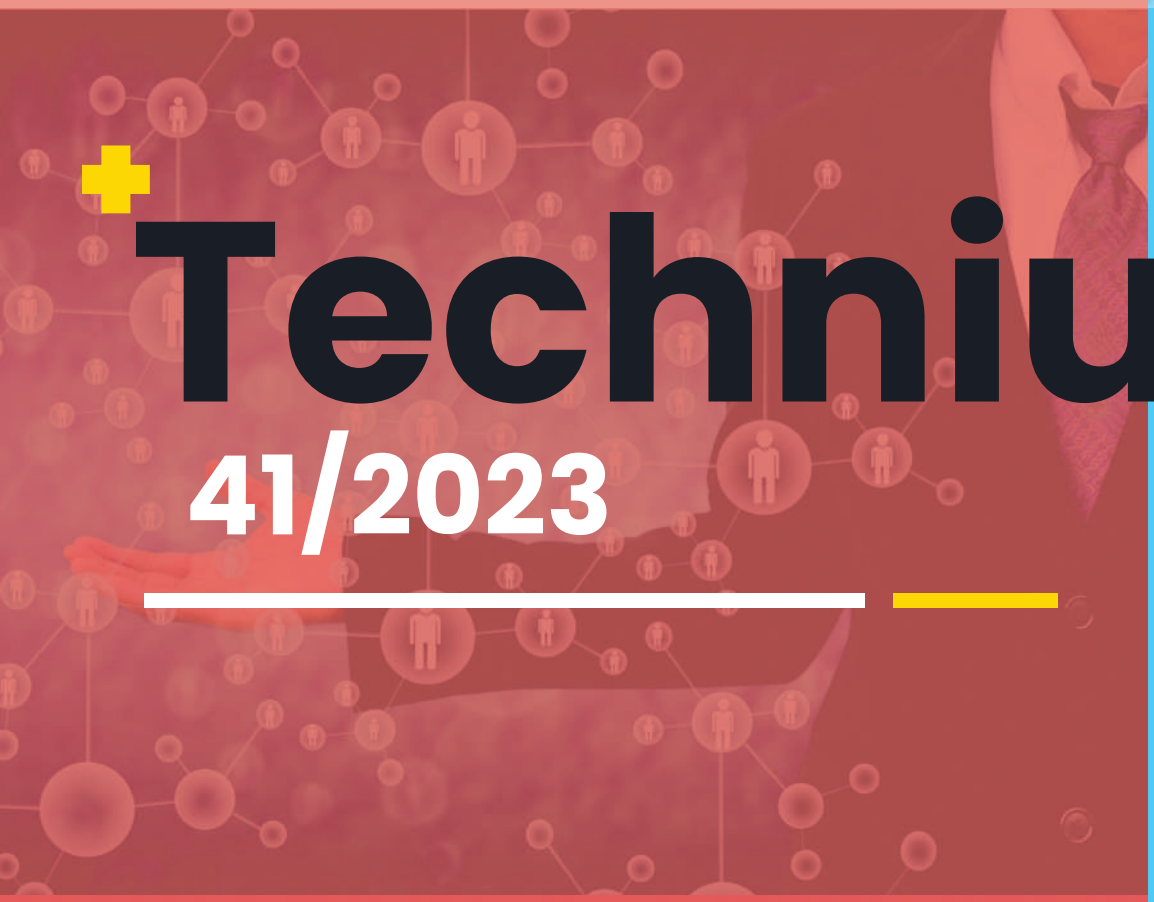
TECHNIUM
SOCIAL SCIENCES JOURNAL



Technium.

41/2023

2023

A new decade for social changes

Technium Social Sciences

Powered by PLUS COMMUNICATION
International Communication & PR

TECHNIUM SOCIAL SCIENCES JOURNAL

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

Item Analysis of Multiple Choice Questions (MCQs) for Dangerous Goods Courses in Air Transportation Management Department

Wiwid Suryono, BB. Harianto

Politeknik Penerbangan Surabaya, Surabaya, Indonesia

Bambangfarzardy@gmail.com

Abstract. Analysis of question items is an assessment of the questions that are evaluated to measure quality. Analysis of question items has two ways, namely quantitative and qualitative question item analysis. The goal to be achieved is to find out the quality of the multiple-choice form of Final Semester Exam questions and also to detect sound and wrong questions. The method used in this study is the quantitative descriptive method. Data collection uses documentation, interviews, and observations. The Excel program uses data analysis to detect the quality of good and bad questions. The results of the quantitative analysis of the reliability of multiple-choice questions of 0.76 mean reliable questions. Difficulty index: 3 medium questions (6%), 47 easy questions (76%), and 9 very easy questions (18%). Discrimination index: 8 questions are very low (16%), 22 questions are low (44%), 9 questions are medium (18%), and t10 questions are high (20%). Distractor efficiency: 125 items of malfunctioning questions (83.33%) and 25 items of functioning questions (16.67%).

Keywords. multiple choice question (MCQs), dangerous goods, air transportation management department, vocational education

1. Introduction

Education is essential in educating the nation into a generation with character and a competitive spirit. (Kristanto et al., 2020) Universities play an important role in educating the nation's youth. Following Minister of Education Issue Number 44 of 2015 concerning National Standards for Higher Education, the process of managing higher education must be standardized. To achieve higher education performance, the accreditation of study programs and universities must be ranked by Minister of Education Issue Number 32 of 2016, Accreditation of Study Programs and Universities. The purpose of this research is to analyze the performance measurement of universities comprehensively. The government established formal educational institutions, one of which is official Education, to realize the goal of educating the nation. According to Law No. 20 of 2003 concerning the National Education System, it is explained that "Vocational education is secondary education that prepares students especially to work in certain fields." According to Law No. 20 of 2003 concerning the National

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

Education System, "Vocational education is secondary education that prepares students specifically for work in certain fields."

