Journal of Engineering Education Transformations ,
Volume 36 , No. 2 , October 2022 , ISSN 2349-2473, eISSN 2394-1707

10.16920/jeet/2022/v36i2/22153

Study on the Correlation Between the Discrimination Index, Facilitation Value and Distractor Efficiency of a Formative Assessment Useful Tools for OBE Practices

Arun Kumar B.R
BMS Institute of Technology and Management
arunkumarbr@bmsit.in

Abstract- Different assessment tools are used to assess engineering knowledge and skills. Multiple choice questions (MCQs) and regular internal tests are commonly used techniques for assessing students. But often the imperfections in the framing of the question can affect students' results and impede the objective of evaluating their knowledge. In this article, the method of evaluating the quality of MCQs and internal test questions framed for PG students in a programming language was considered. For the analysis, Kelly's method was included by considering fast learning group and slow learning groups performances in both the methods of evaluation. Facilitation value, discrimination index, and distraction efficiency were estimated for the items attempted by the group of MCA students. The results obtained show that the item analysis provided necessary data for improvement in question formulation and helped in reviewing the quality of items and tests also. Questions having a lower difficulty index were significantly associated with a higher discrimination index and higher distractor efficiency.

Keywords : Assessment, Difficulty index, Discrimination index, Distractor efficiency, Formative assessment.

Arun Kumar B.R
BMS Institute of Technology and Management
arunkumarbr@bmsit.in

1. Introduction

Present engineering education is based on teaching, learning, and assessment strategies (Nguyen et al., 2020). Assessment strategies are basically two types summative and formative. Summative focus on pass-fail, graduation whereas formative is on how and what to improve on (Serdar et al., 2020) active learning methods in undergraduate or postgraduate engineering programs must address the graduate attributes such as teamwork, creativity, application concepts, design, communication, project management, and entrepreneurship skills, etc., as defined by ABET. Students have the capability of contributing to content creating and act as partners-in learning. So, encouraging students to ask questions and to create examination questions along with solutions is the formative assessment task which facilitates the students in knowledge assimilation (Collier & Tiffany, 2020).

The above flow chart is based on the assessment that enables teachers to motivate learners as per their performance. The design of the assessment should be in line with course goals and course goals' cognitive level should map to the cognitive level of the questions in the assessment. To declare the performance of course goal attainment has to be measured. To measure the course goal attainment, the assessment may be designed keeping in mind skills and cognitive levels set in the course goals. MCQs designed to align with different cognition levels of Blooms 'that is from understanding level to evaluate

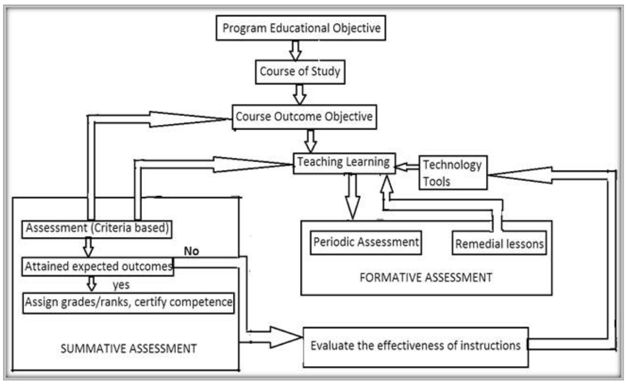

Fig. 1 : Flow chart of achieving successful program educational objectives through summative and formative assessments.

level. MCQ with higher-order thinking refers to the cognition process that includes an application, analysis, evaluation, and creating. It should be noted that the cognitive level of the course outcome should match with the cognitive level of the test item. MCQs test is a common tool both for formative and summative assessment. It is difficult to get data from the MCQ examination conducted for summative assessment. In this work, an analysis of MCQ test items of formative assessment is considered (Denna et al., 2020).

Adopting suitable assessment/evaluation methodology is challenging but essentially significant, since the interpretation of test results should enable correct evaluation of the students' performance, inputs test items preparation skilfully, and TL methodology gets impetus from test items analysis and reinforcement wherever required. Even though no AE the method is error-free the teacher needs to be aware and competent enough to design suitable test items as per the cognitive level at which the performance to be tested (Kaur et al., 2020).

