PSYCHOLOGY AND EDUCATION (2021) 58(5), ISSN 1553 - 6939
Article Received: 22th November,2020; Article Revised: 26th March, 2021; Article Accepted: 26th April, 2021

Effect Of Item Flaws On Difficulty And Discriminatory Indices Of Multiple Choice Questions In A Four-Year Medical Curriculum In Saudi Arabia

Fahad AlDhahri[1*], Aamir Omair[2], Mohi Eldin Magzoub[3]

[1]Pharmaceutical Care Services, King Abdulaziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia, College of Pharmacy, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[2]Department of Medical Education, College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia
[3]College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates
* Corresponding author, Email: faldhahri@yahoo.com

ABSTRACT:
Background
This study reviewed the multiple choice questions (MCQs) written at the College of Medicine, King Saud bin Abdulaziz University for Health Sciences (KSAU-HS) over the four years of its curriculum. It assessed the effect of item flaws on the Difficulty levels and Discriminatory indices of the MCQs.
Methods
All the MCQs used during the four years in all the blocks for the second batch of medical students at KSAU-HS were reviewed to identify the type of flaws and number of distractors that were functioning. The Difficulty levels and Discriminatory indices were obtained from the Assessment Unit, which were compared between the items with and without flaws using Independent samples t-test. Comparison of presence of flaws between different groups was done using the Chi Square test.
Results
The 1412 MCQs reviewed consisted of 938 (66%) recall and 474 (34%) reasoning type of questions, with a difficulty level of 0.69+0.24 and discriminatory index of 0.21+0.22. There were 535 (38%) MCQs in which all the other three options were functioning distractors. There was one non-functioning distractor in 449 (32%) and two or more non-functioning distractors in 428 (30%) MCQs. There were 287 (20%) MCQs which had flaws, with more than half of the flaws i.e. 152 (53%) being Negative statements in the flawed MCQs. The Difficulty level in questions with no flaw was 0.70+0.24, and 0.67+0.23 for questions with a flaw (p=0.06). The Discriminatory Index showed that 42% MCQs with flaws had a satisfactory discriminatory index of >0.3 as compared to 32% of questions without any identifiable flaw (p=0.02).
Conclusions
The overall difficulty and discriminatory indices were satisfactory for the reviewed MCQs. Item flaws were present in 20% of the MCQs with negative statements as the most common flaw. There was no significant difference in the difficulty level with regards to flaws, but questions with flaws had a better discrimination index as compared to those with no flaws.

Keywords:
Multiple choice questions, Difficulty level, Discriminatory index, item flaws, non-functioning distractors

INTRODUCTION

BACKGROUND

Assessment is an important component in the academic process. It is important not only for the examination but is also useful for assessing the students quality of learning.[1] It is considered as one of the major "drivers" of the teaching-learning process.[2] Multiple Choice Questions (MCQs) comprise the most common assessment methods in medical education.[3] One of the strengths of MCQs is that they can efficiently assess a broader range of objectives. The grading process is also faster through the use of computerized checking of the MCQ answer sheets.[4] The computerized analysis also identifies the problem questions and is useful for establishing a MCQ bank for the good quality questions, which can be used in future examinations. [5] One major limitation of writing MCQs is the selection of suitable distractors as the other options, especially for higher cognitive level questions. Another disadvantage is that students might guess the correct answer even with plausible distractors. To overcome these drawbacks the MCQs need to undergo a quality process review with regards to their reliability and validity.[5]

The MCQs can be classified in two ways: True/False type and selecting one best answer from a list of options.[6] The One Best Answer (OBAQ) type of

questions is more popular worldwide. OBAQs consist of a stem, lead-in and a usually three or four other options.[7] The stem is a short problem that is most appropriately related to one of the options provided. The lead-in should be in a form of question related to the stem. The options should be all related to the problem and must be similar in nature to the correct option. Phrases and words like may, could, usually or frequently should be avoided.[6] It has been seen that flawed MCQs interfere with the assessment and can affect the grading of the exam.[8] The faculty should consult general guidelines when constructing MCQs to ensure the quality of test questions. All well-constructed questions will allow the student to form an answer without consulting the options.[6] The common flaws that have been studied in the medical literature are listed in Table 1.[9,10]

