



## Item Analysis of Tool to Examine the Effect of E-Module on the Academic Achievement of Chemistry Students at Secondary Level

Ayesha Maqsood<sup>1</sup>, Dr. Arshad Mehmood Qamar<sup>2\*</sup>, Javaria Bashir<sup>3</sup>

### Abstract

This study was conducted to analyze test items by measuring quantitative characteristics (difficulty level and discrimination index). There were 25 MCQ type test items designed by using first two chapters of Chemistry Class 10<sup>th</sup>. This test was used as a pretest to examine the “Effect of e-module on the Academic Achievement of Chemistry Students at Secondary Level”. Researcher selected 150 students through multistage random sampling from district Gujrat for item analysis. A total of 150 students were used to measure the difficulty index. Researcher arranged the students in descending order and selected 33% high achievers and 33% low achiever by taking a total of 100 students (50 from HEs and 50 from LEs). In this way, 50 top students from high achievers and 50 bottom students from low achievers were selected for Item discrimination. Data were analyzed through descriptive statistics. Findings revealed that 18 items out of 25 were valid. Seven test items (2, 12, 15, 16, 19, 20 and 25) were rejected from the analysis of difficulty index. Out of these seven, difficulty index value for six items was very high (>70). Only one item (19) was difficult and rejected due to the very low value of the difficulty index (< 29). 18 items were retained after necessary revision. Discrimination index of 25 items showed that 18 items were discriminant after the revision of marginal items. It was concluded that 18 items out of 25 were valid based on the Difficulty index and Discrimination index used as a pretest. It was recommended that item analysis is a useful method to ensure the validity and reliability of test items.

**Keywords:** Item Analysis, Difficulty Index, Item Discrimination, Multiple Choice Questions (MCQs)

### 1. Introduction

Evaluation is an integral part of the teaching and learning process. The important thing is to consider the validity, reliability and objectivity of tools during evaluation. MCQs are considered authentic, reliable and objective tools to evaluate the achievement of students. It is considered a preferred tool in different countries to select students for any specific course. These MCQs can assess high-order thinking skills of students if they are constructed properly. They not only assess the recalling of facts but also high cognitive domains according to Bloom's taxonomy like synthesis, evaluation and application of knowledge (Kolte, 2015).

Multiple Choice Questions (MCQs) are considered as one of the best tools for assessment. Well-trained teachers are needed to prepare MCQs and it is a time-consuming process. MCQs should not only assess recall of facts but also high-order thinking and cognitive skills like evaluation, synthesis and application. MCQ type tests are of two types. Norm-referenced tests (NRT) and Criterion-Referenced tests (CRT). NRTs are used to rank the individual as compared to others based on his performance. CRTs are used to check the mastery of an individual against any criteria.

There are two parts of MCQ type test. One is stem which is a statement that shows a question or problem. Another is a set of alternatives in which one is the best answer called key and the remaining are distractors. The quality of MCQ depends on good distractors as these can differentiate informed and not informed students (Burud, Nagandla & Agarwal, 2019).

Constructing good MCQs is a challenging, complex and time-consuming process. Research has shown that item analysis is a quick and time-saving method to revise poorly constructed MCQ to a better one as compared to replace it with a new one. If we replace any item with a new one then it will have a new problem. Item analysis is the simplest way to check the reliability and validity of test items which helps to assess the quality of the test. Multiple Choice questions are the most commonly used test items that are easy to prepare. MCQs are practical to administer, and objective as compared to other type of tests that has subjectivity and bias. These are easy and consistent to score and grade and reduce the burden on teachers. They are easily designed as there are different computer programs available to design MCQ-type test items. These tests can also be graded easily and quickly with the help of computer software such as OMR sheets (Optical Mark Reader) as compared to constructed response questions (Toksöz & Ertunç, 2017).

