all, prepared the blueprint and decided the number of items should be included under each form of objective type test items. Then, the first draft of the test paper was developed by the researcher in order to measure the learning outcomes that mentioned in the unit of fractions of the sixth-grade mathematics syllabus. Blueprint and the draft items were prepared giving due weightage to the learning outcomes, subject content and forms of questions considering the existing learning-teaching resources for the sixth-grade mathematics subject: mathematics syllabus, teacher's guide, students text book and the activity-based mathematics learning guide. In the same time, the scoring rubrics of the test items was also prepared with three different scoring rubrics, namely; Rubric - 1, Rubric - 2, and Rubric - 3. The first draft test paper was reviewed with the help of subject experts: a mathematics subject director, a mathematics education specialist, and a mathematics teacher, the content validity of the questions was ensured and some questions were revised by them. In reviewing they changed the prepared format of some items, such as: matching items were changed into fill-in the blanks items and some options of the multiple-choice items were rearranged and checked whether the content validity of all the items. The final data collection instrument, a mathematics test paper, consists of 65 items with four forms of objective type test items and it comprises thirty-eight fill-in the blanks items, ten true or false items, six matching items and eleven multiple choice questions.

Data Collection procedure

The prepared achievement test was applied to 123 students who studied in grade six classes in four schools of 1AB and 1C types as a pilot study after getting formal permission from the Zonal Education Director, the Principal of the school, class teacher and the parents. Before administering the pilot test, it was confirmed that the unit of fractions has been taught by the teachers of the selected four classes. Although, the allocated time for the test paper is one hour, students were given as much time as they needed to complete the test but, no one requested extra time. Students were instructed to write their answers and the workings only in the question paper it-self. However, all those instructions had been printed on the test paper also. Test was administered by the researcher herself with the help of class teacher.

Data Analysis

The answer scripts of the participants were scored by the researcher herself using the Rubric - 1, Rubric - 2, and Rubric - 3. Minimum score was 21 and the maximum score was 99. The average score (Mean) is 69.9, Median score is 71.7 and Mode of the scores is 76. The standard deviation (SD) is 16.1. Total scores of each student were entered in a Microsoft Excel sheet and those scores were arranged in descending order. After that, those ordered scores were grouped into the class intervals of 21 - 31, 31 - 41, ……., 91 - 101 again. Histogram was created using the grouped frequency distribution. On it, the Frequency Polygon was created by calculating the class value (mid value - x) and by adding two class intervals at the beginning and at the end with 0 frequencies. Grouped Frequency Distribution is illustrated in table 4.1 and the Histogram and the Frequency Polygon is in Figure 4.1 below.

Table 4.1 : Grouped Frequency Distribution of scores

Class Interval	Frequency (f)	Class value (x)
11 – 21	0	16
21 – 31	1	26
31 – 41	4	36
41 – 51	14	46
51 – 61	19	56

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

61 – 71	22	66
71 – 81	29	76
81 – 91	28	86
91 – 101	6	96
101 – 111	0	106

The frequency polygon drawn in figure 4.1 shows that the distribution of scores is almost negatively skewed. It means that the most of test-takers achieve the highest possible scores for the test. The mean, median and mode of the test scores were calculated and those values are 69.9, 71.7 and 76 respectively. If the values of mean < median < mode in a distribution, it has a negative skewness.