An item analysis provides evaluation of an item's difficulty level and its discrimination index. The difficulty level is determined by "calculating the proportion of examinees that answer the item correctly". The discrimination index is calculated by "comparing the proportion of correct answers in the upper 27% with the lower 27% of the test-takers".[11] An item analysis review is conducted to identify MCQs for retention, modification or removal. Another aspect that needs to be considered is the cognitive level of the questions based on Bloom's taxonomy, which classifies questions as Level I (Recall) and Level II (Understanding).[12] The difficulty and discrimination index need to be determined based on the above two levels to see if questions that require levels of thinking (Level II) are more difficult or how well they discriminate. This would be helpful in identifying which type of questions need more improvement in preparation.

There are studies which have assessed the association between difficulty and discriminatory indices, as well as their relationship with non-functioning distractors. But, few studies have assessed the effect of item flaws on the above two psychometric properties of the MCQs. The purpose of this study was to assess the effect of item flaws on the difficulty and discriminatory indices of MCQs prepared at the College of Medicine, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia.

METHODS

This was a cross-sectional study conducted at the College of Medicine (COM) at King Saud bin Abdulaziz University for Health Sciences (KSAU-HS) in Riyadh, Saudi Arabia. The study included review of the MCQs prepared for the first four batches of the COM to compare the effect of item flaws on the difficulty and discriminatory indices of the MCQs. The MCQs from the final exams of all the 14 courses conducted for the second batch of the COM were included in the review. Also MCQs for the Neurosciences block (which is the first block of the medicine program) in Basic Sciences were compared across the first four batches to determine if there was any improvement in the quality of MCQs over the four years. The exams conducted for all the courses consist of one best choice MCQs and the students are required to answer 80 MCQs in 160 minutes (two minutes for each MCQ).

All end-of-block MCQs exams that were administered in all nine blocks in Phase II (Basic Sciences) and five blocks in Phase III (Clinical Sciences) over the four year period were reviewed for the second batch of COM. Also all final exams for batches 1 to 4 Neuroscience block were reviewed. All the questions in the respective examinations were included in the retrospective review.

Data collection was started after receiving approval from the COM-Research Committee and from the Institutional Review Board of King Abdullah International Medical Research Center. After receiving permission from the Assessment Unit, all the available MCQs were reviewed for the above courses. The identification of MCQ item flaws was done by the primary author. In this review common item flaws[9,10] were assessed using the list given in Table 1. The printouts of the MCQs were reviewed in the Assessment Unit and the type of item flaw(s) was identified and noted in the same sheet with the psychometric indices. Ten percent of the questions were reviewed by a second investigator to validate the findings.

Psychometric properties of all the MCQs including item difficulty and item discrimination were available from the Assessment Unit. The Difficulty and Discrimination indices were grouped into categories as used by the College of Medicine at KSAU-HS (Table 2)[11]. The Distractor analysis for the non-functioning distractors (i.e. how many of the other options in the MCQ were selected by less than 5% of the examinees[13]) was identified from the list provided by the Assessment Unit. The items were classified as having zero, one, two, or three non-functioning distractors (out of total four options). This means that if an item had '0' non-functioning distractors then all of the other three incorrect options were selected by more than 5% of the examinees. Or if an item had '3' non-functioning distractors then none of the incorrect options was selected by more than 5% of the examinees.[13]

Data was coded and entered in SPSS v20. Descriptive analysis for different item flaws was reported as mean $\pm$ standard deviation for the Difficulty and Discriminatory indices. The categorical variables like type of items flaws and function of distracter were presented as frequency and percentages. The comparison of type of flaws by Basic (first two years) and Clinical Sciences exams (3rd and 4th year) and

between different batches was done using Chi Square test. The comparison of Difficulty index by type of questions (Recall vs Reasoning) or presence / absence of item flaws was done using the Independent Samples t-test. A p-value <0.05 was considered to show a statistically significant difference.

