

## Item Analysis of Multiple Choice Questions based Undergraduate Assessment in Community Medicine

Parag Chavda<sup>1</sup>, Shobha Misra<sup>2</sup>, Bithika Duttaroy<sup>3</sup>

### Abstract

**Background:** Multiple Choice Question (MCQ) based assessments are ubiquitous now, yet the emphasis on training of teachers in preparing good MCQs is inadequate. There are very few studies reporting use of item analysis of MCQs in medical education from India. Hence, we present this simple psychometric analysis of MCQs.

**Methods:** An item analysis was performed on an internal assessment test paper with 50 MCQs faced by 158 final year undergraduate students (6th and 7th semester) in Community Medicine during the year 2011. The following indicators of item analysis were derived; 1. Response rate for each item, 2. Facility Value (FV), 3. Discrimination index (DI), 4. Distractor efficiency and 5. Cronbach's alpha( $\alpha$ ). Microsoft Office Excel was used for analysis.

**Results & Discussion:** The mean score ( $\pm 2$  SD) of the students was 27.2/50 (15.8 – 38.6), and 54% of the items had a response rate of more than 95%. Out of 50 items on the test, 19 and 22 items were in the acceptable range of FV and DI respectively. Only 12 items were in acceptable range of both. 36% distractors achieved a below optimal response rate (<5%). Cronbach's alpha was 0.66.

**Conclusion:** The analysis helped us detect the technical flaws in the single paper scrutinized. Reporting of matrix of the items falling in different ranges of FV and DI was found to be more informative, as it would help improve the quality of the items in our bank.

**Key Words:** Facility Value, Discrimination index, Distractor efficiency, Cronbach's alpha

### Introduction

Multiple Choice Question (MCQ) based assessments are ubiquitous now, yet not enough stress is laid on the training of teachers on preparing good MCQs. There are very few studies from India (Shah *et al.*, 2011, Sarin *et al.*, 1998) reporting use of item analysis of MCQs in medical education.

Hence, we present this simple psychometric analysis tool used for MCQs which can help improve the quality of MCQ-based assessments.

<sup>1</sup>Assistant Professor, Department of Community Medicine, GMERS Medical College, Vadodara, India

<sup>2</sup>Associate Professor, Department of Preventive & Social Medicine, Medical College, Vadodara, India

<sup>3</sup>Professor & Head, Department of Microbiology, GMERS Medical College, Vadodara, India

#### Corresponding Author:

Dr. Parag Chavda, Assistant Professor, Department of Community Medicine, GMERS Medical College, Gotri, Vadodara -390021, India  
Email: paragchavda@gmail.com

### Methods

This item analysis was performed on an internal assessment test paper containing 50 MCQs faced by 158 final year undergraduate students in Community Medicine in 2011. The paper included single best response type MCQs, with four choices.

Based on their obtained scores, the students were categorized into the high achiever group (top 33%) and the low achiever group (bottom 33%) in order to calculate facility value (FV) and discrimination index (DI). FV and DI were calculated for each item as per the following formula (Singh & Anshu, 2012):

$$FV = [(H+L)/T] \times 100$$

$$DI = [(H-L)/T] \times 2$$

Where, H and L are number of correct responses in the high and low achiever groups respectively, and T is the total number of responses in both the groups.

We also calculated response rate to each item and each distractor. Cronbach's alpha was calculated as a measure of reliability of the

test. Microsoft Office Excel was used for analysis.

## Results & Discussion

The mean score ( $\pm 2$  SD) of the students was 27.2/50 (15.8 – 38.6) for this test paper involving 50 MCQs.

Analysis showed that 54% of the items had a response rate of more than 95%. Of the rest, 38% had a response rate between 81–94%. Overall, the response rate to the items was good.

**Facility Value (FV):** FV (theoretical range: 0-100, acceptable range: 30-70, higher the value, the easier the item) measures the difficulty of an individual item on the test paper. In our test the mean FV was calculated to be 63 which was very near to the desirable. However, only 19/ 50 were in the acceptable range of FV 30-70. 22 items had FV of  $> 70$  and nine items had an FV of  $< 30$ . (Table 1) Thus, easy items outnumbered the difficult ones in our test.