Item analysis is a useful, simple and reliable method to ensure the quality of test items as they are valid and reliable or not. With the help of this analysis, teachers can improve the test to get better results. It helps to identify ambiguous test items in a test and replace them with a better one. It also helps the teachers to get feedback from students about areas that need reinforcement and more emphasis or change in pedagogy. It helps to improve the quality of MCQ test items. There are different concepts used in item analysis to ensure the validity and reliability of test items that are difficulty index and item discrimination (Shenoy, Ravi & Chandu, 2023).

<sup>1</sup> PhD Scholar, Department of Science Education, Allama Iqbal Open University, Islamabad

<sup>2\*</sup> Lecturer Science Education Department, Allama Iqbal Open University Islamabad, Pakistan. [arshad.mehmood@aiou.edu.pk](mailto:arshad.mehmood@aiou.edu.pk)

<sup>3</sup> PhD Education Scholar, University of Gujrat, Gujrat Pakistan

Difficulty index tells to what extent items were easy or difficult and the ratio of correct responses out of total response. It is a measure of the easiness of any item in a test. It is also known as p-value or ease index and its range is between 0-100 percent. A higher percentage of DI means the item was easier. If an item in the test is too easy or too difficult then it fails to distinguish between low achievers and high achievers among students. Item difficulty was calculated by using Kelley's formula by dividing correct responses to the total number of students involved in the test. The average level of DI should be 30%-70%.

Item Discrimination distinguishes high achievers from low achievers. It shows the percentage of students who are performing well as compared to others. It is a type of point-biserial correlation. Its value ranges from -1 to +1 when more students from high achievers respond correctly then it is +1 and it is -1 when more students from low achievers respond correctly. Item has a value near +1 means it is perfectly discriminated between high and low achievers. Researcher is motivated to conduct the item analysis for MCQ test items to assess their quality.

### 1.1. Objectives of the Study

Objectives of this study were:

1. To check the difficulty level of MCQ type test of chemistry at secondary level.
2. To check the Item discrimination of MCQ type test of chemistry at secondary level.

### 1.2. Research Questions

Research questions of this study were:

1. What is the difficulty level of each MCQ item of chemistry at secondary level?
2. What is the Item discrimination of each MCQ item of chemistry at secondary level?

## 2. Literature Review

Evaluation is an integral part of teaching and learning that helps to measure understanding of students and to ensure that learning objectives are achieved or not. It is useful for teachers to measure the progress of students and the overall performance of teachers. Test conduction is a systematic process that requires different steps like planning, construction of test items according to content, pilot testing, administration of test and evaluation to ensure the quality of test items. Evaluation is a way to diagnose the strengths and weaknesses of learners and decide about actions accordingly to improve their understanding and learning. It is not possible to learn about the needs of students without evaluation (Sukendra, 2023).

It is rightly said that "Evaluation drives the curriculum". We can say that if we want to improve the quality of teaching and learning process then assessment can be a starting point. Evaluation is a continuous process that needs to be modified. Different types of tests are used in the evaluation of students. One is Multiple Choice Questions (MCQs) which are objective type tests.

Multiple choice Questions are a widely used type of assessment to test different domains. It has two parts. One is a stem that consists of a statement, problem or question. Another component is four options that consist of a correct response (key) and the remaining options are distractors. MCQ tests are easy to administer with a large number of students and can assess broad areas of knowledge. The important thing is to construct good quality MCQ test items, is not an easy job. According to researchers, pre and post-validation of MCQ test items is compulsory to ensure its quality. Pre-validation can be done with the help of expert opinion. Post-validation is done through item analysis (Asrifan & Raskova Octoberberlina, 2021).

Item analysis is a method to examine students' responses toward each item. It ensures the quality of any whole test. It is beneficial when researchers want to improve the test by removing misleading and ambiguous test items. Teachers can identify a specific content area that needs to be emphasized and their test-building skills become enhanced. It is critical to analyze the test items if we want to achieve educational objectives.