RESULTS

A total of 1412 MCQs were reviewed over the last four years from the end of block exams conducted for Batch 2 of the Medical program as well as the Neurosciences block of the first four batches of students of the medical college (Batches 1, 2, 3 and 4) for comparison over the years. The classification of the questions including the types of distractors and flaws is shown in Table 3. The majority of the questions i.e. 938 (66%) were of the recall type and 474 (34%) were reasoning type of questions. There were "Zero" non-functioning distractors in 535 (38%) of the MCQs, one non-functioning distractor in another 449 (32%) and two or more non-functioning distractors in 428 (30%) of the total 1412 MCQs. The review of the questions for item flaws showed that there were a total of 1135 (80%) questions without any flaws; there were a total of 298 flaws in the 277 (20%) questions with flaws as some question had more than one type of flaw. The main types of flaws that were identified were Negative statements (11%), Convergence (5%), Long correct answer (3%) and Word repeats (1%) while 2% were other, less frequent flaws (e.g. Grammatical cues, None of the above, All of the above cues, vague terms in the options, typing error, and numeric data not stated consistently). The negative statement flaw constituted more than half i.e. 152 (53%) of the total 287 items with flaws.

The MCQs reviewed included 1012 MCQs for batch 2 for all the 14 blocks over their four years. There were 612 MCQs in the nine blocks of Basic Sciences (Phase II) and 400 MCQs in the five blocks of the Clinical Sciences (Phase III). There was an improvement in the quality of MCQs between the two phases as shown in Table 4. Phase III had 373 (93%) questions without any flaws as compared to 439 (72%) in phase II (p<0.001). The significant improvements in the type of flaws was for the 'negative statements' which were 0.3% in phase III as compared to 18% in phase II (p<0.001) and in the 'long correct answer' which was 0.3% in phase III as compared to 5% in phase II (p<0.001). The presence of item flaws was compared across the years for the Neurosciences (NS) block in all of the four batches. There were 80 questions in the final exam of the NS block and it was found that the fourth batch had the lowest number of items with flaws i.e. 15 (19%) as compared to 22 (28%) in batch one, 24 (29%) in batch two, and 32 (40%) in batch three (p=0.04).

The comparison of the difficulty index by type of questions and presence of flaws is shown in Table 5. The recall type questions were found to be more difficult with a Difficulty level of 0.67 ± 0.24, while it was 0.73 ± 0.22 for Reasoning questions (p<0.001). There was no significant difference in the Difficulty level of questions with no flaws (0.70 ± 0.24), as compared to 0.67 ± 0.23 for questions with at least one flaw (p=0.06). With regards to the comparison of the difficulty level between the different types of flaws, it was found that all the common flaws (negative statement, convergence, long correct answer, and word repeats) had similar difficulty indices ranging from 0.63 to 0.69. The only difference found was that MCQs with 'Other' type of flaws were found to be more difficult with a difficulty level of 0.59 ± 0.29 (p=0.04) as shown in Figure 1. The 'Other' type of flaws were less common i.e. 21 (7%) of the 287 items with flaws.

Table 6 shows the analysis for the Discriminatory Index. There was a significant association between the presence of flaws and the discrimination index category (p=0.02). A greater proportion (42%) of questions with flaws had a discriminatory index of 0.3 or higher as compared to questions without any identifiable flaws (32%).