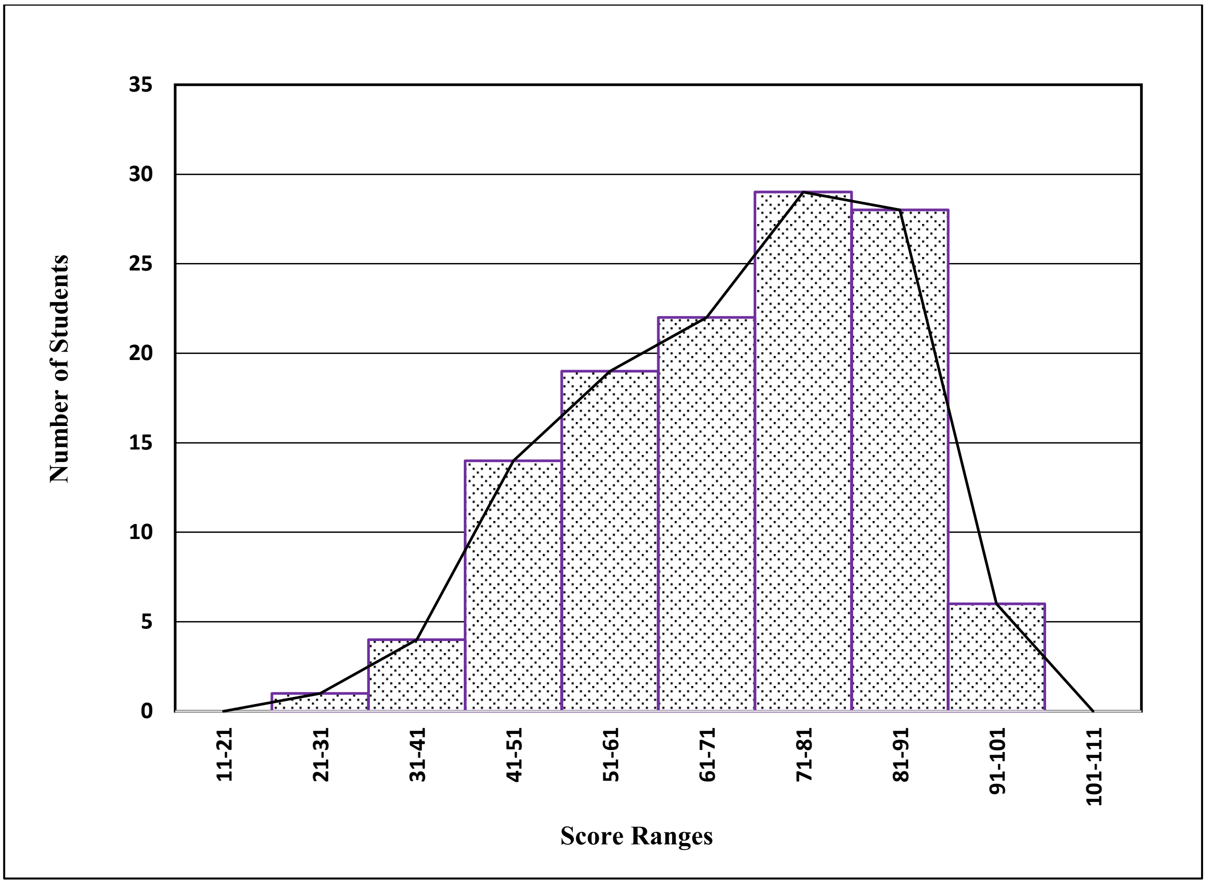

Figure 4.1: Histogram and the Frequency Polygon

The first objective of the study, is analyzing the item difficulty level and discrimination power of objective type individual test items constructed for the unit of fractions. To address it, the item difficulty was calculated by dividing the number of students who answered each item correctly by the total number of respondents. Item discrimination power was calculated using the difference between the number of students who responded correctly in the upper group and the number of students who responded correctly in the lower group was divided by the Number of students in the upper group/lower group.

To calculate the discrimination power of each item, the following steps were followed:

1. arranging the test scores of the participants in descending order
2. Identification of test takers who represent upper 27% and lower 27%, and it was 34 students in each group
3. Excluded the middle 46% of test takers from the analysis with the assumption of that they behave in the similar pattern

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

4. Calculation the discrimination power of each item, using the formula;

Following table 4.2 illustrated the calculated difficulty index and the discrimination index of each item and the decision taken about the item. Graphical representation of the difficulty indices and the discrimination indices are illustrated in the following Figure 4.2.

Table 4.2: Difficulty Index and the Discrimination Index of each item.