Item analysis is conducted through item difficulty index (P), item discrimination (D) and distractor efficiency (DE) analysis of test items. Item difficulty is the proportion to which students responded correctly to any test item. Its acceptable range is 30% to 70%. It is measured by dividing correct responses by total number of responses. Item discrimination is the ability of any test item that clarify how well any question can discriminate a good student from a poor student. Its ideal value is equal or greater than 0.2. Distractor efficiency helps to clarify whether any test item is functional or nonfunctional. It will be functional if more than 5% of students respond it correctly (Shenoy et al., 2023).

## 3. Methodology

This study was a cross-sectional survey design. Researcher selected 150 students through random sampling and collected data with the help of test used in research to find the effect of e-module in chemistry at secondary level. Data was analyzed to do item analysis of test items.

### 3.1. Sample and Sampling Technique

Researcher selected 150 students from district Gujrat through a multistage random sampling technique. As target population for this study was all female regular science students of public secondary schools of district Gujrat. There were total 179 public schools in district Gujrat who studied science subjects at secondary level. There were

total 9904 female regular science students in these schools of district Gujrat who had been enrolled in matric part II during session 2022 in Gujranwala board (BISE GRW Gazette, 2022).

For multistage random sampling, researcher created a list of 179 schools with unique ID. 30 schools were selected randomly from three tehsils of district Gujrat. A list of all sections in these 30 schools was created and then researcher selected one section from each school randomly. Finally, 5 students from each school were selected through random sampling. In this way, 150 students were obtained as a sample.

### 3.2. Instrument of the study

Researcher used the MCQ type test as an instrument. This instrument consisted of 25 test items and it was developed from the first two chapters of Chemistry class X, Punjab Text Book Board Lahore. All MCQ items have four options from which one answer was correct, called the key, and the other three options are very close to the key. These are called distractors. This instrument was used to collect data.

### 3.3. Assumptions of the Study

As students chose for the test was of class X so it was assumed that all students had necessary background knowledge of the content as they had taught these chapters. These students could understand the items easily.

### 3.4. Statistical Analysis and Results

Test of 150 students was marked according to the key and data was analyzed with the help of Excel software. By using descriptive statistics, frequency and percentage were calculated. Difficulty index was calculated by using Kelley's formula. Researcher arranged the marks of students in descending order and selected 33% high achievers and 33% low achiever students by taking a total of 100 students (50 from HEs and 50 from LEs). In this way, 50 top students from high achievers and 50 bottom students from low achievers were selected for Item discrimination.

### 3.5. Difficulty Index

According to item difficulty analysis, items 1, 3, 7, 8, 10, 23 and 24 were average in difficulty so they were selected and added to the test. Items 4, 6 and 9 were difficult so they were added after revision. Items no. 5, 11, 13, 14, 17, 18, 21 and 22 were easy so they were added in the test. Items no. 2, 12, 15, 16, 20 and 25 were easy so they were rejected. Item no. 19 was so difficult so it was also rejected. Result of item difficulty is shown in the table.

**Table 1: Difficulty Index (P value) of Items in Chemistry test**

| Item No.                 | N <sub>CA</sub> | P   | Item No.  | N <sub>CA</sub> | P    |
|--------------------------|-----------------|-----|-----------|-----------------|------|
| Item # 1                 | 71              | 47% | Item # 14 | 94              | 62%  |
| Item # 2                 | 144             | 96% | Item # 15 | 134             | 89%  |
| Item # 3                 | 69              | 46% | Item # 16 | 126             | 84%  |
| Item # 4                 | 54              | 36% | Item # 17 | 104             | 69%  |
| Item # 5                 | 103             | 67% | Item # 18 | 104             | 69%  |
| Item # 6                 | 56              | 37% | Item # 19 | 29              | 19%  |
| Item # 7                 | 77              | 51% | Item # 20 | 121             | 81%  |
| Item # 8                 | 81              | 54% | Item # 21 | 103             | 68%  |
| Item # 9                 | 57              | 38% | Item # 22 | 104             | 69%  |
| Item # 10                | 86              | 57% | Item # 23 | 63              | 42%  |
| Item # 11                | 105             | 70% | Item # 24 | 83              | 55%  |
| Item # 12                | 137             | 91% | Item # 25 | 123             | 82%  |
| Item # 13                | 102             | 68% |           |                 |      |
| Mean                     |                 |     |           |                 | 93.2 |
| Average difficulty level |                 |     |           | 61.88%          |      |