Surabaya Aviation Polytechnic is a vocational education institution that prepares students to work in specific fields to create quality human resources. As an Official University, Surabaya Aviation Polytechnic

Surabaya Aviation Polytechnic, from now on referred to as Poltekbang Surabaya, a state university within the Ministry of Transportation, reports to and is accountable to the Director of the Transportation Human Resources Development Agency. It organizes vocational education programs, research, and community service in aviation (PM 99 TAHUN 2021). In addition to managing vocational Education, the Surabaya Aviation Polytechnic organizes Education and training in transportation. Vocational Education is a higher education diploma program that prepares students for specific applied skill jobs up to the level of applied doctoral programs.

One of them is the Air Transportation Management study program, a three-applied diploma program with multi-disciplines, preparing air transportation professionals and getting to know more about airways, supporting devices, and aviation facilities. In the Diploma three study program of Air Transportation Management, studying the management of transport at airports, either passengers or goods, must be based on international regulations, one of which IATA is the global air transport association. The transportation of dangerous goods is regulated in the provisions of Annex 18, "The Safe Transport of Dangerous Goods by Air," which is a rule that governs the signs, how to pack, and transport dangerous cargo, which Surabaya Poltekbang cadets study as a course of Dangerous Goods Regulation (DGR).

Lecturers are professional educators and scientists whose primary responsibility it is to transform, develop, and disseminate science, technology, and art through education, research, and community service. (Sijabat et al., 2021) Lecturers are professional educators and scientists whose primary responsibility is to transform, develop, and disseminate science, technology, and art through education, research, and community service., who have the status of Civil Servants of the Surabaya Poltekbang who have available positions of Lecturers who work full-time. Lecturers are one of the success factors in producing competent graduates. The task of a lecturer is not only to provide material in front of the class, but a lecturer needs to know about planning, implementing, and evaluating learning. Use teacher observation instruments as part of a new teacher evaluation system that states and districts are considering and implementing. They argue that if teacher observation instruments are to help teachers improve their teaching practices, they should be subject-specific, involve content experts in the observation process, and provide teachers with accurate and valuable information. They discuss the instrument itself, the appraiser, and the design of the system, as well as the timing and feedback from the observations. (Hill & Grossm An, 2013) Evaluation activities that are considered successful if they can improve the process and results of learning scores, then evaluation activities must also pay attention to the quality of good tests, namely those that have good question items**".** Here are three major points: First, evaluations are conducted for a variety of reasons other than eliminating laziness, incompetence, and malpractice; second, organized efforts to provide human services (i.e., programs) can be evaluated; and, third, cooperative evaluations can serve to improve the program and the quality of life.(Posavac, 2016) The evaluation process ensures that the intended learning objectives meet.(Alsoufi et al., 2020)

Evaluation activities are the process of assessing something. If the assessment tries to find what the student has become or has achieved, then the evaluation attempts to do the same for the course, learning experience, or episode each time. Evaluation is an attempt to identify

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

and explain the effects (and effectiveness) of teaching. In such efforts, assessment is an indispensable component. Assessment, whether formal or informal, reveals to us the most important 'effects' – the changes in our student's knowledge and understanding, abilities, and attitudes. (Rowntree, 1987) There are several datasets taken from the Exam, aimed at evaluating the system under the same conditions as how a human being is assessed in school (Lai et al., 2017). To determine the value of something being evaluated, it is necessary to take measurements, and the form of that measurement is testing. This testing is what in Education is known as a test. The Final Exam is known as the Exam related to the teaching and learning process. (Sugianto, 2016)

According to (Nitko, 1996) in his book "Educational Assessment of Students," question item analysis is an evaluation activity carried out by teachers on the results of the implementation of a test to find out whether the questions (items) given to have a good bag cauldron. Analysis activities include collecting, summarizing, and using information from students' answers to make decisions about assessments.

Educators or lecturers need to know about evaluation because the evaluation process is an essential means of measuring cognitive, affective, and psychomotor abilities. Learning, in general, can be categorized into three domains: active, affective, and psychomotor Cogn. (Hoque, 2016) Meaningful learning underlies the constructive integration of thoughts, feelings, and actions that lead to human empowerment, commitment, and responsibility. In other words, for learning to be meaningful, it must integrate cognitive, affective, and psychomotor elements. (Enneking et al., 2019) In general, in the implementation of the evaluation process, lecturers use tests as a means to assess the cognitive learning outcomes of students related to mastery of specific subject matter. The theory of cognitive load is based on the assumption that the working memory capacity of the human being is limited. To acquire new skills, the learner must process further information within the limits of working memory so that it can be integrated with relevant previous knowledge and stored as ema in long-term memory. (Haji et al., 2015)

The test serves as accurate information regarding the learning outcomes of students during the learning process. Testing has a learning assessment role (i.e., achievement testing). (Richards, 2013)There are two functions of the Exam, namely: (1) As an instrument to measure the progress or progress that students have achieved after they have undergone the Teaching and Learning Process within a certain period, and (2) as an instrument to measure the success of the Teaching and Learning Process. Through the test, the teacher will be able to determine the extent to which the material that has been produced can be reached by students. (Sugianto, 2016)

