ITEM ANALYSIS OF AN GENERAL MANAGEMENT AWARENESS/ LEARNING ASSESSMENT TEST IN BUSINESS SCHOOL FOR ITS DIFFICULTY INDEX, DISCRIMINATION INDEX, AND DISTRACTOR ANALYSIS

-Vishal D. Pajankar, National Council of Educational Research and Training, New Delhi

ABSTRACT

Nowadays, Item analysis is an important tool to know the functionality of the set items used in a test, prepared to access the learning level of student. An item which is a multiple-choice question with more than one option out of one is a correct answer called a key and the rest is an incorrect answer(s) called a distractor(s). The main objective of the distractor is to divert the student from the correct answer. Good distractor(s) is always a plausible answer. So to understand the function of the distractor(s), the distractor analysis is very important. In this study, we analyzed the different statistics of item analysis namely difficulty index, discrimination index, distractors, etc. and understand how they are functioning.

Keywords: *Assessment test, item analysis, difficulty index, discrimination index, distractor analysis.*

INTRODUCTION

Item analysis of any assessment test helps to improve the quality of items including its stem as well as distractors. It is an important and relatively simple procedure conducted after the administration of a test. It provides useful information about the reliability and validity of a test item. It tells how the items vary between different levels from difficult to easy. With the difficulty and discrimination index, the questions were able to discriminate between students who performed well on the test, from those who did not. Distractors analysis is an important tool that provides information regarding the individual distractors and the key of a test item. These tools help to modify, improve or remove particular items so they can be further used in subsequent exams.

Multiple choice questions (MCQs) are important tools for any examination or test conducted on a large scale. It is the only source of information for item analysis. (Burud and Agarwal, 2019). Since the item analysis can perform on the responses of these multiple-choice questions. The trend of use of multiple-choice questions has been growing day by day in high stake and low stake examinations/ tests in all countries. The items used in these tests should have a good standard to maintain the fairness, accuracy, and quality of the test (Deepak, at. el., 2015). Carefully

and scientifically constructed multiple-choice questions were able to assess a higher-order cognitive skill such as applications, interpretation, analysis, application, and knowledge.

There are two major types of multiple-choice tests, (i) criterion-referenced tests (CRTs) and (ii) norm-referenced tests (NRTs). Criterion-Referenced Tests (CRTs) are used to decide whether or not an individual can demonstrate mastery in an area of content and competencies while Norm-Referenced Tests (NRTs) where the goal is usually to rank the entire set of individuals to make comparisons of their performances relative to one another (Burud and Agarwal, 2019).

The construction of a quality multiple-choice question or item is an art where the item writer has to use scientific rules to construct the good, quality, and reliable items. The same skill is required for all three levels of items; easy, average, and difficult items. It is not true that easy items can construct by ignoring its rule and only for the construction of difficult items, rules are applicable. Every item has two parts; the first part is a question statement called 'stem' and the second part is options. Minimum two options should be in a multiple-choice question. In common practice, four options of multiple-choice questions are preferred. Options have two parts, the first part is key i.e. correct answer or solution of the given stem and another part is called distractors. Distractors are incorrect answers or solutions of the stem but this incorrect answer closed to the key (correct answer) to distract the respondent. There is no fixed position of the key in a set of options. It may appear at any position. An essential characteristic of distractors is that all options shall present plausible answers and if possible none shall be incorrect. In the overall process, the designing or selection of proper distractors is a very tedious job. Mostly these distractors are responsible to change the nature of an item easily, average, or difficult.

Distractor analysis is an integral part of item analysis. It uses some parameters of item analysis such as item difficulty and item discrimination. The distractor analysis tells how the distractors apart from the key are functioning. Frequencies of the selection of distractors is distracting the students or respondents. A test developer would expect that the distractor is selected by enough of the low achieving candidates for it to be an effective and viable distractor. For analysis of the effectiveness of the distractors, both the item difficulty and item discrimination index can be used. (Asamaah and Ocansey, 2019)

METHODS

The study was conducted in the year 2018-19 in the state of Maharashtra (India) with help of the District Education Authority. The data is collected through the Principal Investigator of this study. In total 60 business schools participated. The schools were selected by sampling procedure 'Probability Proportion to Size' (PPS). From each sampled school maximum of 30 students were selected by a robust random sampling procedure. A criterion was fixed that minimum enrolment must be 10 in that particular school. The test was conducted on General Awareness about Management. The items were prepared on different competency levels considering her/his academic skill in understanding, interpreting, analyzation, etc.

about the content/topic. The test booklet had 15 items with three distractors and one key in each item. In total, 45 distractors and 15 keys were in the test. The test was administered to 659 students. The objective of the test was (i) to know the general awareness of the students in the core contents of Management, and (ii) to identify gaps in their learning. Thirty (30) minutes were allotted to the students to finish the test.