Item No.	Difficulty index	Discrimination index	Decision*	Item No.	Difficulty index	Discrimination index	Decision*
(01)	0.29	0.12	R	(34)	0.59	0.85	S
(02)	0.36	0.12	R	(35)	0.50	0.79	S
(03)	0.26	-0.03	R	(36)	0.63	0.79	S
(04)	0.80	0.41	S	(37)	0.53	0.74	S
(05)	0.82	0.32	R	(38)	0.46	0.68	S
(06)	0.84	0.35	R	(39)	0.50	0.82	S
(07)	0.60	0.71	S	(40)	0.67	0.71	S
(08)	0.48	0.68	S	(41)	0.55	0.68	S
(09)	0.43	0.68	S	(42)	0.70	0.44	S
(10)	0.54	0.56	S	(43)	0.53	0.74	S
(11)	0.80	0.29	M	(44)	0.50	0.76	S
(12)	0.22	0.09	R	(45)	0.68	0.71	S
(13)	0.61	0.65	S	(46)	0.66	0.53	S
(14)	0.52	0.88	S	(47)	0.73	0.59	S
(15)	0.65	0.65	S	(48)	0.59	0.79	S
(16)	0.66	0.74	S	(49)	0.77	0.59	S
(17)	0.67	0.00	R	(50)	0.69	0.65	S
(18)	0.32	0.03	R	(51)	0.45	0.65	S
(19)	0.22	-0.09	R	(52)	0.43	0.65	S
(20)	0.79	0.38	M	(53)	0.84	0.18	R
(21)	0.78	0.44	S	(54)	0.82	0.03	R
(22)	0.76	0.32	M	(55)	0.41	0.71	S
(23)	0.49	0.56	S	(56)	0.41	0.53	S
(24)	0.63	0.47	S	(57)	0.42	0.47	S
(25)	0.76	0.32	M	(58)	0.80	0.09	R
(26)	0.38	0.50	S	(59)	0.80	0.06	R
(27)	0.78	0.50	S	(60)	0.45	0.29	M
(28)	0.79	0.44	S	(61)	0.46	0.41	S
(29)	0.82	0.38	R	(62)	0.71	-0.03	R
(30)	0.80	0.50	S	(63)	0.44	0.50	S
(31)	0.53	0.74	S	(64)	0.73	0.12	R
(32)	0.52	0.82	S	(65)	0.70	-0.09	R
(33)	0.44	0.74	S				