Studies that used knowledge-based learning outcomes measures to test student learning outcomes for declarative tasks showed higher results. (Merchant et al., 2014) In addition, the test can also help lecturers in following up with students, especially students who have poor learning outcomes. Therefore, the test to be tested must be of high quality. Lecturers can conduct question analysis activities to determine the quality of the tests, including the disadvantages and advantages of each item. "The purpose of question analysis is to identify good, bad, and bad questions." So improving the quality of Education involves paying more attention to description and analysis. (Alexander, 2008) Question analysis is critical in enhancing the questions reused in the next test; it can also be used to eliminate misleading items. (Quaigrain & Arhin, 2017) So, if a question meets the forum's standards, it is classified as good; otherwise, we classify it as harmful. (Arora et al., 2015) With the analysis of the questions, information can be obtained about the spectacle of a question and the 'boxer k' to make improvements. (Amelia Suek, 2021)

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

According to (Tinambunan, 1988), in analyzing the question items, three essential things are needed, namely: usually concentrates three vital features: difficulty index, discriminating index, and the effectiveness of each alternative. Item analysis information can tell us if an item was very difficult or very easy, how well it determined between high and low scores on the test, and whether all the alternatives functioned as intended.

The point of the above opinion is that there are three critical things to carry out question item analysis activities: the difficulty index, discrimination index, and the effectiveness of alternative answers to each question. Three things are usually considered in item analysis: difficulty index, discrimination index, and efficacy of the distraction. (Toha, 2010) (Damayanti et al., 2018) Item analysis serves as information to find out the quality of the item, whether the thing is too easy or too complicated, and how well the difference between high and low scores on the test is, as well as to find out if all alternatives are working as intended**.** The second approach to determining quality is the application of statistical item analysis procedures to determine the characteristics of items and use those statistics to decide whether an item can be appropriately included in the determination of test taker performance (Malau-Aduli & Zimitat, 2012)**.** This is strengthened by having good quality if it follows the curriculum, meets the requirements of material, construction, and language aspects, has high validity, reliability, and discrimination index, has a moderate difficulty index, and can measure the achievement of student competencies**.** The question items that are considered feasible are valid questions, reliable, have excellent and good discrimination index, and have a moderate index of difficulty.(Saputra et al., 2018)

The success of dangerous goods course learning can be known by evaluating knowledge in the form of a Final Semester Examination (UAS). Poltekbang Surabaya makes the Final Semester Examination (UAS) an evaluation tool to find learning outcomes and assess the extent of students' mastery of the subject matter for one semester. Learning outcomes are indicators to measure the learning effects of learners(Sharma & Kumar, 2021)

Student learning outcomes are also influenced by the ability of lecturers to ask questions. The good and bad results of the Final Semester Examination (UAS) of students also depend on the quality of the questions each lecturer asks**.** The quality of the questions can be known by conducting question item analysis activities and testing the questions before implementing the Final Semester Examination (UAS). This is so that the results of the Final Semester Examination (UAS) can be maximized. Following the course's learning outcomes, lecturers can find students with high and low abilities. The Final Semester Examination (UAS) questions tested at Poltekbang Surabaya consist of two types: multiple choice and description. In multiple-choice questions, students can choose one of the four (4) alternative answers (a, b, c, d) that have been provided, while for descriptions, students are required to decipher the answers with their own words, and language styles are different from one another. Multiple choice questions (MCQs) answer choice items were used as scoring methods. Well-constructed multiple-choice questions are a very reliable and objective measure of learning. In health professions education, multiple-choice questions are used to assess students' essential knowledge acquisition, evaluate their capacity to apply knowledge, and measure their critical thinking skills when questions are asked in a case-based format (Shaikh et al., 2020)**.** Multiple choice questions are used because, in a short period, various course materials can be efficiently graded and accurately graded (Brown & Abdulnabi, 2017)

Based on this background, the author is interested in conducting research on the analysis of question items with the title "**Item Analysis of Multiple Choice Questions for Dangerous Goods Courses in the Air Transportation Management Department**."

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

2. Methodology

The types of methods used in this study are qualitative and quantitative. Quantitative analysis is carried out using the difficulty index of the question, the discrimination index of the question, and the efficiency of the distractor. The validity and reliability of the question are tested. Qualitative analysis is carried out to determine the validity of the question content. Question study is based on the rules of writing questions regarding material or content, construction, and language. According to (Sugiyono, 2006), "A population is a generalized area consisting of objects/subjects that have certain qualities and characteristics that the researcher sets out to study and then draw conclusions." The population in this study was the entire answer sheet for the Final Semester Examination (UAS) test in semester 3 of the dangerous goods study program of the three diploma study programs in air transportation management Batch 7A and 7B. Poltekbang Surabaya, which amounted to 48 sheets.

"The sample is part of the number and characteristics shared by that population." (Sugiyono, 2006) The sample in this study was all the answers to the Final Semester Examination (UAS) test of 3 courses, the three diploma study programs in air transport management Force 7A and 7B, totaling 48 sheets.

"Research variables are everything in the form of anything set by the researcher to be studied so that information is obtained about it, then conclusions are drawn" (Sugiyono, 2006)**.** The variables in this study are the quality of the final semester exam questions for the dangerous goods course of the three-year diploma study program in air transportation management Batch 7A and 7B for the 2021/2022 academic year, while the indicators are discrimination index, difficulty index, validity, reliability, and distractor efficiency.