During the evaluation, one mark was awarded to the correct answer and a zero mark for the incorrect answer. So the highest score on the test was 15 and the minimum score was 0. With the score earned by the students, difficulty level, difficulty index (DFI), power of discrimination as measured by the discrimination index (DI), and distractor analysis (DE) for all incorrect options had been analyzed.

ITEM ANALYSIS PROCEDURE

Before feeding the data into database software for analysis, the data was cleaned for irrelevant entries. During the cleaning, some entries such as blank or multiple selections of options were reported against a few items. The percentage of such responses was 2.58 ± 0.42. They remain in the dataset for further analysis. Responses of the items for 45 distractors were analyzed and various indices like difficulty index (DIF_I), discrimination index (DI), and distractor efficiency (DE), were analyzed as:

Item Difficulty Index

Item Difficulty Index (DIF_I) can be determined from a certain portion of the group of students. Arrange the students in ascending or descending order according to their performance in the test score. Identify high scorers 27% of students and low scorers 27% of students from upper and lower groups. A general guideline for the interpretation of item difficulty index values is provided in the following table 1 (Adegoke, 2013; Zubairi and Kassim, 2006 and [24]. Thus, the item difficulty index can be defined as below (see, Singh, 2009)

$$DIF\_I = \frac{UpG + LwG}{N_U + N_L}$$

where,

UpG = Group of 27% students with a high ability to respond to the given test item correctly;
LwG = Group of 27% students with low ability to respond to the given test item correctly;
N_U = Number of students who answered the test item correctly in the upper group
N_L = Number of students who answered the test item correctly in the lower group

Table 1: Interpretation of Item difficulty Index.

Difficulty Index (p)	Interpretation
$p \leq 0.30$	Difficult
$0.31 \leq p \leq 0.70$	Moderate Difficult (Average)
$p \geq 0.71$	Easy

Item Discrimination Index

Item Discrimination Index (DI) is defined as an unbiased index of absolute difference in the number of discrimination made between the upper group and the lower group – it is proportional to the net discriminations made by the item between the groups. It is clear from the following formula that the difference between the proportion of correct answers of the top 27% and the bottom 27% of students (Singh, 2009)

$$DIF_I = \frac{(UpG\text{-}LwG)}{N_U} ; \ (\because N_U = N_L)$$

where,

UpG = Group of 27% students with a high ability to respond to the given test item correctly;

LwG = Group of 27% students with low ability to respond to the given test item correctly;

N_U = Number of students who answered the test item correctly in the upper group (= N_L)

Since its proportion ranges from -1 to +1. The item discrimination index can be evaluated in the following term:

Table 2: Interpretation of Discrimination Index.

Discrimination Index (D)	Item Evaluation
$D \leq 0.19$	Poor items; should be removed or completely revised
$0.20 \leq D \leq 0.29$	Marginal item and needs improvement (revision)
$0.30 \leq D \leq 0.39$	Reasonably good item, possibly little or no revision is required
$D \geq 0.40$	Very good item. It is functioning quite satisfactorily

Ref.: Ebel and Frisbie,1991 and Bichi, 2015

Distractor Efficiency

Distractor Efficiency (DE) is calculated as Non Functional distractor (NFD) from the distractor that has been selected by less than 5% of the students. In this study, about in average 2.58 ± 0.42 % (SD) items had either Null or multiple responses. So the range of non-functioning distractors was extended to 10%, as a special case.

RESULT AND DISCUSSION

The assessment test of students was conducted on general awareness about Management subject. A total of 659 students appeared for the test. There were 15 multiple-choice questions (items) in the test booklet. Every item had three distractors with a key (correct answer). The mean score achieved was 5.64 ± 3.06 (maximum 15 marks). Mean scores according to groups were: lower 2.42 ± 0.96;

middle 5.07 ± 0.95; upper 9.83 ± 1.90. For analysis of item statistics, students were classified into two groups based on their performance in the test. High-scored 27% students grouped as an upper group and low scored 27% students grouped as a lower group. The p-value, DI, and DE were analyzed for each item (Table 4).

Table 3: Summary of Test Statistics Variables.

Number of Items	15
Students	659
Mean Score	5.64 ± 3.06
Mean Score (Upper Group)	9.83 ± 1.90
Mean Score (Lower Group)	2.42 ± 0.96
Item Difficulty Index (p)	0.41 ± 0.06

Table 4: Item-wise Difficulty Index and Discrimination Index.