DISCUSSION

The current study assessed the quality of MCQs with regard to psychometric properties and evaluation of items. There were 20% questions with some flaw in them and the greatest number of flaws was of the 'Negative statements' type. These negative statements were significantly decreased in the Clinical Sciences exams of Phase III as compared to the Basic Sciences exams of Phase II. It was also found that the percentage of items with flaws was the lowest for the fourth batch as compared to the first three batches for the Neurosciences block. The study also revealed that recall questions were more difficult as compared to the reasoning type of questions, but there was no significant association of the difficulty index with the presence of item flaws. The discrimination index on the other hand showed that high performing students answered significantly better in items with flaws as compared to the weaker students.

In a study by Tarrent & Ware[14] at an English-language university in Hong Kong, the findings showed that there were 47% flawed items on ten test papers as compared to 20% in this study. In the Hong Kong study the effect of the item flaws seemed to facilitate the students in passing. It was observed that fewer examinees passed the standard scale after removing flawed questions than the total scale. Our study did not identify a major difference in the difficulty level between the items with and without flaws. The most common flaws in our study were negative statement, convergence strategy, long correct answer and word repeats. In another study by Tarrent et al[15] the most common flaws were "ambiguous or unclear information, negative worded stem,

implausible distractors, more than one or no correct answer, longest option is correct and word repeats in the stem".

Tarrent et al[13] also assessed the functioning distractors in MCQs at the same university in Hong Kong. They found the 47% of the MCQs had two or more non-functioning distractors (NFDs). Our study found that there were two or more NFDs in 30% of the total MCQs. In a study from Bahrain from Pediatric MCQs it was reported that 16% of the questions had two or more NFDs[16], while another study from India on Physiology MCQs reported 12.5% of the 40 questions as having 2 or more NFDs[17]. The Hong Kong study also showed that items with "two or more functioning distractors were more difficult and more discriminating".[13]

The comparison of flaws between the Basic Sciences and Clinical exams showed that there were fewer flaws in the Clinical exams as compared to the Basic Sciences exams. The main difference was in the negative statements and long correct answers being less in the clinical exam. This may be due to more effective review of the exam questions in the later years, as certain interventions for improvement of the MCQs quality were carried out during this period. These included faculty development, central control of assessment and continuous review of items by the assessment committee in the College. The negative statements and long correct answer flaws were most reduced between the Basic and Clinical Sciences exams. This may be due to these two type of flaws being easier to recognize and correct.[18]

There was a borderline significance in the difficulty level of the questions with regards to the presence of flaws. Items with flaws were found to be slightly more difficult as compared to those without flaws. This is similar to the study by Tarrant & Ware[14] on nursing assessment which found that the difficulty level of items with flaws had a range from 10% more difficult to 8% less difficult as compared to those without flaws. This may be explained by the fact that good students attempt items with flaws in a better way i.e. they are 'test-wise students'.[19]

Our study also found a significant relationship between the type of question and difficulty level. The lower cognitive level items (recall type) were found to be more difficult as compared to the higher cognitive level items (reasoning type). This may be explained by the fact that the students are more able to answer clinical reasoning questions due to the effect of Problem Based Learning (PBL) curriculum and their clinical judgment. In another study from the United States by Nedeau-Cayo et al,[20] it was found that items with lower cognitive level had more common item flaws as compared to the higher cognitive levels. In the current study it was observed that items with flaws were more discriminatory as compared to items without flaws. A study on medical students from India reported that faulty items were more likely to have a higher discrimination index[21].

This study adds to the previous studies in that it relates the effect of item flaws on the difficulty and discriminatory levels of the MCQs in the medical examination. The main limitation of this study was that only two reviewers assessed the questions bank, so some flaws may have been missed. Also there were a small number of students in each class which ranged between 20 and 37 students. This may affect the reliability of the tests in addition to the difficulty and discriminatory indices.