The data collection methods used in this study were documentation, interviews, and observations. This method is used to obtain a set of student answer sheets, Final Semester Examination (UAS) questions, answer keys, as well as a list of names of cadets of the three air transport management diploma study programs of Batch 7A and 7B of the 2021/2022 academic year. Cadets are students who are registered at Poltekbang Surabaya in the Formation Training in vocational Education.

Data analysis in this study is carried out in2 ways: quantitative analysis of questions and qualitative data analysis. Data analysis is the process of systematically finding and compiling data obtained from research results so that it can be easily understood. The final exam questions for the semester of the dangerous goods course of the three-year diploma study program of Air Transportation Forces 7A and 7B are in the form of multiple choice and description.

To perform the analysis, with the following steps: 1) compile student answer sheets from the highest score to the lowest score, 2) take 27% of the answer sheet from the top, which is after this referred to as the upper group (higher group), and 27% of the answer sheet from below which is from now on referred to as the lower group. The remaining 46% is set aside, 3) to create a table to find out each learner's answers (right or wrong), both for the upper and lower groups. If the learner's answer is correct, it is marked with a 1 (one). On the contrary, if the learner's answer is wrong, it is marked with a 0 (zero).

Qualitative question item analysis is an analysis that is carried out before the question is used or tested. Aspects considered in the qualitative analysis include material, construction, language or culture, and answer keys. (Daniel, 2016)(Lampert et al., 2013) Quantitative question item analysis analyzes question items based on empirical data from items obtained from questions that have been tested. In quantitative analysis, there are two approaches: classical and modern.

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

Meanwhile, classical analysis of question items is a process of analysis through information from students' answers using classical test theory. Meanwhile, modern sticking is the analysis of question items using Item Response Theory (IRT) or question item answer theory. This theory is a theory that uses mathematical functions to connect the chances of answering correctly with the ability of students. (Daniel, 2016) (Lampert et al., 2013)

The level of difficulty is the percentage of students who correctly answered a question; it ranges from 0 to 1.0, with a higher score indicating that more students correctly answered the question. For five-level multiple-choice questions, the best difficulty index is 0.60. (Kaur et al., 2016)(Saputra et al., 2018)**.** The degree of difficulty of the question item is usually denoted by p. The greater the p-value means, the greater the participant who answered correctly. This means that the higher the difficulty level index calculated from the results, the easier the question item. The question difficulty index is the percentage of high- and low-achieving students who correctly answered the questions. It ranges from 0% to 100%. It is calculated using the formula DIF I or P = H + L 100/N, where H is the number of students who correctly answered the questions in the high-achieving group, L is the number of students who correctly answered the questions in the low-achieving group, and N is the total number of students in both groups (including those who did not answer).(Saputra et al., 2018) The interpretation of the difficulty level of the problem is summarized as follows (Karim et al., 2021) a) difficult, value-p ranges from 0.00 – 0.30, b) medium, p-value ranges 0.31 – 0.70, c) easy, p-value ranges 0.71 – 1. 00. 2) The discrimination index of a test item is its ability to separate good students from those who are not good (Arta et al., n.d.)(Saputra et al., 2018)**.** One item of this question aims to distinguish between students who can do the question material (in this case, it is indicated that they are learning) and students who are unable to do it (in this case, not / lack of learning). DI = 2 (H-L)/N, where the symbols H, L, and N represent the same value g as previously stated. (Kaur et al., 2016) discrimination index was measured using point-biserial correlation (PBC), which ranged from −1.0 to 1.0, with higher values indicating more discrimination. It describes the extent to which students who have high quiz scores answer items correctly, and those who have low quiz scores answer them incorrectly. PBC values of 0.20 and higher are within the acceptable range. (Kaur et al., 2016) By knowing the discrimination index, we can know the breadth with which a student can receive learning and an evaluation of the learning model implemented by an educator (Ertikanto et al., 2022)**.** The interpretation of the discrimination index of the question is summarized as follows: a) low, 0.00-0.19, b) medium, 0.19-0.29, c) good, 0.30-0.39, d) excellent 0.40-1.00. (Olutola, 2015) 3) Deceptive function. The construction of the question item consists of two parts: the main question and alternative answers (answer keys and deception)**.** The dissemination of answer choices is used as a basis for the study of questions. It is intended to know whether or not the available answers work. If the set of all strategies is reduced to two—confident/uncertain responses and random guesses—then the distractor selection frequency distribution pattern is sufficient to describe all possible quantitative combinations between these strategies. However, in actual test behavior, at least five strategies are allocated (confident knowledge, uncertain knowledge, partial knowledge, random guessing, cheating) so that a model sufficient to describe the entire variation of the test behavior has not yet been created**.**(Olutola, 2015) We examined the distribution of selected answers among respondents and the significance of the distractors with the highest percentage. We used various criteria to compare the results vertically (between values) and horizontally (between questions). Following this comparison, we observed a distribution pattern of student responses per class/question.(Olutola, 2015) Deceptive answers are said to work when many test takers choose deceptive answers or fewer test takers choose**.** The distractor is an important

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

component of an item and greatly impacts the test's total score. Student performance depends on how the distractor is designed. For this reason, Distractor efficiency (DE), which shows whether the distractor in the question is well chosen or fails to distract students from choosing the right answer, is critical. (Burud et al., 2019)**.** Distractor efficiency, the number of non-functional distractors (NFD) per item, and the number of items with NFD are all calculated. NFD is the option selected by 5% of students. A non-functioning distractor occurs when 5% of students choose the incorrect answer (NFD). A functional distractor (FD) is a distractor chosen by more than 5% of students (Mahjabeen et al., 2017). As the amount of NFD in an item decreases, the efficiency of the distractor increases from 0% to 100%. The average standard deviation is calculated for all three parameters (Diff I, DI, and DE). The percentage of items belonging to different Diff I and DI categories is also calculated.(Rehman et al., 2018)