Items	Item Difficulty Ratio	Lower Group (Item Difficulty)	Upper Group (Item Difficulty)	Item Difficulty Index	Item Discrimination Index
Item_1	0.31	0.13	0.58	0.36	0.44
Item_2	0.48	0.21	0.79	0.50	0.57
Item_3	0.37	0.11	0.65	0.38	0.54
Item_4	0.47	0.25	0.64	0.44	0.39
Item_5	0.45	0.17	0.79	0.48	0.62
Item_6	0.37	0.13	0.71	0.42	0.57
Item_7	0.44	0.24	0.67	0.46	0.43
Item_8	0.30	0.10	0.65	0.37	0.56
Item_9	0.27	0.12	0.49	0.30	0.37
Item_10	0.38	0.10	0.78	0.44	0.68
Item_11	0.30	0.20	0.40	0.30	0.21
Item_12	0.40	0.15	0.71	0.43	0.56
Item_13	0.42	0.18	0.67	0.42	0.49
Item_14	0.34	0.11	0.70	0.41	0.59
Item_15	0.34	0.21	0.59	0.40	0.38

The analysis of item difficulty indices and discrimination indices for all 15 items was given in table 5. A workout for distractor efficiency was also conducted. About 87% of items had an average level (of difficulty) with a mean p-value 0.42 ± 0.04 (SD) whereas 2 (13.33%) items were too difficult. On other hand, 73.77% of items had an excellent discrimination index. Their mean discrimination index was 0.51 ± 0.12 (SD). They properly discriminate between high scorer and low scorer students. Distractor analysis gives an opportunity to study the responses made by the students on each alternative of the item. The analysis of all these items explained how the options were functioning and how distractors distract students to select the proper key. Almost all items were discriminated between high scorer students and low scorer students. In total 45 distractors out of which 4 (26.67%) were non-functioning distractors (NFDs). Only item 11, is weak to discriminate. This item required some revision.

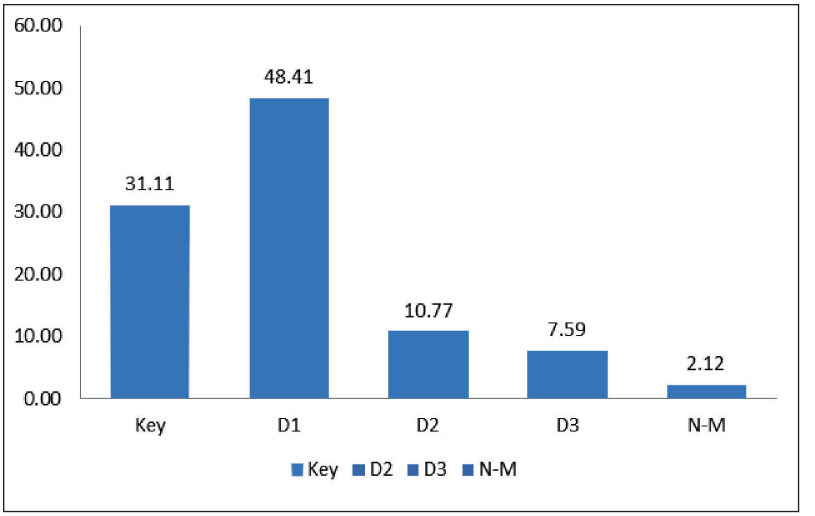
Table 5: Classification of Items according to difficulty and discrimination indices.

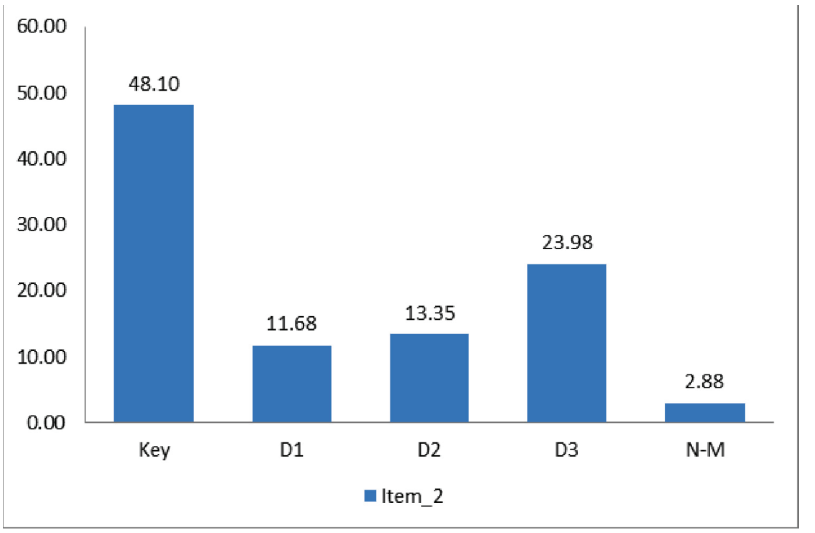
(A) Difficulty Indices					
p-value	**Interpretation**	**Mean p-value**	**Standard deviation**	**DE (%age)**	**Items (n=15)**
p > 0 .70	Easy	--	--	--	0 (n=0%)
0.3 – 0.70	Average	0.42	0.04	92.31%	13 (86.67%)
< 0.30	Difficult	0.30	0.00	100.00%	2 (13.33%)
(B) Discrimination Indices					
D	Interpretation	Mean D	Standard deviation	DE (%age)	No. of Items (n=15)
D ≥ 0.40	Excellent	0.55	0.07	90.91%	11 (73.77%)
0.3 ≤ D ≤ 0.39	Good	0.38	0.01	--	1 (13.33%)
0.2 ≤ DI ≤ 0.29	Marginal	0.21	--	100%	1 (6.67%)
D ≤ 0.19	Poor	--	--	--	--
(C) Between Difficulty Indices (p) and Discrimination Indices (D)					
p value/D	p ≤ 0.30	0.31 ≤ p ≤ 0.70		p ≥ 0.71	Total
D ≤ 0.19	--	--		--	--
DE (%)	--	--		--	--
0.2 ≤ DI ≤ 0.29	1 (6.67%)	--		--	1 (6.67%)
DE (%)	75%	--		--	75%
0.3 ≤ D ≤ 0.39	1 (6.67%)	13 (86.67%)		--	14 (93.33%)
DE (%)	75%	83.33%		--	79%