Even today there exists a large scope for enhancing the quality of teaching which is a continuous process, in fact, quality of teaching also reflects in test items. It is essential to engineer the test items design, administer the tests, and verify/validate assessment and evaluation processes before arriving at valid conclusions. The widely used MCQ has a set of rules to be governed during its construction. Well-defined scientific procedures and principles are to be strictly adopted in test item analysis to assess/evaluate students 'performance as well as the quality of test items (Ebel, 1939; Kaur et al., 2020; Kelley, 1939; Quaigrain & Arhin, 2017).

This work analyses the test items which are designed to assess engineering knowledge and skills of the postgraduate program namely Master of Computer Applications (MCA) offered in Visvesvaraya Technological University, Belgaum, India. The approach for the analyses was considered and implemented for the object-oriented programming with C++ course. The procedures of analysis were based on the articles documented test item analysis of various kinds of students' responses (Rehman et al. 2018).

2. Methodology

Fifty-three number of first-year Post Graduate in Computer Applications (MCA) students who appeared for MCQ test in the course "object-oriented programming with C++" (C++). MCQ test was conducted and considered for a maximum of fifteen marks without negative marking where each question of one mark each with four options were designed. Among the four options, one answer is correct and three other options serve as distractors, the rationale that explains the correctness of the options given. Internal Assessment test (IA) in the course C++ was conducted for a maximum of 40 marks. There are five questions to be answered where each question carries 8 marks. Both the MCQ test and IA were formative in nature and analyzed in this research work. To analyze the data, a one-way analysis of variance is adopted for providing the statistical whether there is a significant difference between the means of MCQs and IA test results.

The correlation analysis is a Statistical procedure; by which we can determine the degree of association or relationship between two variables. Each of the samples is considered as a simple random sample. Populations from which the samples are drawn are normally distributed. Each one of the samples is independent of the other samples. The level of significance is fixed to 0.05 and the critical reason for acceptance was fixed to 0.025. total variation present in the sample data will split up into new two components as follows:

(A) Variation within the subgroups of samples and

(B) Variation between the subgroups of the samples.

Test statistic F is calculated based on

B $F_{cal} = \frac{Variance\ between\ the\ FLG\ and\ SLG}{variance\ within\ the\ FLG\ and\ SLG}$:levant

Building an unpredictable but appropriate relevant question bank (QB) is another difficult task. However, item analysis carried out on each question based on parameters such as difficulty index, discrimination index and distractor efficiency enables us to build a suitable database.To build a QB of a particular course, it is necessary to analyze the existing test items based on the response given by the students in the past examination to decide whether the test item can be retained in the QB or removed from it. To make such decisions, it is required to analyze how a set of students responded to the test item of the particular assessment of the course. Item response theory (IRT) is the major psychometric paradigm with a family of models for constructing, scoring, and analyzing assessments. Dichotomous IRT Models are the most suitable models for analyzing MCQ test items where there is a possibility of either correct or incorrect scores only exists (Burud et al., 2019; Chalmers, 2020). Amongst the three dichotomous models: 1PL/Rasch, the 2PL, and the 3PL, the 3PL model uses difficulty, discrimination, and guessing as parameters. In the case of more than two possible scores, polytomous models are used for analyzing essay item assessments. The integers values are typically considered in all IRT models (Chalmers, 2020; Lopez, 1998). The difference between the proportions of high and low scorers answering a dichotomous item correctly is given by the item discrimination index, where high values flag good items, low values indicate bad items.

This research work adopts the conventional approach by extracting sample of 'n' students of high scorers called fast learning group (FLG) and low scorers tagged as slow learning group (SLG) (Jari, 2018). The distribution of students typically follows Kelley's method, 27% of the high scorers and low scorers are the FLG and SLG respectively, the remaining 46% of students are considered normal (Kelley, 1939). Then the item discrimination index is DI = p(FLG) - p(SLG), where p(FLG) and p(SLG) are the proportions of correct responses (answers) given by FLG and SLG respectively. The maximum value of DI, Max(DI), is 1.0 and occurs when all the students in FLG respond with the correct answer, and all the students in SLG fail to give the correct answer.The value of Max (DI) is limited by the easiness of an item for the whole sample, e-value, e(G). When $0.27 < e(G) < 0.73$, i.e., when between 27% and 73% of the sample succeed, this leads to the highest possible discrimination that is when all the students of FLG succeed, i.e., p(SLG) = 1.0, and all the students in SLG fail i.e., p(SLG) = 0, so that Max(d) = 1.0. The item is most discriminating when $e(G) < 0.27$, that is p(SLG)=0 and p(FLG) = e(G)/0.27, so that Max(d) = e(G)/0.27. Similarly, when $e(G) > 0.73$, Max(DI) = (1 - e(G))/.27. An item is not discriminating when p(SLG) = p(FLG) = e(G). Since negatively discriminating items contradict the test as a whole, they are eliminated.