**Table 1: Distribution of items over the range of Facility Value and Discrimination Index**

		Facility Value			Total
		< 30	30 – 70	> 70	
Discrimination Index	> 0.2	1	12	12	25
	0.15 – 0.2	1	3	6	10
	0 – 0.14	4	4	4	12
	Negative	3	-	-	3
Total		9	19	22	50

**Discrimination Index (DI):** The theoretical range of DI is from negative to 1.0, with  $>0.35$  being acceptable: and higher the DI more differentiating is the item. (Singh & Anshu, 2012) DI is a measure of how well an individual item can differentiate between the high and low performing students (top and bottom 33%) The mean DI in our test was calculated to be 0.2. The detailed analysis (Table 1) shows that there were three items in our test that had a negative DI. A total of 22 items had DI of  $< 0.2$  and these items need to be revised. The remaining 25 items were considered in the acceptable range.

Preparing a table of distribution of items over the ranges of both FV and DI makes more sense than analyzing the isolated figures of FV and DI. Item analysis in this manner (Table 1) indicated that only 12 out of 50 items were such that they satisfied the acceptable ranges of both FV and DI. All others needed to be revised either due to low FV or DI. As identified by the researchers, the reasons for items that had DI  $< 0.2$  were: too easy question (16), too difficult question (7), confusing wording (1) and several correct answers (1). Thus large numbers of easy items were lowering the DI in this test.

Thus abiding strictly by ideal range of DI, our test fares poorly. Yet, it is important to understand that although very difficult or easy

items will have low ability to discriminate, such items are often needed to adequately sample course content and objectives. Further, an item may show low discrimination if the test covers a wide range of content areas at different taxonomic levels of cognitive skills. (Mehrens & Lehmann, 1983) This is very much a case with Community Medicine curriculum.

**Distractor Efficiency:** Distractors used in the items need to be plausible so that they attract at least 5% students. The mean distractor response rate in this test was 11% and as many as 36% (54/150) distractors achieved a response rate of  $< 5\%$ . During the attempt to modify these distractors it was observed that even after much effort, creating three plausible distractors was found to be difficult for some items. Hence we were left with no choice but to offer as plausible distractors as possible.

**Reliability Coefficient:** Cronbach's alpha (range 0 to +1) was calculated as a measure of reliability of the test. Whereas  $\alpha \geq 0.8$  is considered good, it was found to be 0.66 for this test. In our test though the length was fairly good with 50 items, low DI of many items and the diverse subject matter has probably lead to low alpha.

## Conclusion

Traditionally considered as tedious and time consuming task, the use of software applications such as Microsoft Excel has made it relatively easy to calculate various item and test statistics. As many as half of the items with a DI of  $<0.2$  helped us identify the reasons for same and correct them. The distractor response rate also directed us towards making necessary corrections to improve the quality of distractors. We expect that with use of improved DI the reliability coefficient will also improve in future tests. Reporting of matrix of the items falling in different ranges of FV and DI is more informative than merely providing mean FV and DI values as has been done in previous published articles.

## References

- Mehrens, W & Lehmann, I (1983) *Measurement and Evaluation in Education and Psychology*, 3rd edn. New York: Holt, Rinehart & Winston.
- Sarin, Y.K., Khurana, M., Natu, M.V., Abraham G. Thomas & Tejinder Singh (1998) Item Analysis of Published MCQs, *Indian Paediatrics*, 35, 11, pp. 1103-1105.
- Shah, C.J., Baxi, S.N., Parmar, R.D., Parmar, D. & Tripathi, C.B. (2011) Item Analysis of MCQ from Presently Available MCQ Books, *Indian Journal for the practicing doctor*, 6, 1, pp. 26-30.
- Singh, T. & Anshu (2012) *Principles of Assessment in Medical Education*, New Delhi: Jaypee Brothers Medical Publishers.