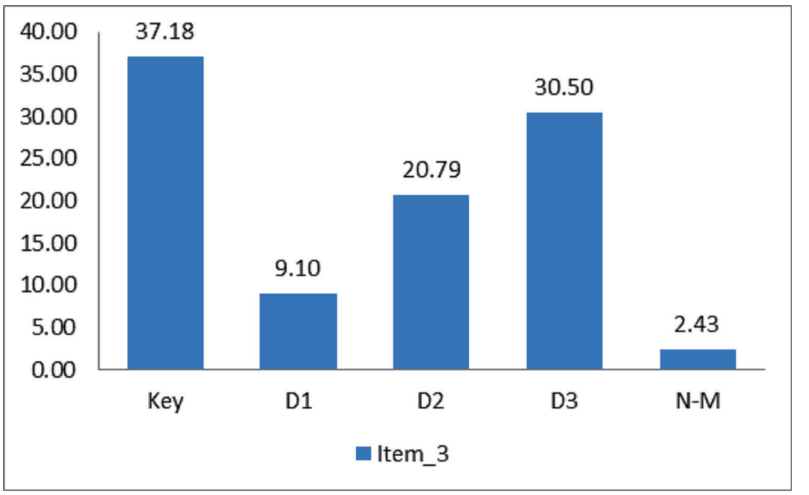
D ≥ 0.40	--	--	--	--
DE (%)	--	--	--	--
Items (15)	2 (13.33)	13 (86.67%)		15 (100%)