3. Mcq Test Items Analysis

FLG is identified based on scores in the MCQs by considering those who have scored 60% and above to a maximum of 27% of the total students in the upper-performance group. The test items are analysed and their corresponding facility value (FV) and discriminating index (DI) are computed. Fig. 2 shows the FLG group's performance range of marks obtained and the test items.

Slow learning group is identified based on the score in MCQ quiz by considering those who have scored less than 60% to a maximum of twenty-seven percent of the total students in the lower performance group. Fig. 3 shows the SLG group's performance range of marks obtained and the test items. FV and DI were computed and tabulated in Table 1 as per the standard scientific procedures adopted in many

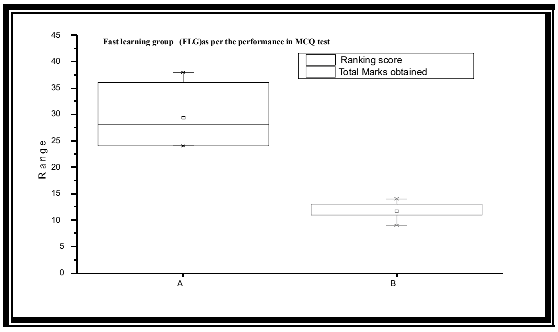

Fig.2 : FLG group's performance range of marks obtained and the test items.

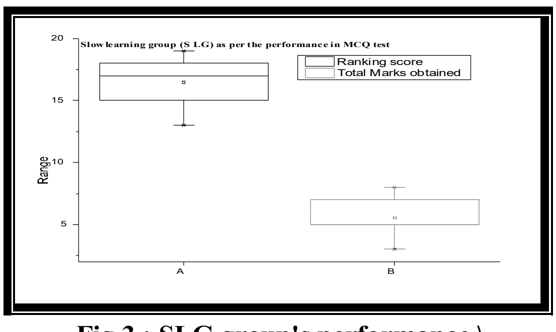

Fig.3 : SLG group's performance \ range of marks obtained and the test items.

Table 1 : Estimated Fv And Di Of Test Items For the Mca Students Who Scored Marks In Their Mcqs

Q.no.	No. of students responded with correct answer in FLG (RF)	No. of students responded with correct answer in SLG (RS)	Max. marks	Facilitation value (FV) = [(RF +RS)] / (N)	Discrimination Index (DI) = (RF– RS) / group size
1	0	0	1	0	0
2	13	7	1	0.66	0.4
3	15	7	1	0.73	0.53
4	14	9	1	0.76	0.33
5	14	8	1	0.73	0.375
6	12	5	1	0.56	0.466
7	12	10	1	0.73	0.13
8	14	3	1	0.56	0.73
9	11	10	1	0.70	0.066
10	14	11	1	0.83	0.2
11	8	1	1	9/30=0.3	(8-1)/15=0.46
12	13	3	1	16/30 =0,53	(13-3)/15=0.66
13	13	3	1	16/30 =0.53	(13-3)/15=0.66
14	15	5	1	20/30=0.66	(15-5)/15=0.66
15	8	1	1	9/30=0.3	(8-1)/15=0.46

	N total	Mean	Standard Deviation	Sum	Minimum	Median	Maximum
FLG	15	11.73333	3.89994	176	0	13	15
SLG	15	5.53333	3.62268	83	0	5	11
FV	15	0.57556	0.22306	8.63333	0	0.66667	0.83333

scientific reports.

It is found that average FV and DI is 0.572 and 0.408 respectively. The computed values are interpreted based on the interpretation table and

Table 2 : Interpretation and Computation of Difficulty Index/facilitation Value [15]

Cut off points (%)	Interpretation	Action	Test items (%)	Remarks / Item no.
<30	Too Difficult	Revise/ Discard	6%	1
30-70	Good/ Acceptable	Store	60%	2,6,8,9,11,12,13, 14,15
50-60	Excellent / ideal	Store	26%	6,8,12 and 13
>70	Tool easy	Revise/ Discard	33%	3,4,5,7 and 10

corresponding action is taken to refine the MCQ bank of the particular course.