**Table 2: Evaluation and Interpretation of Difficulty Index**

| Sr. # | DI Range | N/25 | Percentage | Item Evaluation | Interpretation    |
|-------|----------|------|------------|-----------------|-------------------|
| 1.    | ≤ 30     | 1    | 4%         | Too difficult   | Rejected          |
| 2.    | 31-40    | 3    | 12%        | Difficult       | Revised and added |
| 3.    | 41-60    | 7    | 28%        | Average         | Added             |
| 4.    | 61-80    | 8    | 32%        | Easy            | Added             |
| 5.    | 81-100   | 6    | 24%        | Too easy        | Rejected          |

Based on item difficulty, only 4% of items were too difficult and they were rejected. Difficult items were 12% so they were revised and added to the test. 28% of items were average in difficulty and they were sustained. Similarly easy items were 32% and also added to the test. 24% of items were too easy and they were rejected. Results were more obvious graphically.

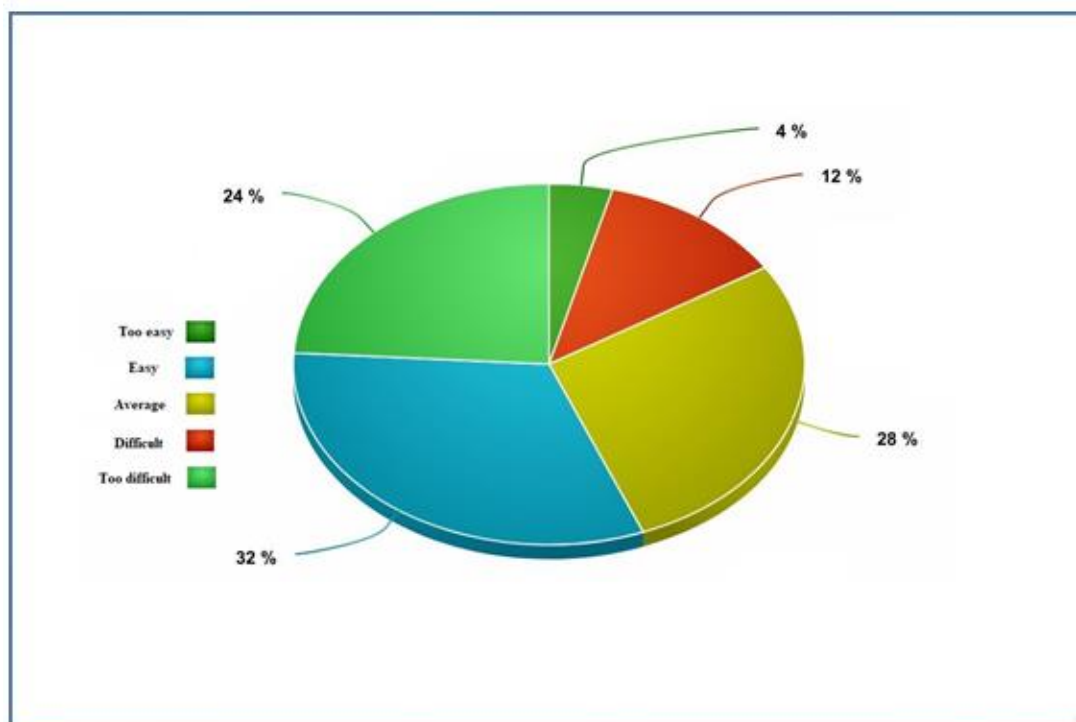


Figure 1: Percentage of difficulty Index

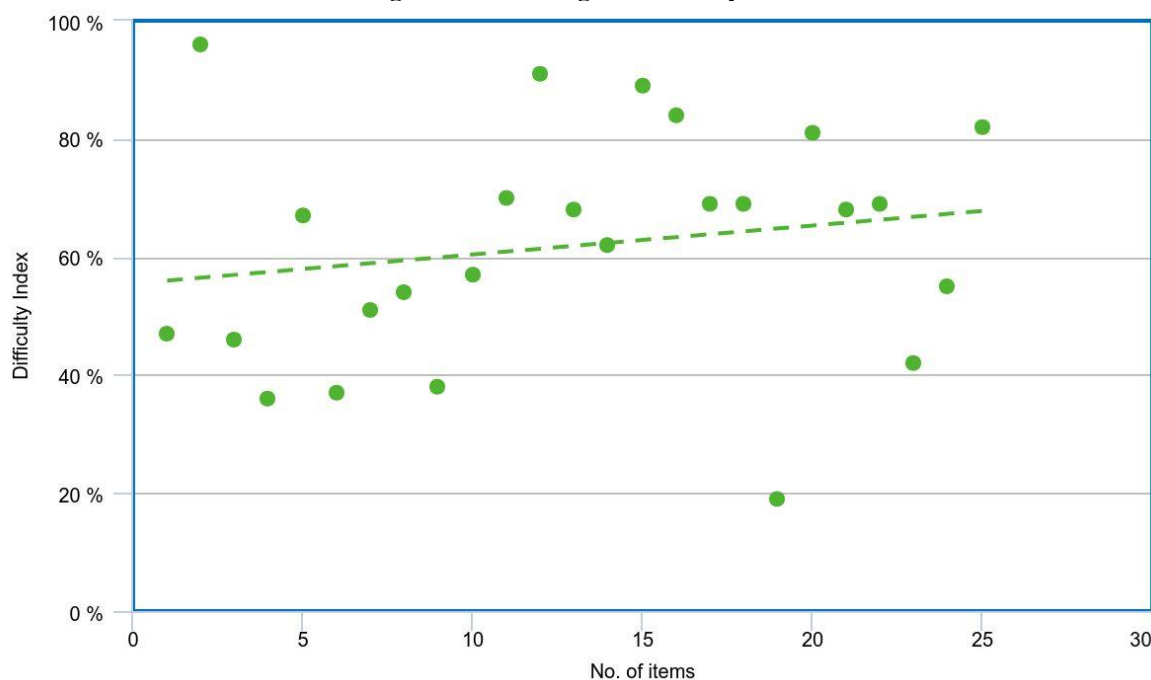


Figure 2: Scatter Plot of difficulty Index

### 3.6. Item Discrimination

Researcher calculated the item discrimination by arranging the scores in descending order. Divided respondents into three groups, high achievers, middle and low achievers by selecting the top 33% and bottom 33% to discard 34%, middle respondents. Researcher counted correct responses from high achievers and from low achiever groups, subtracted these and divided them by the total no. of one group. Item discrimination was calculated by Kelly’s method. Results are shown in Table 3.