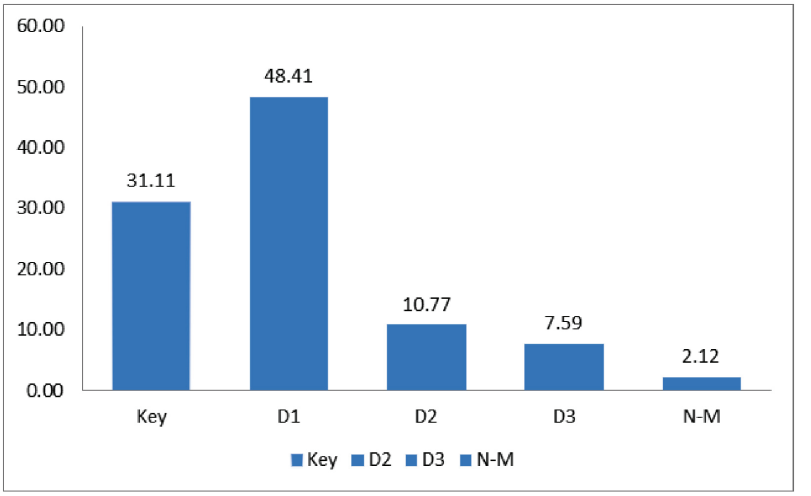
Ref.: Hingorjo, and Farhan, 2012

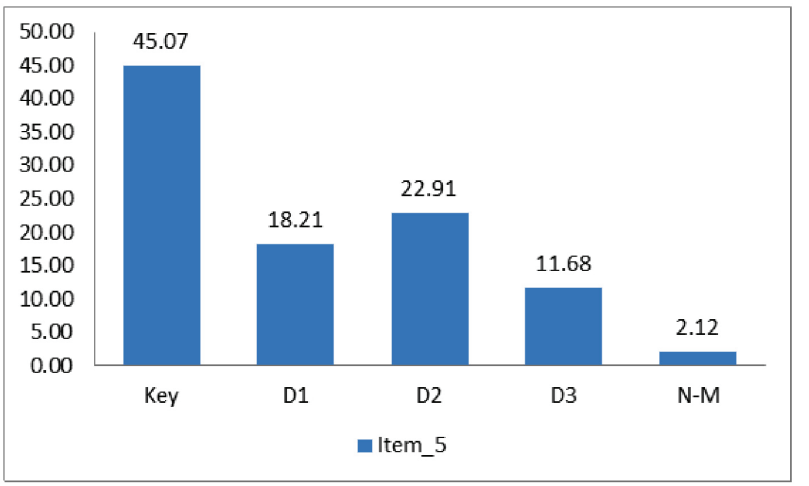
In items 1, 8, 9, and 11, one distractor was much closed to the key so the maximum number of students selected that distractor instead of the key. However, the rest two distractors were unable to distract students, maybe they were far away from the correct answer. Similarly, items 3, 6, and 14 had one distractor close to the key but score less than the key. Detail description of keys and distractors is given below in the form of bar charts for each item.

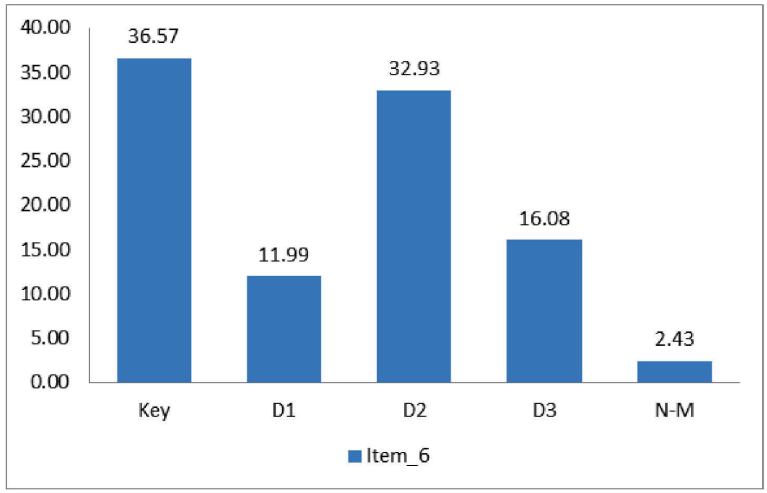


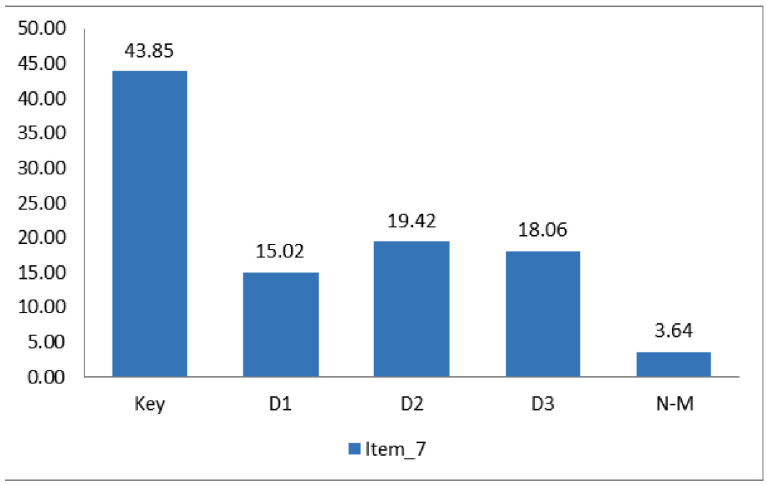


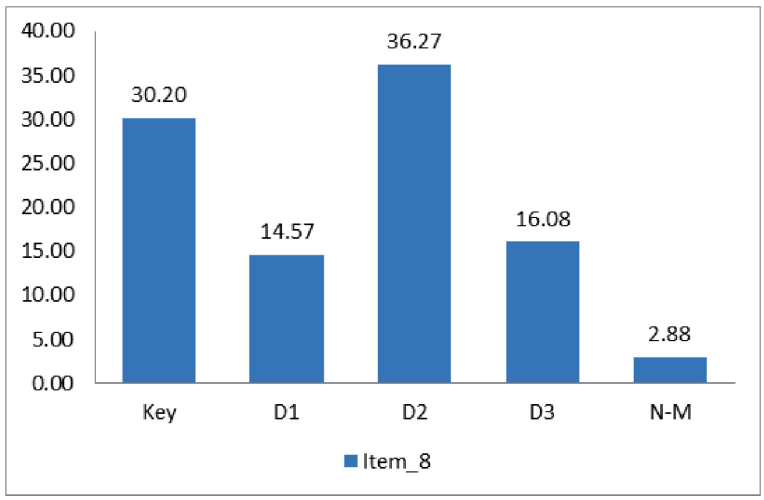

40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
37.18
9.10
20.79
30.50
2.43
Key
D1
D2
D3
N-M
Item_3