As indicated in Table II, one test item that is 7% of the test items are to be revised/discarded since no students attended and also too difficult to attend as DI< 30%, 14 test items that are 93 % of the test items are in acceptable condition since test items are in the range of FV of 30%- 70%. Further as per the response of the students, 12 items are excellent and hence can be continued in the question bank. It is necessary to revise too easy questions also, such 5 questions are found in the analysis with FV> 70%. The interpretation table for DI is prepared and tabulated in Table 5 based on the methods reported in the scientific literature.

Table 3 :interpretation and Computation of Discrimination Index

Cut off points (%)	Interpretation	Action	Test items (%)	Remarks/ item no.
If DI <=0.19	Poor	Eliminate/ completely revise	20	1, 7 and 9
If 0.20 ≤ DI ≤ 0.29	Marginal	Needs revision	7	9
If 0.30 ≤ DI ≤ 0.39	Good	Little or no revision required	7	4
If DI ≥ 0.40	Excellent , item is functioning satisfactorily	Store	66	2,3,5,6,8,11,12,13,14 and 15

As per the Table III, 66% of the test items shall be retained in the MCQ bank, 20% of the items must be little revised or eliminated. 14 % of the items (2 items10 who are in the range of <= DI <= 0.39 need

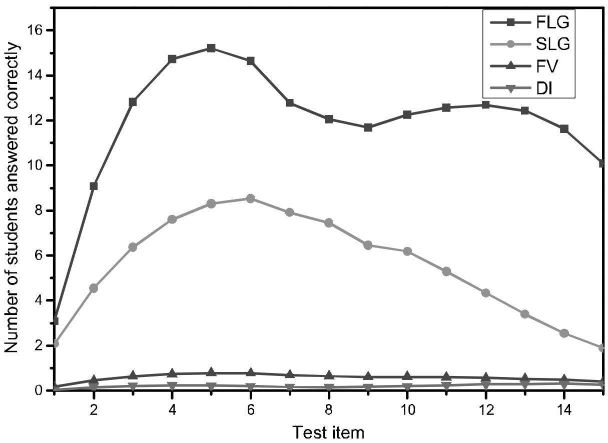

Fig. 4 : Variation of FLG, SLG, FV and DI with respect to test items (15 Nos.)

little revision or no revision required. The variation of FLG, SLG, FV and DI were showed as in Fig. 4 and Fig.5 shows mean value of the influence factor for the FV and DI items.

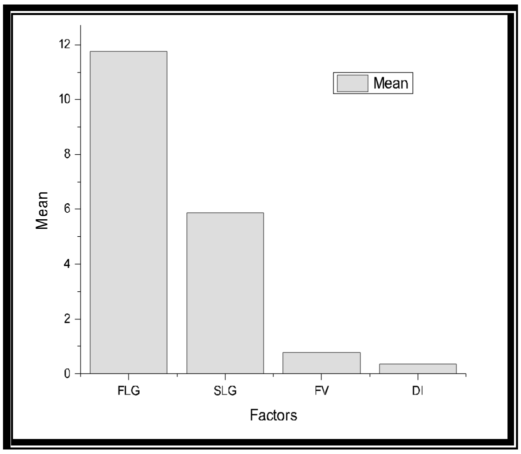

Fig. 5 : Mean value of the influence factor for the FLG, SLG, FV and DI items

Table 4 : Anova Analysis Output For the Mcqs

Sample Size	Mean	Standard Deviation	E of Mean	
A	15	38.93333	1.16292	0.30026
B	15	35.26667	3.5146	0.90746
C	15	24.93333	4.38287	1.13165
D	15	20.13333	10.92747	2.82146
	DF	Sum of Squares	Mean Square	F Value Prob>F
Model	3	3456.45	1152.15	30.25528 9.24161E-12
Error	56	2132.53333	38.08095	
Total	59	5588.98333		
R-Square	Coeff Var	Root MSE	Data Mean	
0.61844	0.20696	6.17098	29.81667	

Table 5 :anova Analysis Output For The Ia Marks

Sample	Size	Mean	Standard Deviation	SE of Mean
A	8	66.375	40.02834	14.15215
B	8	37.75	22.2951	7.88251
C	5	0.6908	0.16408	0.07338
D	5	0.337	0.13673	0.06115
	DF	Sum of Squares	Mean Square	F Value Prob>F
Model	3	19630.23336	6543.41112	30.25528 9.24161 E-12
Error	22	14695.55747	667.97989	
Total	25	34325.79083		
R-Square	Coeff Var	Root MSE	Data Mean	
0.57188	0.80175	25.84531	32.23612	