CONCLUSION

The improvement of quality of questions from the pre-clinical to clinical exams with regard to flaws may be attributed to more effective review of the question bank. Questions with some flaw in them had a better discrimination index which may indicate that, students who are in the upper 27 percentile are able to attempt flawed item in a better manner as compared to students in the lower 27 percentile. Some recommendations for any institution to have good quality of MCQs include central control of assessment, continuous review of items, faculty development activities, shared item bank, alignment of curricular activity to assessment and use MCQs that test higher cognitive orders (clinical reasoning).

LIST OF ABBREVIATIONS

COM: College of Medicine

KSAU-HS: King Saud bin Abdulaziz University for Health Sciences

MCQ: Multiple choice questions

NFD: Non-Functioning Distractors

NS: Neurosciences

OBAQ: One Best Answer Question

PBL: Problem Based Learning

Declarations

Ethics approval

The study did not involve any human subjects so consent to participate was not required.

The approval for the study was obtained from the King Abdullah International Medical Research Center before the MCQ data was obtained from the Assessment Unit of the College of Medicine, King Saud bin Abdulaziz University for Health Sciences

Consent for publication

Not applicable

Availability of data and materials

The data extraction sheets for identifying the flaws and recording the difficulty and discrimination index are available with the corresponding author on reasonable request. They can be obtained by sending an email to faldhahri@yahoo.com

Competing interests
The authors declare that they have no competing interests

Funding
There was no funding required for this research

Authors' contributions
FA was responsible for all aspects of the research process including the following: preparation of the research proposal, literature review, data collection, data entry, writing of the manuscript and final submission to the journal

AO was responsible for data analysis, writing the results section, reviewing and editing the manuscript for final publication

MEM was responsible for overseeing the research from the concept to the publication, he was involved in proposal development, development of the data collection instrument, final approval of the manuscript

ACKNOWLEDGEMENT
The authors thank Dr. Michael Seefeldt, Imran Zafar, Annabelle Borja, Clara Hernandez and Roseminda Asensi for their help and support.

REFERENCES

1. Fisher MR. Student assessment in teaching and learning. [Online]. Vanderbilt University 2020. Available from: https://cft.vanderbilt.edu/student-assessment-in-teaching-and-learning/ Accessed 11 Aug 2020
2. Conley DT, Darling-Hammond L. Creating systems of assessment for deeper learning. 2013.Stanford, CA: Stanford Center for Opportunity Policy in Education.p.18. Available from: https://edpolicy.stanford.edu/sites/default/files/publications/creating-systems-assessment-deeper-learning_0.pdf Accessed 11 Aug 2020
3. Pham H, Trigg M, Wu S, O'Connell A, Harry C, Barnard J, Devitt P. Choosing medical assessments: Does the multiple-choice question make the grade? Educ Health 2018; 31:65-71
4. Fellenz MR. Using assessment to support higher level learning: The multiple choice item development assignment. Assess Eval High Educ 2004;29(6): 703-19
5. Tangianu F, Mazzone A, Berti F, Pinna G, Bortolotti I, Colombo F, et al. Are multiple-choice questions a good tool for the assessment of clinical competence in internal medicine? Italian J Med 2018; 12(2):88-96
6. Paniagua MA, Swygert KA. Constructing written test questions for the basic and clinical sciences. Philadelphia, PA: National Board of Medical Examiners, 2016. Available from: https://www.unmc.edu/facdev/_documents/ConstructingWrittenTestQuestions_WritingManual.pdf. Accessed 11 Aug 2020
7. Linn RL. Measurement and assessment in teaching. 9th ed. Old Tappan: Pearson Prentice Hall; 2005
8. Salam A, Yousuf R, Abu Bakar SM. Multiple choice questions in medical education: how to construct high quality questions. Int J Human Health Sc 2020; 4(2): 79-88. doi:http://dx.doi.org/10.31344/ijhhs.v4i2.180
9. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Meas Educ 2002; 15(3): 309-34
10. Gronlund NE. Assessment of student achievement. 8th ed. Boston: Allyn and Bacon; 2006.
11. Zafar I. Item analysis assumptions (Difficulty & Discrimination indexes). Fact Sheet No. 24A. King Saud bin Abdulaziz University for Health Sciences. [Online].2008 Available from: https://com.ksau-hs.edu.sa/images/stories/documents/FactSheet24A_and_B.pdf Accessed 6 Aug 2020
12. Agarwal PK. Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? J Educ Psychol 2019; 111(2): 189-209. doi:https://psycnet.apa.org/doi/10.1037/edu0000282
13. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Med Educ 2009; 9: 40. doi: https://doi.org/10.1186/1472-6920-9-40
14. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ. 2008; 42(2):198-206
15. Tarrant M, Knierim A, Hayes SK, Ware J. The frequency of item-writing flaws in multiple-choice questions used in high-stakes nursing assessments. Nurse Educ Today 2006; 26:662–71
16. Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item analysis of multiple choice questions at the Department of Paediatrics, Arabian Gulf