Table 3: Item Discrimination of Items of Chemistry test

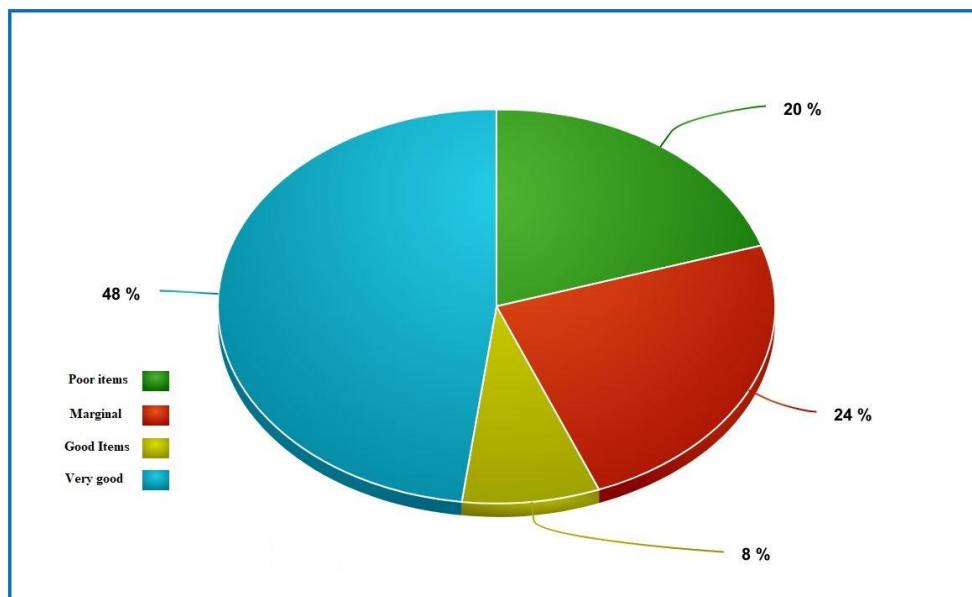
| Item no. | NH | NL | N  | $D = \frac{NH - NL}{N}$ | Decision  | Interpretation |
|----------|----|----|----|-------------------------|-----------|----------------|
| Item # 1 | 00 | 9  | 50 | 0.78                    | Very good | Accepted       |
| Item # 2 | 50 | 44 | 50 | 0.24                    | Marginal  | Revised        |
| Item # 3 | 48 | 36 | 50 | 0.24                    | Marginal  | Accepted       |

| Item no.  | NH | NL | N  | $D = \frac{NH - NL}{N}$ | Decision  | Interpretation |
|-----------|----|----|----|-------------------------|-----------|----------------|
| Item # 4  | 22 | 10 | 50 | 0.24                    | Marginal  | Accepted       |
| Item # 5  | 44 | 30 | 50 | 0.28                    | Marginal  | Accepted       |
| Item # 6  | 31 | 4  | 50 | 0.54                    | Very good | Accepted       |
| Item # 7  | 42 | 6  | 50 | 0.72                    | Very good | Accepted       |
| Item # 8  | 36 | 20 | 50 | 0.32                    | Good      | Accepted       |
| Item # 9  | 23 | 18 | 50 | 0.1                     | Poor      | Revised        |
| Item # 10 | 43 | 16 | 50 | 0.54                    | Very good | Accepted       |
| Item # 11 | 45 | 24 | 50 | 0.42                    | Very good | Accepted       |
| Item # 12 | 49 | 44 | 50 | 0.1                     | Poor      | Rejected       |
| Item # 13 | 42 | 25 | 50 | 0.34                    | Good      | Accepted       |
| Item # 14 | 46 | 20 | 50 | 0.52                    | Very good | Accepted       |
| Item # 15 | 50 | 38 | 50 | 0.24                    | Marginal  | Revised        |
| Item # 16 | 46 | 41 | 50 | 0.1                     | Poor      | Rejected       |
| Item # 17 | 44 | 23 | 50 | 0.42                    | Very good | Accepted       |
| Item # 18 | 47 | 25 | 50 | 0.44                    | Very good | Accepted       |
| Item # 19 | 13 | 4  | 50 | 0.18                    | Poor      | Rejected       |
| Item # 20 | 44 | 34 | 50 | 0.20                    | Marginal  | Revised        |
| Item # 21 | 44 | 21 | 50 | 0.46                    | Very good | Accepted       |
| Item # 22 | 47 | 25 | 50 | 0.44                    | Very good | Accepted       |
| Item # 23 | 30 | 10 | 50 | 0.40                    | Very good | Accepted       |
| Item # 24 | 44 | 14 | 50 | 0.60                    | Very good | Accepted       |
| Item # 25 | 42 | 37 | 50 | 0.10                    | Poor      | Rejected       |