Tables IV and V show the ANOVA output for the MCQ's and IA tests. In both the cases there is a alternative hypothesis suggested with F-Values as positive values. Further, coefficient variance of MCQ is smaller than the Coefficient variance of IA marks. Hence, the influence of IA marks is more significant than the MCQ's. But still the MCQs made the studies of the students as more significant in sharpening of their thinking, extra activity through self and additional reading.

A. Effectiveness of distractors of MCQ test item

Each test item has the correct option (answer key) and incorrect alternatives that are called distractors. The ability of distractions is the essential character of the MCQ and considered as effective if at least 5% of the students choose correct response (Pearson et al., 2018). Several research articles have demonstrated that three distractors are as efficient as four/five distractors and recommended having only one option as correct and distractors as plausible.

While determining the effectiveness of the distractors all MCQs of formative assessment having three distractors are considered (consistency maintained). Out of 15 test items, a sample consisting of first five test items are computed to show their distractor efficiency by computing no. of students from SLG and FLG responded by selecting that particular option, and effectiveness of each distractor is shown in Table 5 where NL (Number of lower group students) and NU (Number of upper group students).

Table VI shows the Interpretation of the effectiveness of distractors and also indicates the

Table 6 : Computation Of Effectiveness Of Distractors

Test Item	Lower group- 27%				Upper group- 27%				ED= (NL~NU)/n			
	NL-A	NL-B	NL-C	NL-D	NU-A	NU-B	NU-C	NU-D	ED-A	ED-B	ED-C	ED-D
1	3	7	5	0	1	13	0	1	0.133	-0.4	0.3	-0.06
2	3	3	9	0	2	0	13	0	0.06	0.2	-0.26	0
3	3	6	4	2	1	14	0	0	0.133	-0.53	0.26	0.133
4	3	10	2	0	1	14	0	0	0.133	-0.26	0.133	0
5	6	3	6	0	1	1	13	0	0.33	0.133	-0.46	0

computation of functional and non-functional distracters (NFD). NFD is an item in the incorrect option chosen by less than 5% of students and the functional or effective distractors is the option selected by 5% or more.

On the basis of the number of NFDs in an item, DE ranges from 0 to 100%. If an item contains three or two or one or nil NFDs then (DE) would be 0. The effectiveness or difficulty of the distractor also depends on the given set of students in the particular batch. While designing the quality MCQ distractors need to be plausible on the other hand the NFDs to be reducing.

The difficulty of a distracter is depending on its attractiveness to a given population of individuals. Easy distracters can be discarded by almost all examinees. On the other hand, difficult distracters have high effectiveness and response frequency. Designing plausible distracters and reducing NFDs is an important aspect for framing quality MCQs.

Table 7 :
Interpretation Of De For The Sample [19][20]

Test item no.	No. of NFDs	No. of FDs	Action
1	1	2	Discard distractor (d)
2	1	2	Discard distractor (d)
3	1	2	Discard distractor (d)
4	2	1	Discard distractors and (d)
5	1	2	Discard distractor (d)

By referring to Table VI, test items, 1, 2, 3 and 5, has one NFD each, due to the distractors, that is option (d) in test item 1, 2, 3 and 5, distractors options (c) and (d) of test item 4.

If DE =0, then such item has to be discarded as it fails to attract less than 5% of the examinee.

Table 8 : Interpretation Of De [15]

DE (item)	Action
0	Discard
33.3	Revise
66.6	Store
100	Store

Table 9 : Analysis Of Distractors' Efficiency of Multiple Choice Questions [18] [19][20][21][22]

Indicators	Values
Number of MCQs	15
Total no. of Distracters (wrong answers with total MCQs)	45
Items with only Functional distracters (%) or with no NFDs that is DE with 100% efficiency	2/15 =14%
Overall NFD (%)	16/45= 36%
Overall FDs (%)	36/45=64%
Items with 1 NFD	10/15= 66 %
Items with 2 NFDs	3/15 = that is 20%
Item with 3 NFDs	0

Following table DE table shown in Table VII can be used in decision making.