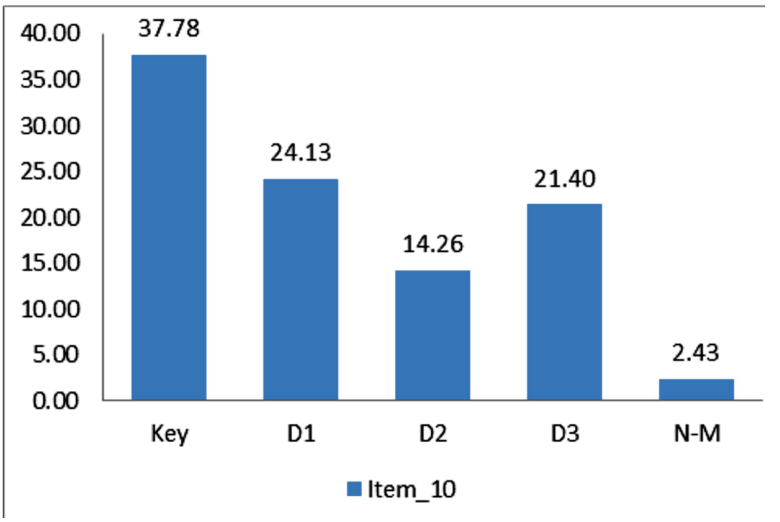

60.00
50.00
40.00
30.00
20.00
10.00
0.00
31.11
48.41
10.77
7.59
2.12
Key
D1
D2
D3
N-M
Key D2 D3 N-M

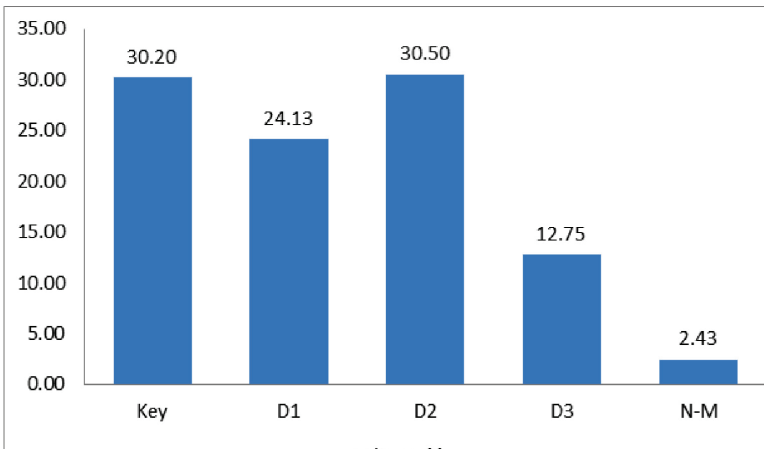

50.00
45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
45.07
18.21
22.91
11.68
2.12
Key
D1
D2
D3
N-M
Item_5

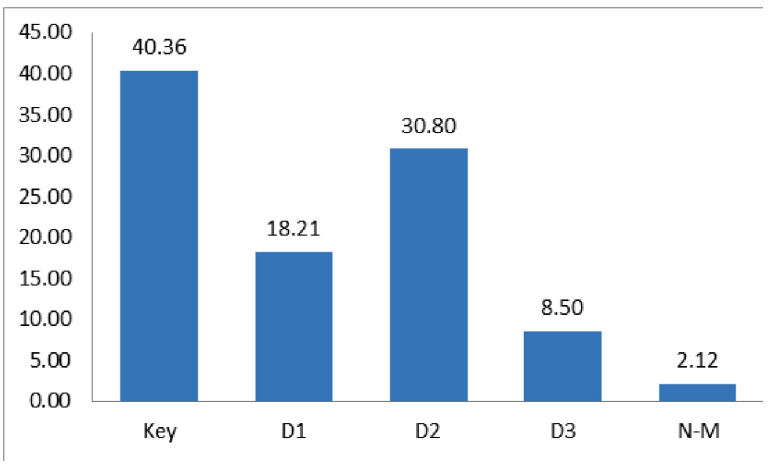

40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
36.57
11.99
32.93
16.08
2.43
Key
D1
D2
D3
N-M
Item_6