University, Manama, Bahrain. Qaboos Univ Med J 2018; 18(1):e68-74. doi: 10.18295/squmj.2018.18.01.011

17. Kolte V. Item analysis of multiple choice questions in Physiology examination. Indian J Basic App Med Res 2015; 4(4):320-6
18. Shank P. How to Fix 3 Common Flaws in Multiple-Choice Questions. [Online]. 2019. Available from: https://learningsolutionsmag.com/articles/how-to-fix-3-common-flaws-in-multiple-choice-questions. Accessed 19 Sept 2020.
19. Ibbett NL, Wheldon BJ. The incidence of clueing in multiple choice testbank questions in accounting: Some evidence from Australia. e-Journal Business Educ Scholarship Teach 2016; 10(1): 20-35.
20. Nedeau-Cayo R, Laughlin D, Rus L, Hall J. Assessment of item-writing flaws in multiple-choice questions. J Nurses Prof Dev 2013;29(2):52-7. doi: 10.1097/NND.0b013e318286c2f1
21. Chandrika R, Krishan PHL, Sajitha K, Harish P, Jayaprakash S. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. Int J Educ Psychol Res 2016; 2(4): 201-4

Table 1: Types of Flaws in MCQs[9,10]

1. Grammatical cues
2. Absolute terms - terms such as “always” or “never”
3. Long correct answer
4. Word repeats - a word or phrase is included in the stem and in the correct answer
5. Convergence strategy - the correct answer includes the most elements in common with the other options
6. None of the above and All of the above Cues
7. Numeric data are not stated consistently
8. Terms in the options are vague (e.g., “rarely,” “usually”)
9. Stems are tricky or unnecessarily complicated
10. Negative statement (e.g. “NOT”, “Except”, Which is not correct”)
11. Typing error

Table 2: Classification of the Difficulty and Discriminatory Indices[11]

Index Range	Inference to question
Interpretation of Difficulty index:	
0.85-1.00	Very Easy
0.70-0.84	Easy
0.30- 0.69	Optimum
0.15-0.29	Hard
0.00-0.14	Very Hard
Interpretation of Discrimination index:	
Below zero	Negative
0-0.19	Poor
0.20-0.29	Dubious
0.30- 1.00	Good

Table 3: Types of Questions, Distractors, and Flaws in the total items reviewed (N=1412)

	n	**%**
Type of Question		
Recall	938	66%
Reasoning	474	34%
Distractors		

Zero non-functioning Distractors	535	38%
One non-functioning Distractors	449	32%
Two non-functioning Distractors	270	19%
Three non-functioning Distractors	158	11%
Type of Flaw		
No flaws	1135	80%
Negative statement[a]	152	11%
Convergence strategy[b]	73	5%
Long correct answer	34	3%
Word repeats[c]	18	1%
Others[d]:	21	2%

[a] e.g. NOT, Except, [a] which i[a] which is not correct

[b] the correct answer includes the most elements in common with the other options