With reference to the Table IX, 10 test items have 1NFD (66% DE), 3items have 2NFDs (33%DE) and 2 items have only FDs (100% DE). In other words, 12 test items can be restored without any revision, however, remaining 3 items needs revision to meet minimum of 66% DE. In the next section, internal test items are analysed.

B. Item analysis of internal test

Similar to MCQ test item analysis, internal test items are analysed with respect to their FV and DI. The test was conducted for maximum of 40 marks with 8 questions where there exists a choice between

Table 10 : Worksheet Of Flg Ranking

Sl. no	Student Name/USN	Ranking Score (40)	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Total marks
1.	FLG1	40	8	0	7	0	0	8	8	8	39
2	FLG2	40	0	7	0	8	0	8	8	8	39
3	FLG3	40	0	7	0	8	0	6	8	5	34
4	FLG4	40	8	0	7	8	0	0	7	8	38
5	FLG5	40	0	8	0	8	0	7	7	8	38
6	FLG6	40	7	8	0	8	0	0	8	8	39
7	FLG7	39	0	7	5	7	0	0	6	6	31
8	FLG8	39	0	6	0	7	0	8	8	5	34
9	FLG9	39	7	0	0	8	0	8	8	3	34
10	FLG10	39	0	8	0	8	0	8	7	8	39
11	FLG11	39	0	7	0	7	0	8	7	8	29
1	FLG12	38	5	7	0	8	0	0	8	3	31
13	FLG13	37	6	0	0	7	0	3	7	8	31
14	FLG14	37	8	0	0	5	0	8	8	8	37
15	FLG15	37	5	0	0	7	0	8	8	8	36

Table 11 : Work Sheet of Slg Ranking

Sl. no	Student Name/USN	Ranking Score (40)	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Total marks
1.	SLG1	17	1	0	0	2	0	1	0	0	4
2	SLG2	18	2	0	0	0	0	0	0	0	2
3.	SLG3	21	3	7	2	6	0	5	7	0	30
4.	SLG4	21	5	2	0	4	0	0	5	1	17
5.	SLG5	23	0	7	0	0	0	0	0	0	07
6.	SLG6	24	0	2	5	0	0	0	4	8	19
7.	SLG7	24	2	2	0	0	0	0	1	2	07
8.	SLG8	25	0	6	5	8	0	0	7	1	27
9.	SLG9	25	0	6	0	7	0	7	6	5	31
10.	SLG10	28	4	5	0	6	1	0	6	0	22
11.	SLG11	29	0	3	0	6	0	4	3	2	18
12.	SLG12	29	6	0	0	7	0	7	7	8	35
13.	SLG13	29	6	0	6	8	0	0	7	7	34
14	SLG14	30	5	0	0	7	0	2	7	2	23
15	SLG15	31	5	0	0	4	0	2	7	8	26

questions 1and 2, 3 and 4 and 5 and 6, whereas questions 7 and 8 are compulsory to attend. Test was taken by 53 students, by considering 27% of the top performers in the upper group following FLG is formed which is shown in Table 9. Score in FLG ranges from 29 to 39 out of maximum of 40 marks.

Similar to FLG, 27% of total no. of students in the lower performance group are considered in forming SLG whose score ranges from 2 to 35. Table XI indicates SLG in the internal test conducted of the course namely "Object Oriented Programming with C++ ".

After forming the FLG and SLG for each test item, FV and DI are computed and shown in the Table XI. The average values of FV and DI are 0.690 ± 0.05, and

Table 12 : Computation of Fv and Di For 8 Test Items

Item	TMF	TMS	Facilitation value (FV) = (TMF + TMS) / MM * N	Final FV	DI
Question 1	55	39	94/8*30 = 0.39	0.83	0.13
Question 2	65	40	105/8*30 = 0.43		
Question 3	19	18	37/8*30 = 0.15	0.85	0.33
Question 4	104	65	169/8*30 = 0.70		
Question 5	0	1	1/8*30 = 0.0041	0.454	0.42
Question 6	80	28	108/8*30 = 0.45		
Question 7	105	67	172/8*30 = 0.71	0.71	0.31
Question 8	103	44	147/8*30 = 0.61	0.61	0.49

0.337± 0.03 respectively.