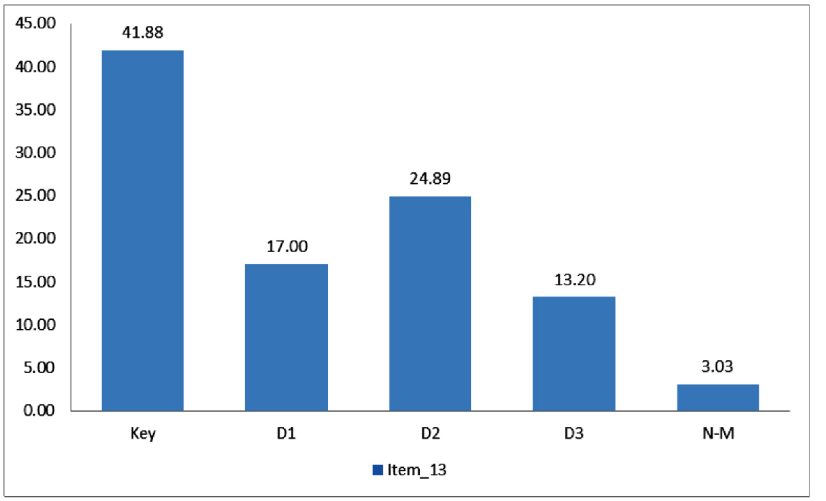

50.00
45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
43.85
15.02
19.42
18.06
3.64
Key
D1
D2
D3
N-M
Item_7

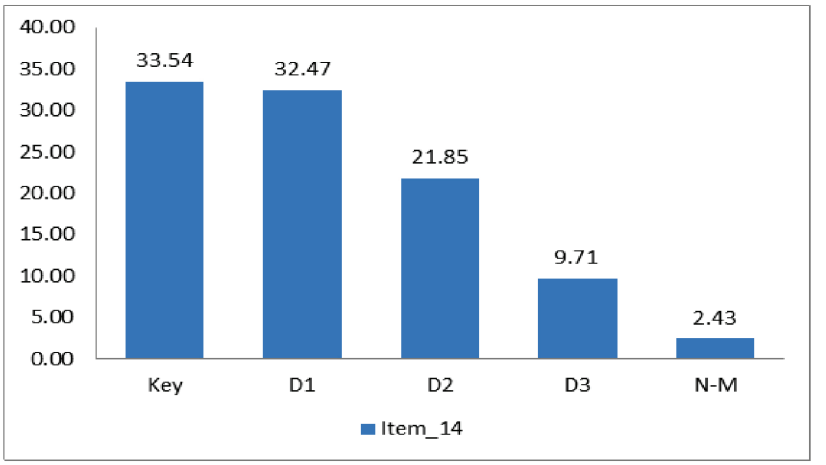

40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
30.20
14.57
36.27
16.08
2.88
Key
D1
D2
D3
N-M
Item_8

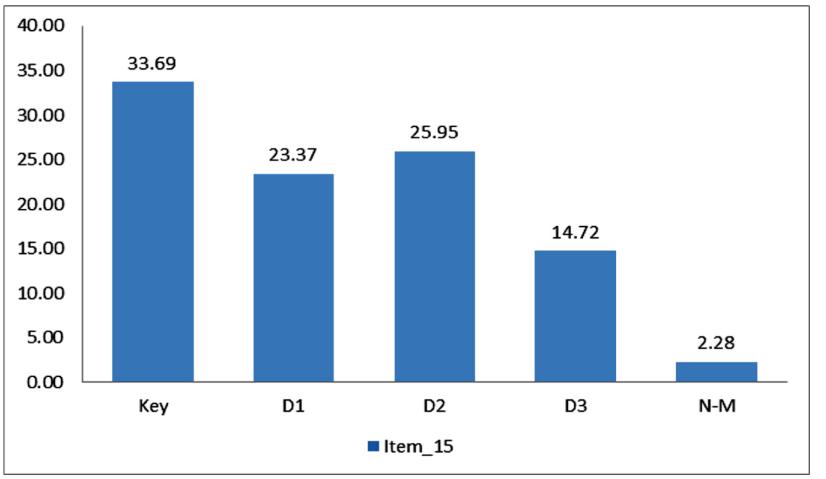

40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
37.78
24.13
14.26
21.40
2.43
Key
D1
D2
D3
N-M
Item_10

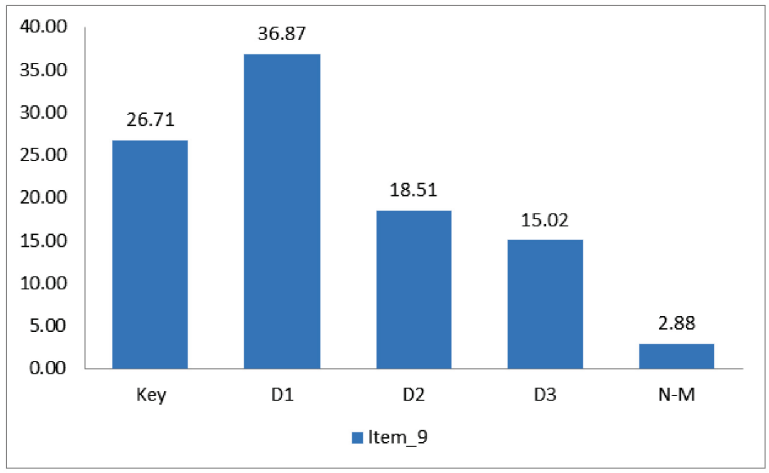

35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
30.20
24.13
30.50
12.75
2.43
Key
D1
D2
D3
N-M