[c] a word or phrase is included in the stem and in the correct answer

[d] includes: Grammatical cues, None of the above and All of the above, Terms in the options are vague (e.g. "rarely," "usually"), Typing error, Numeric data are not stated consistently

Table 4:

Comparison of flaws between Phase II and III for Batch 2 over the four years (N=1012)

	Phase II (Basic Sciences) (n=612) n (%)	**Phase III (Clinical Sciences) (n=400) n (%)**	**p-value**
No flaws	439 (72%)	373 (93%)	<0.001*
Negative statement[a]	109 (18%)	1 (0.3%)	<0.001*
Convergence strategy[b]	31 (5%)	15 (4%)	0.33
Long correct answer	28 (5%)	1 (0.3%)	<0.001*
Word repeats[c]	13 (2%)	5 (1%)	0.31
Others[d]	9 (1.5%)	6 (1.5%)	0.96

*

Significant at p<0.05

[a] e.g. NOT, Except, Which is not correct

[b] the correct answer includes the most elements in common with the other options

[c] a word or phrase is included in the stem and in the correct answer

[d] includes: Grammatical cues, None of the above and All of the above, Terms in the options are vague (e.g. "rarely," "usually"), Typing error, Numeric data are not stated consistently

Table 5: Comparison of Difficulty index by Type of Question and by Presence of Item Flaw (N=1412)

	N	Mean ± Sd	p-value
Type of Question			
Recall	938	0.67 ±0.24	**<0.001***
Reasoning	474	0.73 ±0.22	
Flaws			
No	1135	0.70 ±0.24	0.06

Yes	277	0.67 ±0.23	

* Significant at p<0.05

Table 6: Comparison of Discriminatory index by Presence of Flaws (N=1412)

Flaws	**Discrimination Index Categories**			
	Negative (< 0)	Poor (0-0.19)	Dubious (0.20-0.29)	Okay (0.3-1.0)
No (n=1135)	115 (10%)	469 (41%)	185 (16%)	**366 (32%)**
Yes (n= 277)	19 (7%)	103 (37%)	39 (14%)	**116 (42%)**
	p-value: 0.02*			

* Significant at p<0.05

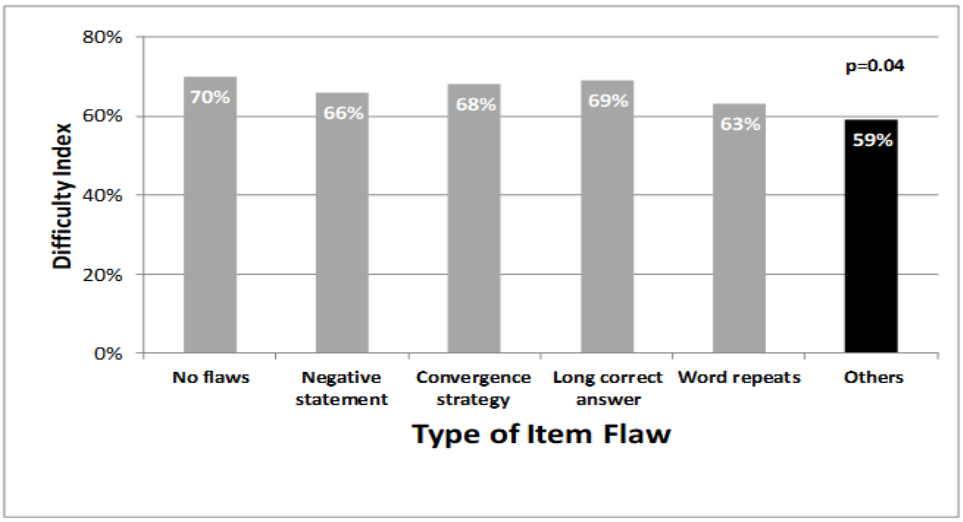

Fig 1. Comparison of Difficulty Index by the types of item flaws