With reference to Table XII, TMF stands for Total marks obtained by students in FLG with upper performance, TMS: Total marks obtained by students in SLG with lower performance, n (FLG) : No. of students in FLG, in this case n (FLG) =15, n (SLG) : No. of students in SLG, in this case n (SLG) = 15, N: Total no. of students in groups that is N= n (FLG) + n (SLG) = 15+15= 30.

With reference to Table XII, TMF stands for Total

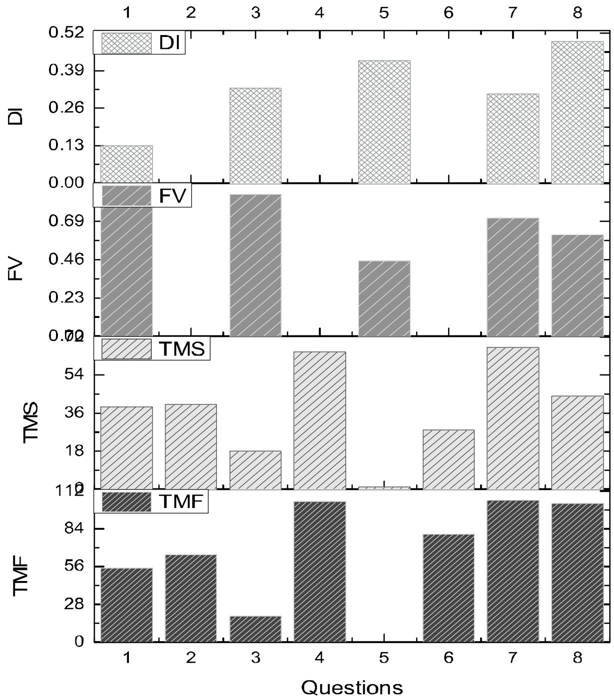

Fig. 3 : Graph shows the variations of TMF, TMS, FV and DI values with respect to test items

Table 13 : Interpretation And Computation of Difficulty Index/facilitation Value [15] for Internal Test Items

Cut off points (%)	Interpretation	Action	Test items (%)	Item no.
<30	Too Difficult	Revise/Discard	-	-
30-70	Good/Acceptable	Store	40	5, 6 and 8
50-60	Excellent / ideal	Store	-	-
>70	Tool easy	Revise/Discard	60	1,2, 3,4 and 7

Table 14 : Interpretation And Computation of Discrimination Index [15]

Cut off points (%)	Interpretation	Action	Test items(%)	Item. No.
If DI <=0.19	Poor	Eliminate/ completely revise	20	1,2
If 0.20 ≤ DI ≤ 0.29	Marginal	Needs revision	--	-
If 0.30 ≤ DI ≤ 0.39	Good	Little or no revision required	40	3,4 and 7
If DI ≥ 0.40	Excellent , item is functioning satisfactorily	Store	40	5, 6 and 8

marks obtained by students in FLG with upper performance, TMS: Total marks obtained by students in SLG with lower performance, n (FLG) : No. of students in FLG, in this case n (FLG) =15, n (SLG) : No. of students in SLG, in this case n (SLG) = 15, N: Total no. of students in groups that is N= n (FLG) + n (SLG) = 15+15= 30.

By referring to Table XIII and Table XIV, it is implied that question 1 and 2 may be eliminated from the question bank or completely revised. The questions 5, 6 and 8 are having DI >= 0.4 with good FV can be retained in the question bank. However, questions 3, 4 and 7 may be discarded or with suitable revision may be added back into the question bank. It shall be noted that 40% of the test items can be accepted without any changes (items 5/6, and 8).

4. Conclusions

From the reported study, it was observed that in the case of IA score ranges from 72.5% to 97.5% falls in FLG whereas score in SLG ranges from 5% to 32.5%. In the MCQ test, the performance of the FLG score ranges from 60% to 93.3%, whereas in SLG score ranges from 20% to 33.3%.

In the study, for the course of C++ the MCQs and the internal tests conducted for PG students were analyzed by the IRT method. In both the techniques, FLG and SLG students' categories based on their ranking score and estimated and correlated the values of DI, FV, and DE for setting the standards. For the present study, 73% of MCQ questions are storable and 27 % of the items need to be revised based on the computed FV and DI values. It is found that DI and FV values for MCQs and for IA test items were found to be 33%. Around 14% of the overall MCQs have effective distraction efficiency. FV and DI values for the internal test items were in the range of 45.4 % to 85% and 13% to 49% respectively. Among the questions, 40 % of the questions considered acceptable, and 60% of the items considered too easy because of the compulsory regular items. Based on the DI value, 40 % of the questions were excellent and 40% of them need no or little revision. Which shows that the quality of the questions in both the assessments meet the global educational standard. It is required to put in more effort in terms of time, knowledge and should possess an interest in designing plausible distractors which leads to the contribution of valid and real test conduction. In this connection, more training and focus is required from the institution to adopt continuous improvement in TLP. The analysis enabled to maintain the question bank of test items with three functional distractors having high discrimination, the moderate difficulty which is useful for subsequent refined AE processes that can discriminate FLG and SLG.