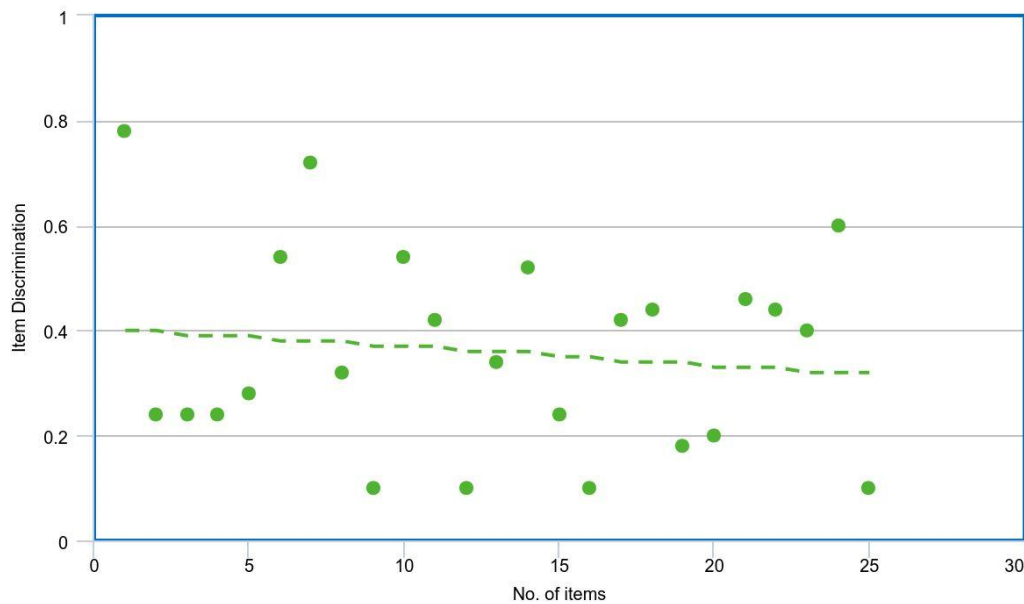
**Table 4: Evaluation and Interpretation of Item Discrimination**

| Sr. # | Index Range | N/25 | Percentage | Evaluation | Interpretation    |
|-------|-------------|------|------------|------------|-------------------|
| 1.    | $\geq 0.19$ | 5    | 20%        | Poor item  | Rejected          |
| 2.    | 0.20-0.29   | 6    | 24%        | Marginal   | Revised and Added |
| 3.    | 0.30-0.39   | 2    | 8%         | Good item  | Added             |
| 4.    | $\leq 0.40$ | 12   | 48%        | Very Good  | Added             |

According to item discrimination analysis, items 9, 12, 16, 19, 25 were poor. They were rejected except item no. 9 as it was accepted according to item difficulty so it was revised and added to the test. Items 2, 3, 4, 5, 15 and 20 were marginal so they were also added to the test. Items 8 and 13 were good and included in the test. Items 1, 6, 7, 10, 11, 14, 17, 18, 21, 22, 23 and 24 were very good as their value of item discrimination was greater or equal to 0.40. These were also added to the test. Item discrimination analysis showed that poor items were 20%, marginal items were 24%, and good items were 8% and 48% items had very good discrimination index value. Results were more obvious graphically.