3. Result and Discussion

The results of the odd semester analysis of energy conversion machine subjects for the 2012/2013 school year were analyzed for a total of 50 questions. The results of the difficulty analysis of the question found that:

Table 1. Recapitulation of Difficulty Levels

Difficulty Level	Interpretation	Problem Number	Sum	Percentage
0.39	Medium	39	1	2 %
0.50	Medium	42	1	2 %
0.61	Medium	47	1	2 %
0.71	Easy	11,49	2	4 %
0.75	Easy	23	1	2 %
0.79	Easy	30,40,41,43,50	5	10 %
0.82	Easy	28,29,34,44	4	8 %
0.86	Easy	15,18,20,33	4	8 %
0.89	Easy	10,13,35,36	4	8 %
0.93	Easy	4,6,9,14,17,21,27,31,38,46	10	20 %
0.96	Easy	2,12,16,19,26,37,45,48	8	16 %
1.00	Very easy	1,3,5,7,8,22,24,25,32	9	18 %
TOTAL			50	100%

The analysis results in table 1 show a recapitulation of the problem's difficulty level. Regarding difficulty level, it is revealed that three questions are classified as moderate, 38 questions are easy, and nine are straightforward. It can be known that three questions with sensible interpretation have a percentage of 6%, 47 questions with specific interpretation have a percentage of 76%, and nine questions with straightforward interpretation have a percentage of 18%. In terms of difficulty level, the questions are analyzed into medium, easy, and straightforward questions. The data is then visualized in the form of a histogram, as shown in the following figure:

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

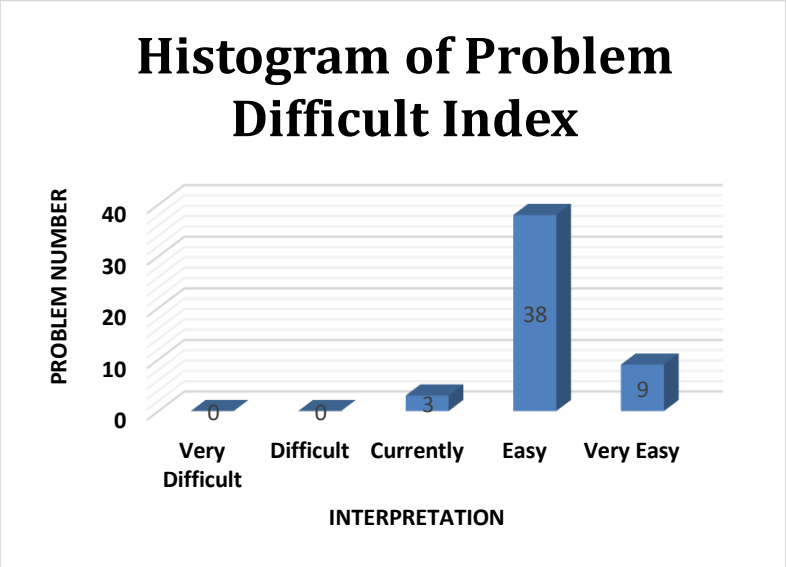


Figure 1. Histogram of Difficult Index

Figure 1 shows that the problem's difficulty index is very difficult and difficult is 0, the medium is 3, the easy is 38, and the very easy is 9.

Table 2. Recapitulation of Discrimination Index

Difficulty index	Interpretation	Problem Number	Sum	Presentase
0,00	Very Low	1,3,5,7,8,22,24,25,32	8	16 %
0.07	Low	2,10,12,13,16,19,26,35,37,45,48	11	22%
0,14	Low	4,6,9,14,17,18,21,27,31,38,46	11	22%
0,21	Enough	34,36	2	4%
0,29	Enough	15,20,22	3	6%
0,36	Enough	28,29,39,44	4	8%
0,43	High	11,30,40,41,43,50	6	12%
0,50	High	23	1	2%
0,57	High	49	1	2%
0,64	High	47	1	2%
0,70	High	42	1	2%
TOTAL			50	100%

The analysis results in table 2 show a recapitulation of the discrimination index of the problem. Regarding the level of discrimination index, it was revealed that eight questions are classified as very low, 22 questions are low, nine questions are enough, and ten questions are high. Eight questions with shallow interpretation have a percentage of 16%, 22 with low interpretation have a percentage of 44%, 9 with sufficient interpretation have a percentage of 18%, and 10 with high interpretation have a percentage of 20%. In terms of the level of discrimination index of the questions, the question items are analyzed into very low, low, medium, and high question items. The data is then visualized in the form of a histogram as shown in the following figure:

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

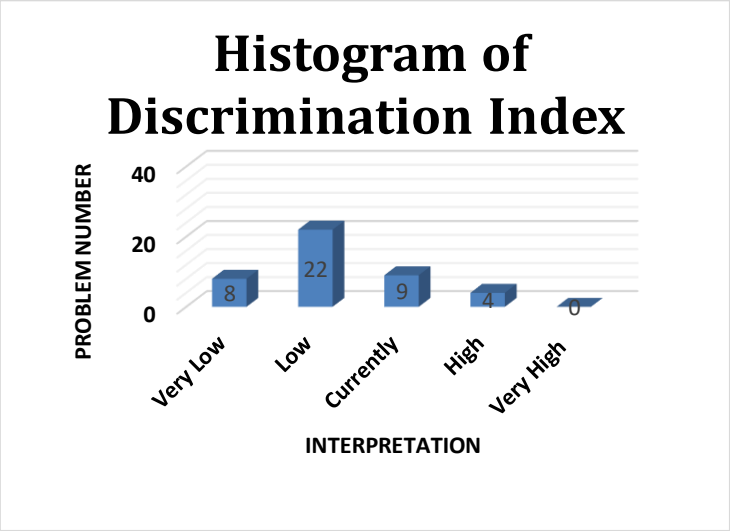


Figure 2. Histogram of Discrimination Power

Figure 2 explains that the histogram of the very low level of discrimination index of the question is 8, the low is 22, the medium is 9, the high is 4, and the very high is 0.