Item analysis is helpful to assess the quality of both test items and test as a whole. It enhances the skills of the teachers in bringing clarity in teaching or focused content deliberation and test conduction skills.

Acknowledgement

Author, Dr. Arun Kumar B.R. is thankful for the encouragement and support extended by the dept. of Computer Science & Engineering, BMSIT&M Principal and Management in carrying out this research work.

References

[1] Anwar, F. (2019). Activity-based teaching, student motivation and academic achievement. Journal of Education and Educational Development, 6(1), 154–170.

[2] 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.

[3] Chalmers, R. P.. (2020). Partially and Fully noncompensatory response models for dichotomous and polytomous items. Applied

Psychological Measurement, 44(6), 415–430.

[4] Collier, D. R., & Gallagher, T. L. (2020). Blogging in elementary classrooms: Mentoring teacher candidates' to use formative writing assessment and connect theory to practice. Teaching/Writing: The Journal of Writing Teacher Education:, 9, article 11.

[5] Deena, G., Raja, K., Nizar Banu, P. K., & Kannan, K. (2020). Developing the assessment questions automatically to determine the cognitive level of the E-learner using NLP techniques. International Journal of Service Science, Management, Engineering and Technology, 11, 16.

[6] Ebel, R. L. (1979). Essentials of educational measurement (3rd ed). Prentice Hall.

[7] K A Nguyen. (2020). C. Lucas, D. Leadbeatter, Student generation and peer review of examination questions in the dental curriculum: Enhancing student engagement and learning. European Journal of Dental Education, 00, 1–11.

[8] Fozzard, N., Pearson, A., du Toit, E., Naug, H., Wen, W., & Peak, I. R. (2018). Analysis of MCQ and distractor use in a large first year Health Faculty Foundation Program: Assessing the effects of changing from five to four options. BMC Medical Education, 18(1), 252.

[9] Kaur, M., Singla, S., & Mahajan, R., Item analysis of multiple choice questions in Pharmacology, International Journal of Applied and Basic Medical Research Published by Wolters Kluwer – Medknow.

[10] Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. Journal of Educational Psychology, 30(1), 17–24.

[11] Metsämuuronen, & Jari. Generalized discrimination Index 2018.

[12] Namdeo, S. K., & Rout, Sushi D. (2016). Assessment of functional and non-functional distracter in an item analysis. International Journal of Contemporary Medical Research ISSN: 2393-915X; (Print): 2454, 7379 | ICV: 50.43 | Volume 3 | Issue 7 | July, 1891–1893.

[13] Pho, V., Ligozat, A., & Grau, B. (2015, June 21–25). Distractor quality evaluation in multiple choice questions [Conference session]. International Conference on Artificial Intelligence in Education, Madrid, Spain.

[14] 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.

[15] Ramirez-Mendoza, R., Morales-Menendez, R., Iqbal, H., & Parra-Saldivar, R. (2018). Engineering education 4.0: Proposal for a new curriculum. In The IEEE global engineering education conference, 2018 (pp. 1273–1282).

[16] Rehman, A., Aslam, A., & Hassan, S. H. (April–June 2018). Item analysis of multiple choice questions, Pakistan Oral and Dental Journal, 38(2) (pp. 291–293).

[17] Sahoo, D. P., & Singh, R. Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. International Journal of Research in Medical Sciences, 5(12).

[18] Serdar, S., Mustafa, H. B., & Mohammad, A. (2020). Computer based evaluation to assess students' learning for multiple-choice question-based exams:CBE-MCQs software tool. Computer Applications in Engineering Education, 1–15.

[19] Tang, T., Vezzani, V., & Eriksson, V. (2020). Developing critical thinking, collective creativity skills and problem solving through playful design jams. Thinking Skills and Creativity, 37.

[20] Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Medical Education, 9, 40.