*Decision : S - Selected; M - Moderated; R - Rejected

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

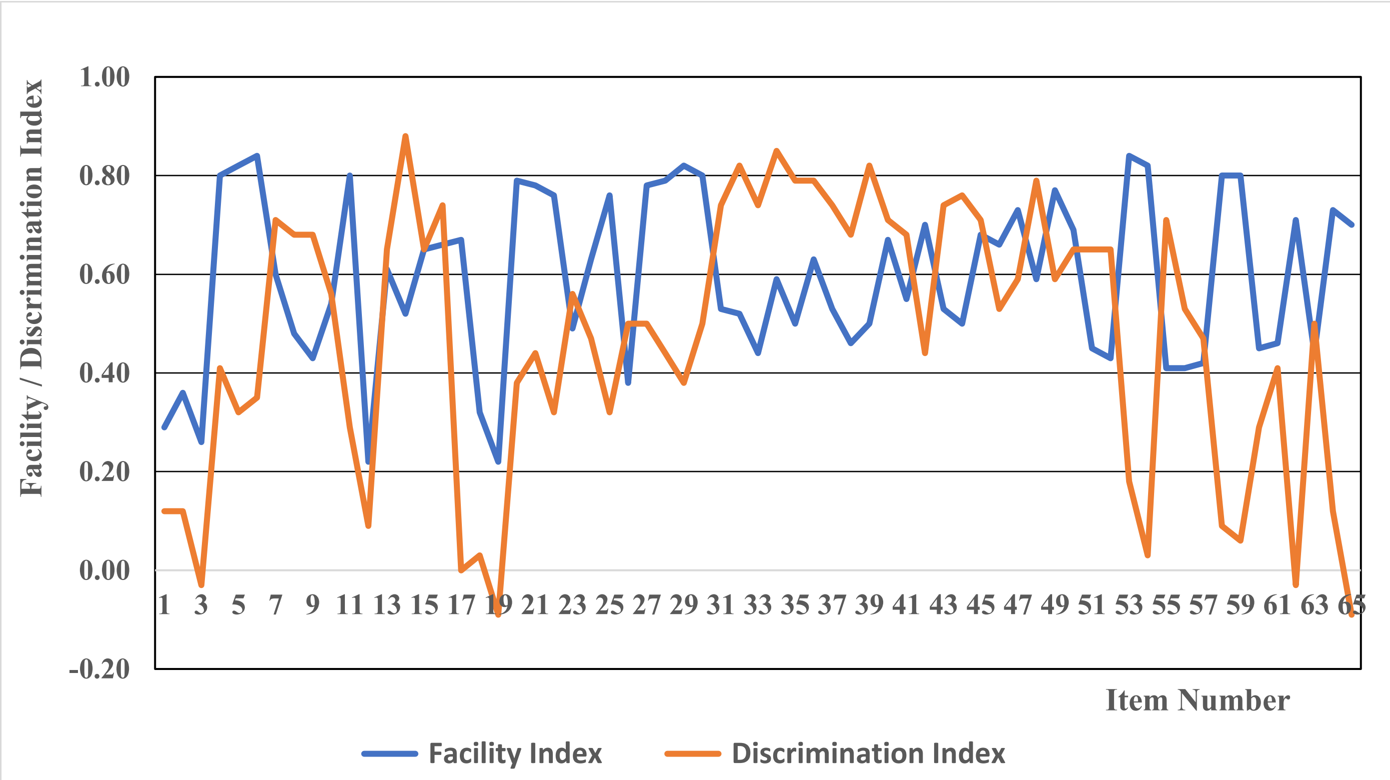

Figure 4.2 : Graphical representation of difficulty and discrimination indices in table 4.2

DISCUSSION

Niwlikar (2024) explained that excessively easy items may cause a ceiling effect, while Sugianto (2020) and Niwlikar (2024) recommended an optimal item-difficulty range of 0.30 - 0.70. In the present pilot test, the mean difficulty index was 0.59, and 41 items (63%) falling within the acceptable range. Similarly, Cakir and Ari (2022) and Tavakol and Dennick (2011) considered items with difficulty indices between 0.30 and 0.80 appropriate, while items above 0.80 or below 0.30 were considered too easy or too difficult and were excluded. Therefore, the present study retained items within the 0.30 - 0.80 range.

Regarding the discrimination index, Sabri (2013), Marie and Sreekala (2015), and Sugianto (2020) considered 0.40 or above to indicate very good items, while values between 0.30 - 0.39 were reasonably good and 0.20 - 0.29 were marginal. Items below 0.20, particularly those with negative discrimination, were considered poor and recommended for rejection. Accordingly, the present study retained items with discrimination indices of 0.40 or above, revised five items with values between 0.20 and 0.39 when their difficulty indices were acceptable, and rejected items with discrimination indices below 0.20 or negative.

Accordingly, seventeen items out of 65 (26%) were rejected either due to difficulty level or discrimination index was not into the optimal level. Forty-three items (66%) were accepted without revision while 05 (8%) items were accepted after adjusting necessary revisions. The following table 5.1 is showed that the selected, moderated and rejected items of this study.

Table 5.1 : Selected, Moderated and Rejected items

Discrimination Index / Difficulty Index	0.40 or above	Between 0.20 - 0.39	0.20 or below	Total
Above 0.80	-	5 6 29	53 54	**05**

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VIII August 2026

Between 0.30 - 0.80	4 7 8 9 10 13 14 15 16 21 23 24 26 27 28 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 55 56 57 61 63	11 20 22 25 60	2 17 18 58 59 62 64 65	**56**
Below 0.30	-	-	1 3 12 19	**04**
Total	**43**	**08**	**14**	**65**

The following two diagrams (Figure 5.1 and Figure 5.2) show the relationship of difficulty index and discrimination power of the rejected items and the relationship of difficulty index and discrimination power of the accepted items as Figure 5.1 and Figure 5.2 respectively.