The effectiveness/function of each distractor is calculated using the number of testees who choose the option divided by the number of all testees multiplied by 100%, and then the interpretation is carried out. Odd semester exam questions of Dangerous Goods. Each question has four options consisting of one answer option and three deception options (distractor), meaning that there are 150 distractors on the n test questions. Based on the analysis of the calculation of the number of testees who choose *options* / alternative answers to the test questions on the summative test, information can be obtained as stated in the following this table :

Table 3. Recapitulation of Distractor efficiency

No	Answer Key	Distractor Function	
		Good	Bad
1	B	-	A, C, D
2	C	-	A, B, D
3	D	-	A, B, C
4	B	-	A, C, D
5	A	-	B, C, D
6	C	-	A, B, D
7	B	-	A, C, D
8	B	-	A, C, D
9	C	-	A, B, D
10	B	-	A, C, D
11	D	A, B	C
12	D	-	A, B, C
13	C	B	A, D
14	C	-	A, B, D
15	C	-	A, B, D

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

16	B	-	A, C, D
17	A	-	B, C, D
18	B	A	C, D
19	B	-	A, C, D
20	C	-	A, B, D
21	C	-	A, B, D
22	D	-	A, B, C
23	B	-	A, C, D
24	A	-	B, C, D
25	C	-	A, B, D
26	B	-	A, C, D
27	B	C	A, D
28	C	B	A, D
29	C	B, A	D
30	D	A	B, C
31	A	C	B, D
32	C	-	A, B, D
33	C	-	A, B, D
34	D	-	A, B, C
35	D	-	A, B, C
36	D	-	A, B, C
37	A	-	B, C, D
38	B	-	A, C, D
39	D	-	A, B, C
40	B	D	A, C
41	A	B, C, D	-
42	D	A, C	B
43	B	C	A, D
44	D	B	A, C
45	B	A	C, D
46	D	A, B, C	-
47	D	-	A, B, C
48	A	B	C, D
49	A	C	B, D
50	D	A	B, C
TOTAL	50	25	125

Table 4. Frequency distribution of functioning distractor

	Frequency	Percentage
Number of items	50	
Number of distractors	150	
Distractor with frequency <5%	125	83,33%
Distractors with discrimination ≥ 5%	25	16,67%

The analysis results in table 3 show a recapitulation of the distractor efficiency. Of the 50 questions with 150 distractors, 48 samples were tested. Where 27% of the answer sheets will

Technium Social Sciences Journal
Vol. 41, 44-57, March, 2023
ISSN: 2668-7798
www.techniumscience.com

be taken from the top, which is from now on referred to as the upper group (higher group), and 27% of the answer sheet from the bottom, which is from now on referred to as the lower group. The remaining 46% is set aside. Regarding discrimination index level, 25 questions about distractor functioning and 127 questions about distractors not working. Of the 23 distractors that functioned, the distractor rate was 6% as much as 7, 8% as many as 3, 10% as many as 3, 12% as many as 3, 14% as many as 2, 16% as many as 2, 18% as many as 2, 20% as many as 1, 22% as many as 1, 27% as many as 1, and 43% as many as 1. At the same time, the 125 malfunctioning distractors have a distractor rate of 0%, as much as 60, 2%, as much as 45, and 4%, as much as 19.

(Arikunto, 2008) stated that it is essential to analyze the questions that have been made to find out the good, bad, and wrong questions to correct further the questions that are considered harmful or inaccurate**.** In addition, from the results of the analysis, it can be seen that students' weaknesses in the learning indicators that the teacher has taught.

The difficult index analysis (Sudijono, 2011) stated that for items based on the results of the study are included in the excellent category (in the sense that the degree of difficulty of the item is sufficient or moderate), the item should be immediately recorded in the question bank book. The question items can be reissued in the learning outcomes tests in the future. Then for items that fall into the easy category, the tester (teacher) should re-examine, track and trace so that almost all testees can know the factors that cause the item to be answered. According to (Kementrian Pendidikan Nasional, 2010), If the question item belongs to the easy category, then the prediction of the question is: the deception of the question item does not work; Most students answered the question correctly item, meaning that most students had understood the material in question. Furthermore, for things that fall into the difficult category, the tester (teacher) should re-examine, track and trace so that it can be known the factors that cause the item in question be answered by the testee / student. According to (Kementrian Pendidikan Nasional, 2010), If an item of question belongs to the category of difficult, then the prediction is: the item of the question may be the wrong answer key; that item has two or more correct answers; the material in question has not been taught or has not completed the learning, so that the minimum competencies that must be mastered by students have not been achieved; the measured material is not suitable to be asked using the given form of the question; Or the sentence soal is too complex and long.