45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
40.36
18.21
30.80
8.50
2.12
Key
D1
D2
D3
N-M


45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
41.88
17.00
24.89
13.20
3.03
Key
D1
D2
D3
N-M
Item_13


40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
33.54
32.47
21.85
9.71
2.43
Key
D1
D2
D3
N-M
Item_14


40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
33.69
23.37
25.95
14.72
2.28
Key
D1
D2
D3
N-M
Item_15



Note:
Key: Correct Answer
D1: Distractor 1
D2: Distractor 2
D3: Distractor 3
D4: Distractor 4
N-M: Students response either Null or Multiple

Figure 1: Percentage of Responses of Keys and Distractors

Note: For simplicity and easily understanding the position of the key is shifted to first in all the below graphs

CONCLUSION

Analysis and findings of this study imply that item analysis has an important role in determining the quality of assessment tools, especially regarding the multiple-choice questions (items). Well-constructed multiple-choice questions (items) are a good assessment tool to assess the level of cognition. Two item statistics i.e. item difficulty indices and discrimination indices can be used to judge the quality of the items since these two indices produce almost the same item characteristics.

The study was conducted on 659 students to assess their general awareness about their core subject Management at the terminal year in the business schools. The quality test items were used in this assessment exercise. Item analysis should be maintained in test development and evaluation, because of its importance in the investigation of reliability and in minimizing measurement errors.

It is observed that almost all items were in the range of $0.31 \leq p \leq 0.70$ (difficulty index) and two items were a little more difficult but none of the items had an easy level. The discrimination indices of almost all items had good levels and were easily discriminated the items. Only a few items required a minor revision to function properly.

Hence, the study has been able to establish that an individual item in a test with moderate difficulty and a good positive discrimination power is ideal for a good test. Item analysis results that are generated may be influenced by many factors which include examinees having a poor understanding of difficult topics, ambiguity in wordings of the items or even inappropriate key, instructional procedure applied, it may also be due to personal variations in students' intelligence level (Bichi, 2015).

The limitation of the study was the assessment test had only limited test items to assess the general awareness about Management subject in the students. This limited number of items is unable to cover all the topics of the particular level curriculum. Due to this limitation, this assessment test exercises is unable to analyze the effect of Bloom's level of questions effects on distractor efficiency and discriminatory index. The present findings are encouraging demonstrating that a good item will have a good discrimination index and difficulty index (Burud and Agarwal, 2019)..

Finally, it is concluded that item analysis is an important aspect of quality assurance and quality improvement of any tests. The indices of item analysis must be analyzed holistically. Revision of test items (multiple choice questions) must be based on the good difficulty index, discriminatory index, and functional distractors. Also, it is recommended that a suggested that training on test development and construction should be regularly organized for item developers and teachers to enhance more skills in test construction.

Reference

1. Burud I., Nagandla K. and Agarwal P. 2019 Impact of distractors in item analysis of multiple-choice questions. International Journal of Research in Medical Sciences. 7(4), 1136-1139.
2. Deepak, K.K., at. el. 2015. Psychometrics of Multiple Choice questions with Non-Functioning Distractors: Implication to Medical Education. Indian Journal Physiol Pharmacol, 59(4), 428 – 435.
3. Asamaah, D. and Ocansey M.K.K. 2019. International Journal of Research and Scientific Innovation. VI(9), 23 – 33.
4. Singh, A.K. 2009. Tests, Measurement and Research Methods in Behavioural Sciences. New Delhi: Bharati Bhawan.
5. Adegoke, B. A. 2013. Comparison of Item Statistics of Physics Achievement Test using Classical Test and Item Response Theory Frameworks. Journal of Education and Practice. 4(22).
6. Zubairi, A. M. and Kassim, N.L.A. 2006. Classical and Rasch analysis of dichotomously scored reading comprehension test items. Malaysian Journal of ELT Research. 2, 1-20.
7. Bichi, A.A. 2015. Item Analysis using a Derived Science Achievement Test Data. International Journal of Science and Research. 4(5).
8. Ebel, R.L. and Frisbie, D.A. 1991. Essentials of Educational Measurement. 5th Edn. Prentice-Hall, Englewood Cliffs, New Jersey.
9. Hingorjo, M.R. and Farhan J. 2012. Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index, and Distractor Efficiency. Journal of Pakistan Medical Association. 62(2), 142 – 147.