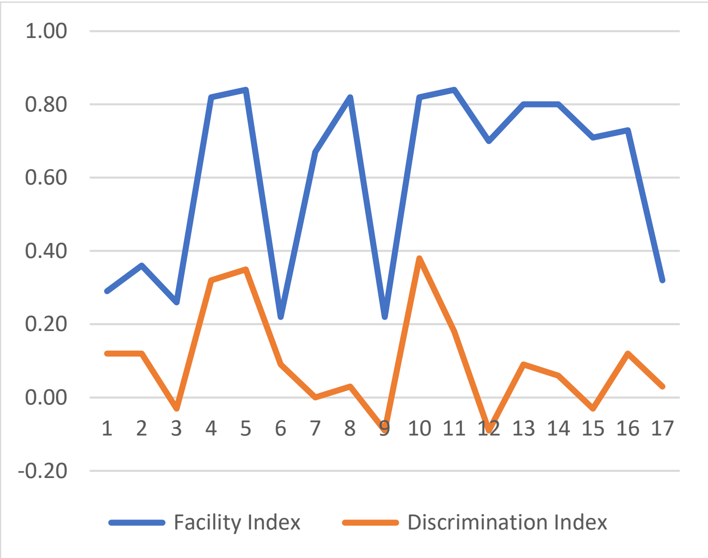

Figure 5.1: The relationship of difficulty index and discrimination power of the rejected items

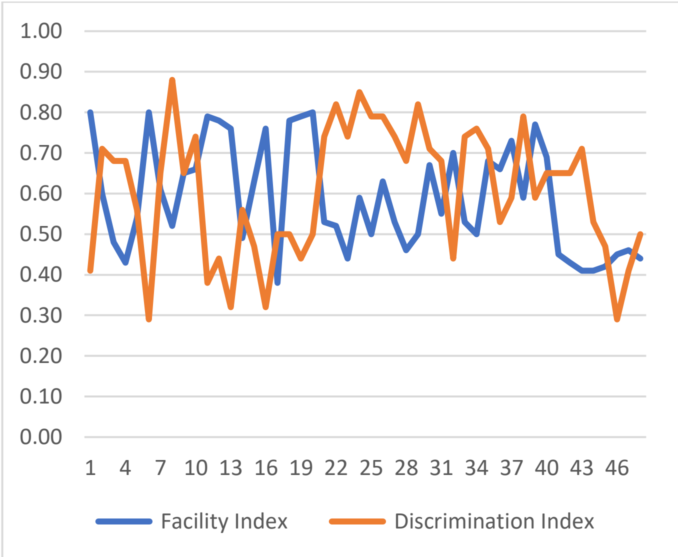

Figure 5.2: The relationship of difficulty index and discrimination power of the accepted items

According to the table 5.1 and figures 5.1 & 5.2 above, it is illustrated that if an item difficulty index is above 0.80 or below 0.30 the discrimination power of that item will not exceed 0.40 or above. Always the discrimination power of those items was below 0.40. Therefore, when an item is too easy or too difficult the discrimination power (separate students those who perform well from those who do not) of that item will be decreases. The Pearson correlation coefficient between difficulty index and the discrimination index of these items is 0.44. So, it has positive correlation between difficulty index and the discrimination index. It is clearly showed in figure 5.1 above which shows rejected items from the test. On the other hand, if the item difficulty index is in between 0.30 - 0.80, most of the items' discrimination power (43 items) is above 0.40. Those items are selected without any revision. Only five items have the discrimination power in between 0.20 - 0.39. Those are the items that moderated and selected. However, these items have significant relationship. When an item difficulty index is high the discrimination index is lower and vice versa. The Pearson correlation coefficient between difficulty index and the discrimination index of these items is -0.39. So, it has negative correlation. It is clearly showed in figure 5.2 above. These finding are almost similar to the findings of the study of Marie & Sreekala (2015).