Regarding discrimination index, (Sudijono, 2011) stated that items that already have good discrimination index (*satisfactory, good, and excellent*) should be included in the bank book about learning outcomes tests. Then for items whose discrimination index is still low (*poor*), the tester (teacher) should search for later improvement, and after being corrected, can be submitted again in the upcoming learning outcomes test; later, the item is analyzed furthermore, whether the discrimination index increases or not. Especially for items whose item discrimination numbers are marked negative, we recommend that in the future, learning outcome tests do not need to be reissued. According to (Kementrian Pendidikan Nasional, 2010), If a question item cannot distinguish between the two students' abilities, then the possibility that occurs in the question is: the answer key to the question item is not correct; that question item has two or more correct answer keys; the measured competence is unclear; the deceptor does not work; The material asked is too difficult, so many students are guessing; or most students who understand the material in question think there is misinformation in the question item.

Then in terms of the function of the distractor, according to (Sudijono, 2011) that distractors that can already perform their roles properly can be used again in future tests, while

distractors that have not been able to function properly should be repaired or replaced with other distractors

4. Conclusion

Based on the analysis of the validity test results on 50 questions, it was found that 19 questions had valid scores for understanding. These questions can be categorized into two groups: firstly, 13 questions had a moderate correlation (0.4<rxy<=0.6), and secondly, 6 questions had a high correlation (0.6<rxy<=0.8). Furthermore, the reliability test analysis, which used the Spearman Brown formula, resulted in a high reliability value of 0.76 (0.61<rxy<=0.8). In terms of the level of difficulty, 3 questions were of medium difficulty (6%), 47 questions were considered easy (76%), and 9 questions were very easy (18%). In addition, the discrimination index revealed that 8 questions had very low discrimination (16%), 22 questions had low discrimination (44%), 9 questions had medium discrimination (18%), and 10 questions had high discrimination (20%). Lastly, the distractor efficiency analysis showed that 125 malfunctioning questions (83.33%) and 25 functioning questions (16.67%) were included in the assessment.

References

[1] Alexander, R. J. (2008). *Education for All, The Quality Imperative and the Problem of Pedagogy*. www.robinalexander.org.uk

[2] Alsoufi, A., Alsuyihili, A., Msherghi, A., Elhadi, A., Atiyah, H., Ashini, A., Ashwieb, A., Ghula, M., Ben Hasan, H., Abudabuos, S., Alameen, H., Abokhdhir, T., Anaiba, M., Nagib, T., Shuwayyah, A., Benothman, R., Arrefae, G., Alkhwayildi, A., Alhadi, A., … Elhadi, M. (2020). Impact of the COVID-19 pandemic on medical Education: Medical students' knowledge, attitudes, and practices regarding electronic learning. *PloS One*, *15*(11), e0242905. https://doi.org/10.1371/journal.pone.0242905

[3] Amelia Suek, L. (2021). ITEM ANALYSIS OF AN ENGLISH SUMMATIVE TEST. *Pattimura Excellence Journal of Language and Culture*, *1*(1), 9–18. https://ojs3.unpatti.ac.id/index.php/pejlac

[4] Arikunto, suharsimi. (2008). *Dasar-dasar Evaluasi Pendidikan*. Bumi Aksara .

[5] Arora, P., Ganguly, D., & Jones, G. J. F. (2015). The Good, the Bad and their Kins. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 1232–1239. https://doi.org/10.1145/2808797.2809318

[6] Arta, W. S., Asib, A., & Sri Wahyuni, D. (n.d.). *A Content-Form Analysis of English Final Test Items for The Second Semester of The Eleventh Grade Students of SMA Negeri*.

[7] Brown, G. T. L., & Abdulnabi, H. H. A. (2017). Evaluating the Quality of Higher Education Instructor-Constructed Multiple-Choice Tests: Impact on Student Grades. *Frontiers in Education*, *2*. https://doi.org/10.3389/feduc.2017.00024

[8] 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. https://doi.org/10.18203/2320-6012.ijrms20191313

[9] Damayanti, A., Wennyta, & Munawwaroh, K. (2018). AN ANALYSIS ON THE ITEMS DIFFICULTY LEVEL OF ENGLISH SEMESTER TEST AT THE TENTH GRADE STUDENTS OF SMAN 3 JAMBI CITY ACADEMIC YEAR 2016/2017. *Journal of English Language Teaching (Jelt)*, *2*(1).

[10] Daniel, E. (2016). The Usefulness of Qualitative and Quantitative Approaches and Methods in Researching Problem-Solving Ability in Science Education Curriculum. *Journal of Education and Practice* , *7*(15), 91–100. www.iiste.org

[11] Enneking, K. M., Breitenstein, G. R., Coleman, A. F., Reeves, J. H., Wang, Y., & Grove, N. P. (2019). The Evaluation of a Hybrid, General Chemistry Laboratory Curriculum: Impact on Students' Cognitive, Affective, and Psychomotor Learning. *Journal of Chemical Education*, *96*(6), 1058–1067. https://doi.org/10.1021/acs.jchemed.8b00637

[12] Ertikanto, C., Distrik, W., Putu, I., & Nyeneng, D. (2022). Implementation of Written Assessment Higher Order Thinking Skills in Physical Learning with a Scientific Approach based Blended Learning. *Jurnal Pembelajaran Fisika (JPF)*, *10*(1), 11–22. https://doi.org/10.23960/jpf.v10.n1.202202

[13] Haji, F. A., Khan, R., Regehr, G., Drake, J., de Ribaupierre, S., & Dubrowski, A. (2015). Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. *Advances in Health Sciences Education : Theory and Practice*, *20*(5), 1237–1253. https://doi.org/10.1007/s10459-015-9599-8

[14] Hill, H. C., & Grossm An, P. (2013). Learning from Teacher Observations: Challenges and Opportunities Posed by New Teacher Evaluation Systems. In *Harvard Educational Review* (Vol. 83, Issue 2).

