Islamic Foundations of Education: Item Analysis of a Teacher Made Test at Master Level

* Zunaira Fatima
** Muhammad Ijaz Latif

Abstract:

This study is aimed at constructing and analyzing a researcher made Multiple Choice Test in the subject of Islamic foundations of Education. Sixty items were finalized and administered on the selected sample which consisted of the students of Education at Masters' level in the four HEC recognized public sector universities in Pakistan i.e., Bahauddin Zakariya University Multan, The Islamia University of Bahawalpur, University of Sargodha and University of Education (Multan and Vehari campus). After pilot testing, test was administered on 385 male and female students selected through multistage sampling technique. Item analysis was performed by using classical test theory. According to discrimination and difficulty indices 8 items in Form A and 6 items in form B needed modifications because the low and high achiever students were unable to attempt those items. This study recommends that a larger size of sample should be included to attain higher reliability of the items.

Key Words: Traditional Item Analysis, Discrimination index, Difficulty index, Educational Assessment Test, Islamic Foundations.

Introduction and Review of Literature

The foremost aim of education is to provide harmonious development of body, mind and soul in a way that an individual is able to achieve what is best in life. The concept of learning is mainly motivated at providing relevant security. It is more concerned with attitude and understanding then with mere teaching. It exalts people for their motives and also judges them by their actions.[1]

* Lecturer, Department of Education, University of Sargodha, Sargodha.
** Professor and Chairman, Department of International Relations, The Islamia University of Bahawalpur.

[1] L. Alkanderi (2001*). Exploring Education In Islam: Al-Ghazali's Model Of The Master-Pupil Relationship Applied To Educational Relationships Within The Islamic Family,* PhD Thesis,

Islam has put greater emphasis on the importance of acquisition and dissemination of knowledge (*ilm*) concerning human actions and activities. Islam has mentioned it as compulsory (*farz*) upon its adherents, regardless of gender, to learn and disseminate knowledge. The obligation of seeking out knowledge is binding upon every Muslim by the command of the Qur'an and Sunnah of the Holy Prophet (PBUH). Education from Islamic perspective is often defined by Muslim scholars in three different dimensions which are reflected in different concepts introduced, important among them are; *tarbiyyah* – the process of education that gives emphasis on physical and intellectual development of an individual; *ta'dīb* – the process of education that gives emphasis on nurturing good human being with noble codes of conduct and ethics approved by Islam, so that he might conduct and position himself in society with justice; and *talīm* – the process of education that is based on teaching and learning.[2]

In today's system of education all the three processes have always been examined through a justified and comprehensive process of testing and examination. In the past, renowned Muslim scholars used to award 'San'ad' or 'Dastar' on achieving some merit. Following the same past tradition, today's modern system of education is trying to develop a justified system of examination to judge the current achievement of their students. This teacher made achievement test is also an attempt to contribute towards the process of testing and analyzing, to check; how much our students are aware of their basic and fundamentals of Islamic principles and the history of historians and their philosophy of education.

Chronology of Educational Testing

In order to evaluate capabilities of individuals various types of tests have been developed for different disciplines. In some cases these tests are to educate men and are of different types and for different purposes. Usually they are for evaluating different capacities and learning skills. Teachers are the right persons who are responsible to examine their students. So they have to be sure about the subject matter, the type of questioning, the exact answer (objective type) or the most relevant response (essay type) for the question and at the end the use of the test scores. This present study is aimed to construct and analyze an achievement test in the subject area of Islamic foundations of education.

Tests have ancient origins like many other elements of education. It is interesting to review this early history and trace the development of ideas about what

---

Pennsylvania State University retrieved from http://www.latefah.net/artic3%5cexploring_education_in_islam.pdf on December 15, 2015.