According to the above analysis and the recommendations of Ramatu & Kongnyuy (2025), 26 fill-in the blanks items out of 38, 08 true/false items out of 10, all the 06 matching items and 08 MCQ items out of 11 were selected. Ramatu & Kongnyuy (2025), suggested that a broader range of item formats beyond multiple-choice questions should be integrated to improve the content validity and engagement of the mathematics examinations. They emphasized that this incorporation of diverse types of items can assess different cognitive skills and provide a more comprehensive measure of student understanding. Following table 5.2 illustrated the prepared items and the selected items with the form of objective type tests and the learning outcome.

Table 5.2 : Prepared items and selected items for the achievement test

Form of the Item	Fill-in the blanks		True/False Items		Matching Items		MCQ	
Learning Outcome	**Items prepared**	**Items selected**	**Items prepared**	**Items selected**	**Items prepared**	**Items selected**	**Items prepared**	**Items selected**
3.1 Identifies unit fractions and proper fractions	13	**10**	05	**04**	-	**-**	03	**-**
3.2 Selects equivalent fractions.	10	**08**	04	**04**	06	**06**	08	**08**
3.3 Compares fractions.	15	**08**	01	**-**	-	**-**		**-**
Total number of Items	38	**26**	10	**08**	06	**06**	11	**08**

FINDINGS AND CONCLUSION

The findings of this paper have significance for all test developers. To select items for an achievement test using Classical Test Theory (CTT), the most appropriate range of difficulty index is 0.30 - 0.80 and the most logical discrimination index is 0.40 or above. The items which are in these ranges are the suitable items for an achievement test. Items in this range has a common characteristic that the item difficulty index is high the discrimination index is lower than that and item difficulty index is low the discrimination index is higher than that. The study proposes that test items with good positive discrimination power and medium difficulty level are the perfect items for a good achievement test.

The common characteristic of rejected items is when the item is too easy or too difficult the discrimination power becomes lower or negative. So, in preparing test items for an achievement test conducting an item analysis is a must. The final test paper should consist only with high-quality test items which have measured the learning outcomes of the relevant unit. According to the main objective of this study, the final achievement test paper for the unit of fractions of the sixth-grade math syllabus was developed with validated 48 test items.

Another finding of this study is that there is a positive correlation between difficulty index and discrimination index of the rejected items. On the other hand, there is a negative correlation between difficulty index and discrimination index of the accepted items.

It is recommended that the item creators' attention should be drawn to include various item formats such as; fill-in-the-blanks, true-false, matching, multiple-choice test items to enhance the content validity of a test. And finally, this work should be repeated in each and every unit of the sixth-grade mathematics syllabus and the other subjects also, to develop a good item bank for each subject for the betterment of students' community.