[15] Hoque, M. E. (2016). Three Domains of Learning: Cognitive, Affective and Psychomotor. In *The Journal of EFL Education and Research* (Vol. 2). www.edrc-jefler.org

[16] Karim, S. A., Sudiro, S., & Sakinah, S. (2021). Utilizing test items analysis to examine the level of difficulty and discriminating Power in a teacher-made test. *EduLite: Journal of English Education, Literature and Culture*, *6*(2), 256. https://doi.org/10.30659/e.6.2.256-269

[17] Kaur, M., Singla, S., & Mahajan, R. (2016). Item analysis of in use multiple choice questions in pharmacology. *International Journal of Applied and Basic Medical Research*, *6*(3), 170. https://doi.org/10.4103/2229-516X.186965

[18] Kementrian Pendidikan Nasional. (2010). *PANDUAN ANALISIS BUTIR SOAL*.

[19] Kristanto, T., Maulana Hadiansyah, W., & Nasrullah, M. (2020). Analysis of Higher Education Performance Measurement Using Academic Scorecard and Analytical Hierarchy Process. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–6. https://doi.org/10.1109/ICIC50835.2020.9288628

[20] Lai, G., Xie, Q., Liu, H., Yang, Y., & Hovy, E. (2017). *RACE: Large-scale ReAding Comprehension Dataset From Examinations*. http://arxiv.org/abs/1704.04683

[21] Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping It Complex. *Journal of Teacher Education*, *64*(3), 226–243. https://doi.org/10.1177/0022487112473837

[22] Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Annals of PIMS*, 310–315.

[23] Malau-Aduli, B. S., & Zimitat, C. (2012). Peer review improves the quality of MCQ examinations. *Assessment & Evaluation in Higher Education*, *37*(8), 919–931. https://doi.org/10.1080/02602938.2011.586991

[24] Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J.

(2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher Education: A meta-analysis. *Computers & Education*, *70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

[25] Nitko, A. J. (1996). *Educational Assessment of Students. Second Edition*. Prentice-Hall .

[26] Olutola, A. T. (2015). Item Difficulty and Discrimination Indices of Multiple Choice Biology Tests. *Liceo Journal of Higher Education Research*, *11*(1), 16–30. https://doi.org/10.7828/ljher.v11i1.890

[27] Posavac, E. J. (2016). *Program evaluation : methods and case studies* (Vol. 08). Prentice Hall.

[28] Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1), 1301013. https://doi.org/10.1080/2331186X.2017.1301013

[29] Rehman, A., Aslam, A., & Hassan, S. H. (2018). ITEM ANALYSIS OF MULTIPLE CHOICE QUESTIONS. *Pakistan Oral & Dental Journal Vol 3*, *38*(2), 291–293.

[30] Richards, J. C. (2013). Curriculum Approaches in Language Teaching: Forward, Central, and Backward Design. *RELC Journal*, *44*(1), 5–33. https://doi.org/10.1177/0033688212473293

[31] Rowntree, D. (1987). *Assessing students : how shall we know them?* K. Page.

[32] Saputra, M. D., Joyoatmojo, S., & Wardani, D. K. (2018). The Assessment of Critical-Thinking-Skill Tests for Accounting Students of Vocational High Schools ARTICL E IN FO ABSTRACT. *International Journal of Educational Research Review (IJERE)*, *3*(4), 85–96. www.ijere.com

[33] Shaikh, S., Kannan, S. K., Naqvi, Z. A., Pasha, Z., & Ahamad, M. (2020). The Role of Faculty Development in Improving the Quality of Multiple-Choice Questions in Dental Education. *Journal of Dental Education*, *84*(3), 316–322. https://doi.org/10.21815/JDE.019.189

[34] Sharma, D., & Kumar, S. (2021). Expected learning outcomes by New Media usage for Senior secondary school students-A Structural Equation Modelling perspective. *Turkish Journal of Computer and Mathematics Education*, *12*(11), 3604–3613.

[35] Sijabat, A., Febrianty Sianipar, H., Siahaan, T. M., & Sijabat, O. P. (2021). Legal Protections for Lecturers. *Proceedings of the 2nd International Conference on Law and Human Rights 2021 (ICLHR 2021)*, 525–531.

[36] Sudijono, A. (2011). *Pengantar Evaluasi Pendidikan*. PT Rajagrafindo Persada. https://opac.perpusnas.go.id/DetailOpac.aspx?id=498690

[37] Sugianto, A. (2016). AN ANALYSIS OF ENGLISH NATIONAL FINAL EXAMINATION FOR JUNIOR HIGH SCHOOL IN TERMS OF VALIDITY AND RELIABILITY. *Journal on English as a Foreign Language*, *6*(1), 29–40. http://e-journal.iain-palangkaraya.ac.id/index.php/jefl

[38] Sugiyono. (2006). *Statistik untuk Penelitian* (E. Mulyatiningsih (ed.)). CV.ALFABETA.

[39] Tinambunan, W. (1988). *Evaluation of student achievement*. Depdikbud. https://onesearch.id/Record/IOS3107.7598/Details

[40] Toha, M. (2010). *AN ITEM ANALYSIS OF ENGLISH SUMMATIVE TEST FOR THE FIRST YEAR OF JUNIOUR HIGH SCHOOL*.