[2] F. B. Yasin and S. M. Jani. (2013) *Islamic Education, Philosophy, and Features*, International Journal of Education and Research, 1(10).

constitutes good practice. History of testing spans more than 4,000 years and have undergone many changes during this time. Numerous "new" testing procedures currently being advocated have already been tried several times in the past, and problems associated with them are well known, although some advocates of "new" practices choose to ignore them.[3]

Etymology of Testing

Testing plays an important role in today's educational system. The results of a test often serve as primary source of making public perception regarding the quality of assessments and tests. These tests are being used for multiple purposes, as a primary tool for the teachers, administrators and the policy makers. It might help in two different ways; firstly, to improve the quality of teaching and learning, secondly, to assess the programs and schools. Linn and Millar[4] describe that testing is a process of putting the subjects to experimental test in order to determine the individual's progress. According to Frizbie,[5] testing is a tool of measurement or might be considered as a standardized procedure to determine a kind of behavior to be measured. Similarly Raza[6] quotes the definition from the Glossary of Testing and Education (2003) that testing is an activity of measurement, to determine a person's behavior, competence, knowledge and skills in a given area of disciplines. Furthermore tests and assessments are widely in use to evaluate the system of education, to improve the instructional process, to measure the academic achievement of the students and to assess student's mastery of knowledge and skills.

In brief, test is a tool or a process that purposes a sequence of tasks to which students are to respond.[7] Test is a tool of measurement that evaluates and describes the scores in numerical form of the data; the purpose is to evaluate the extent of learning and teaching. In educational setting tests mostly contain a set of items which intend to measure a domain of knowledge and skills.[8] Test is also

---

[3] A. Ward and M. Ward. (1999), *Assessment in Classroom,* London: Wadsworth.

[4] R. L. Linn and M. D. Miller. (2008). *Measurement and Assessment in Teaching*, 9th ed., Delhi: Dorling Kindersley Pvt. Ltd.

[5] R. L. Ebel and D. S. Frisbie. (1986). *Essentials of Educational Measurement*, 5th ed., New Delhi: Prentice Hall of India Private Limited.

[6] M. A. Raza (2009). *Standardization of an Aptitude Test in the subject of Science at Elementary Level, Unpublished PhD Thesis*, Bahauddin Zakariya University, Multan.

[7] L. R. Aiken (2000). *Psychological Testing and Assessment*, 10th ed., Boston: Allyn and Bacon.

[8] W. A. Mehrens and I. J. Lehmann. (1991). *Measurement and Evaluation in Education and Research*, Fort Worth, TX: Harcourt Brace College Publishers.

defined by Aiken[9] as an assessment procedure that includes a number of measures which help to create a more comprehensive picture of a persons' performance.

Multiple Choice Items

Multiple choice items are the most common and widely studied type of objective type tests. In the following, focus is being narrowed down to this sort of tests.

Multiple choice items or *Multiple Choice Questions* (MCQs) are also referred as selection response item. These are the most commonly used and the familiar items being used in education. The format presents two or more possible answers among which only on is the correct response while others represent the common errors. These items produce highly reliable and well established and efficient test results. Multiple choice items still are limited to the measurement of achievement or aptitude sometimes, because they can only measure simple learning outcomes rather than the complex learning outcomes. For example measurement of performing an experiment cannot be checked from these types of items. Despite of its drawbacks, these tests highly recommended due to their maximum control on valid and reliable standardized tests.

Multiple choice test formats are extremely recommended in measuring knowledge and comprehension level outcomes. Knowledge outcomes are highly concerned with vocabulary, facts, principles, methods and procedures. Different levels of comprehension like understanding, application and interpretation can be measured with this type of questions.

Multiple choice items are widely applicable to ensure different phases of achievement testing. It does not possess many of the problems which other objective type items present. These can easily and comprehensively define the problem which a short answer item can never present. These reduce the problem of homogeneous material like matching items and also minimize the issue of guessing problem which true false items possess.

On the other hand, the major limitation of these items is that they are only the selection type paper and pencil test items which just can measure problem solving behavior at the verbal level only due to its requirement of selection type question. They are unable to measure the complex learning outcomes.

According to Linn and Gronlund, (2000),[10] traditionally an MCQ consists of a:

Stem: the text of the item

Options: provided by the examiner from which examinee has to choose any one

---

[9] L. R. Aiken (2000). op.cit.

[10] R. L. Linn and N. E. Gronlund. (2000). *Measurement and Assessment in Teaching*, 7th ed., New Jersey: Prentice Hall. Inc.

Correct response: the key answer present in the list of options.

Distracters: the wrong or the incorrect answer in the list of the options.