REFERENCES

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing. Washington, D.C.: American Educational Research Association.
2. Boopathiraj. C. & Chellamani, K. (2013). Analysis of test items on Difficulty level and Discrimination Index in the test for Research in Education. International Journal of Social Science and Interdisciplinary Research. Vol. 2(2), pp. 189 -193.
3. Çakir, H. & Ari, A. G. (2022). Development and Validation of an Achievement Test in Biology. Journal of Social Sciences and Education, 5 (1), 64-75.
4. Gajjar, S.; Sharma, R.; Kumar, P. & Rana, M. (2014). Item and Test Analysis to Identify Quality Multiple Choice Questions (MCQs) from an Assessment of Medical Students of Ahmedabad, Gujarat. Indian J Community Med. 39(1):17-20. doi: 10.4103/0970-0218.126347.
5. Gregory, R.J. (2015). Psychological Testing: History, Principles, and Applications. Global Edition. Pearson Education Limited.
6. Hogan, T.P. (2019). Psychological Testing : A Practical Introduction (Fourth Edition). Hoboken, NJ: John Wiley & Sons, Inc. https://books.google.ca/books?id=K7s6EAAAQBAJ&printsec=copyright#v=onepage&q&f=false
7. Instructional Assessment Resources. (2011). Item Analysis. Retrieved November 9, 2013 from University of Texas at Austin, Instructional Assessment Resources, IAR Web site: http://www.utexas.edu/academic/ctl/assessment/iar/students/report/itemanalysis.php
8. Marie, S.M.J.A. & Sreekala, E. (2015). Relevance of Item Analysis in Standardizing an Achievement Test in Teaching of Physical Science in B.Ed Syllabus. i-manager's Journal of Educational Technology, Vol. 12, No. 3
9. Niwlikar, B. A. (2024). What Is Item Analysis? 2 Methods to identify Best Test Items for Test. Careershodh. https://www.careershodh.com/item-analysis-2/
10. Odukoya, J. A., Adekeye, O., Igbinoba, A. O. & Afolabi, A. (2017). Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. 52(3):983–997. doi: 10.1007/s11135-017-0499-2
11. Ramatu, N. & Kongnyuy, P. (2025). An Analysis of Item Difficulty and Item Discrimination of Mathematics Multiple Choice Items (2023 and 2024) in Bamboutos Division of Cameroon. East African Journal of Education Studies, 8(3), 505-511. https://doi.org/10.37284 /eajes.8.3.3600
12. Rezigalla, A. A. (2022). Item Analysis: Concept and Application. Medical Education for the 21st Century, Edited by Michael S. Firstenberg and Stanislaw P. Stawicki. DOI: 10.5772/intechopen.100138
13. Sabri, S. (2013). Item Analysis of Student Comprehensive Test for Research in Teaching Beginner String Ensemble Using Model Based Teaching Among Music Students in Public Universities. International Journal of Education and Research Vol. 1(12).
14. Sajjad, M., Iltaf, S., and Khan, R. A. (2020). Nonfunctional distractor analysis: An indicator for quality of multiple-choice questions. Pak J Med Sci, 36(5): p. 982-986.
15. Sharma, L. R. (2021). Analysis of Difficulty Index, Discrimination Index and Distractor Efficiency of Multiple-Choice Questions of Speech Sounds of English. International Research Journal of MMC (IRJMMC) Vol. 2 Issue 1 (February, 2021) ISSN 2717-4999 (Online) 2717-4980 (Print), Retrieved from https://doi.org/10.3126/irjmmc.v2i1.35126
16. Sireci, S. G. (2020). Standardization and UNDERSTANDardization in Educational Assessment Educational Measurement Issues and Practice, National Council on Measurement in Education (NCME). https://doi.org/10.1111/emip.12377Digital Object Identifier (DOI)
17. Suek, L. A. (2021). Item Analysis of an English Summative Test. Language Studies Centre, Pattimura Excellence Journal of Language and Culture, ISSN 2808-2265. Journal Home page: https://ojs3.unpatti.ac.id/index.php/pejlac
18. Sugianto, A. (2020). Item Analysis of English Summative Test: EFL Teacher-Made Test. Indonesian EFL Research and Practices, 1(1), 35–54. Retrieved from https://jurnal.iaima.ac.id/i-efl/article/view/4
19. Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education. 2011; 2:53-55, ISSN: 2042-6372. DOI: 10.5116/ijme.4dfb.8dfd

20. Thorndike, R.L. & Hagen, E.P. (1964). Measurement and Evaluation in Psychology and Education. Second Edition. New York, London, Sydney, John Wiley & Sons, Inc.
21. Urbina, S. (2016). Psychological testing. Encyclopedia of mental health, 2nd edition, Academic Press. pp.357-364. Retrieved from: https://www.researchgate.net/publication/286780211_Psychological_testing
22. Zulaiha, R., Fahmi, F., Rahdiani, D., Rahman, A. & Anfal, M. F. A. (2020). Analysis of Difficulty Level and Discriminating Power Between Multiple Choices and Essay Items on Math Test. Advances in Social Science, Education and Humanities Research, volume 545. Proceedings of the International Conference on Educational Assessment and Policy (ICEAP).