**Figure 3: Percentage of Item Discrimination**



**Figure 4: Scatter Plot of Item Discrimination**

#### 4. Findings and Discussion

Findings showed that item analysis is a powerful process to ensure the quality of MCQ test items. Based on item difficulty and item discrimination, seven items 2, 12, 15, 16, 19, 20 and 25 were rejected and excluded from the test. The remaining 18 items were revised according to need and added to the test. The findings of the current study are consistent with previous studies. A study was conducted by Kusumawati & Hadi in 2018, titled “An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school.” The findings of their study showed that according to item discrimination, 8.57% of items were of low category, 31.43% of items were of medium category, 5.71% of items had a high value of discrimination index and 2.86% of items were of very good discrimination index. They found that all distractors were functional. Another study was conducted by Obon & Rey (2019, December), titled “Analysis of Multiple-Choice Questions (MCQs): Item and Test Statistics from the 2nd Year Nursing Qualifying Exam in a University in Cavite, Philippines.” They concluded that out of 194 items, item difficulty of 115 items (59.3%) were right and 79 items (40.7%) were difficult. In the case of DI, 17 items (8.8%) were very good to discriminate between high and low achievers. Researchers recommended that item analysis is a good way to assess the quality of test items.

#### 5. Conclusion

This study concluded that item analysis is an important method to ensure the validity and reliability of items in any exam. It provides a holistic way to analyze items. Decisions to revise test items can be made based on the difficulty index and discrimination index. Revision of test items based on item analysis helps to strengthen and update the item bank of MCQs. As in this study, 18 items out of 25 were selected after the necessary revision. Examiners can select suitable difficulty levels for MCQs according to the need and purpose of the assessment. Results of item analysis can help faculty as well as students to improve learning outcomes. Teachers can assess any modification in their teaching strategies.

#### 6. Recommendations

It is recommended that institutions should conduct item analysis of their MCQ type test over ten years this will provide a powerful item bank for future use. The government should conduct in-service training for teachers regarding item analysis so they can be able to construct valid and reliable test items. For further research, it is recommended that item distractors may also be found. Item analysis can be carried out for test items of other subjects. Moreover, interviews with the students should be conducted about the difficulty of test items, discrimination power and distractor analysis. This will provide qualitative data to strengthen the results of quantitative analysis and provide a broader picture.

#### References

- Asrifan, A., & Raskova Octaberlina, L. (2021). An Item Analysis Of English Summative Test in EFL Classroom (A case study at Elementary School in Indonesia). *ScienceOpen Preprints*.
- Burud, I., Nagandla, K., & Agarwal, P. (2019). Impact of distractors in item analysis of multiple choice questions. *International Journal of Research in Medical Sciences*, 7(4), 1136-1139.

- Kolte, V. (2015). Item analysis of multiple choice questions in physiology examination. *Indian Journal of Basic and Applied Medical Research*, 4(4), 320-326.
- Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REiD (Research and Evaluation in Education)*, 4(1), 70-78.
- Obon, A. M., & Rey, K. A. M. (2019, December). Analysis of Multiple-Choice Questions (MCQs): Item and test statistics from the 2nd year nursing qualifying exam in a University in Cavite, Philippines. In *Abstract Proceedings International Scholars Conference (Vol. 7, No. 1, pp. 499-511)*.
- Qamar, A. M., Kanwal, W., & Nadeem, H. A. (2022). Item analysis for test to examine the effect of e-module on the academic performance of 7th class science students in Islamabad. *Jahan-e-Tahqeeq*, 5(2), 188-196.
- Qamar, A. M., Kanwal, W., & Perveen, R. (2022). Item Analysis of Tool used for Examining the Effectiveness of E-modules for Academic Performance of 7th Grade Students. *Archives of Educational Studies (ARES)*, 2(2), 205-222.
- Shenoy, V., Ravi, P., & Chandy, D. (2023). A cross-sectional study on Item Analysis of Prevalidated and nonvalidated anatomy multiple-choice questions. *National Journal of Clinical Anatomy*, 12(2), 94-97.
- Sukendra, I. (2023, November). ITEM ANALYSIS OF ENGLISH TEACHER-MADE TEST. In *UNNES-TEFLIN National Conference (Vol. 5, pp. 390-401)*.
- Toksöz, S., & Ertunç, A. (2017). Item analysis of a multiple-choice exam. *Advances in Language and Literary Studies*, 8(6), 141-146.