Objectives of the Study

Specific objective of the research study was to construct and analyze an achievement test (Islamic Foundations of Education) consisted of multiple choice items. Beside these specific objectives some general objectives are as follows;

1. **Determining the reliability of the test.**
2. **Estimating the validity of the instrument.**

Procedure of the Study

An achievement test in the subject of Islamic foundations of education, comprising 60 multiple choice items was constructed. The test was divided into two parallel forms to estimate reliability. Table of specification was made to estimate validity of the test. The researcher collected the course outlines from the concerned universities to construct the relevant items. After construction the test was administered on 100 students of M. A. Education initially, for the pilot study. After doing item analysis through traditional method, the selected items were arranged in the form of the final draft. Multi-stage sampling technique was used to select the Master Level students of Education at all the universities of Punjab. The administration of test was performed with the consent and permission of the concerned university officials. Scoring keys for test was prepared and scoring was done by using these keys. After scoring, the scores of all the students were recorded in the form of tables. Item analysis was done through traditional method.

Definition of the Terms

Students: The students of Education at master level in the three universities of Punjab i.e. Bahauddin Zakariya University Multan (BZU), The Islamia University of Bahawalpur (The IUB), University of Sargodha (UOS).

Candidates: The students on whom the test was administered

High Achievers: Candidates who got maximum marks in the test

Low Achievers: Candidates who got minimum marks in the test

Data Analysis

Analysis of this teacher made test has been performed by using item difficulty and discrimination method. A brief introduction of these terms is discussed below.

Consideration of Question Difficulty

An item’s difficulty level is important because it tells the instructor something meaningful about the comprehension, or performance, or material or tasks contained in the item. This paradox is comprehensible when we recall that the difficulty index represents the percentage of the total number of respondents answering the item correctly. In other words, an inverse relationship exists between the magnitude of the index and what it purports to represent. In any event, an instructor might be justified in concluding that nearly everyone had command of the material for Item 1. Extremely high difficulty indices, however, can indicate a structural defect in the item. The data for Item 1 could have been obtained from the following item:

Item 1: Among the major contributors to low reliability are;
1. An appropriate time limit
2. Inadequate samplings of content and individuals. (Keyed as Correct)
3. Lack of content heterogeneity in the test.
4. Differential weighting of alternatives in scoring each item.
5. Poor lighting in the testing room.

When we examine the content and structure of the test, it is obvious that a grammatical clue exists. The stem calls for a plural response, and the only plural response is “2” —which in this case happens to be the correct answer A student who noticed this clue could respond correctly to the item without knowing the answer This irrelevant clue could alone account for the high difficulty index, particularly where the low group is concerned. The lesson here is obvious. In selecting items for a test, considering content alone or item analysis data alone can be very misleading. Both factors need to be considered when you accept items for the final form of the test.

The difficulty index is only an estimate of the “real” difficulty level of an item and is based on the responses of only the high- and low-scoring groups. The middle groups, usually from 50 percent to 33 percent of the total, have been eliminated.

Considerations of Question Discrimination

Item discrimination has been defined as the degree to which an item differentiates the high achievers from the low achievers. Actually, the term differentiation is probably more descriptive than discrimination, but we will stick with the latter term rather than trying to change many decades of psychometric history. A perfect positively discriminating item would be answered correctly by the entire high group and none of the members of low group; the discrimination index

would then be + 1.00. If the entire low group and none of the member of high group respond correctly, the index would be —1.00. In a sense, one might interpret the discrimination index as the correlation of the item with the total test score. Extreme values are almost never observed on a classroom test, rather, items in the middle range of positive discrimination are usually found.

Examining Distracter Effectiveness

An ideal item, at least from a statistical item analysis standpoint, is one that all students in the high group answer correctly and all students in the low group could not reply correctly. In addition, the responses of the low group should be evenly distributed among the incorrect alternatives. However, this rarely happens in practice.

Item-Analysis; Purposes and Procedure

Item-analysis is a procedure for evaluating the effectiveness of an item in a test. So here item-analysis was conducted for four general purposes;

i. To select the best available items for the final form of a test.
ii. To identify any structural and content defects in any of the items.
iii. To detect learning difficulties of the sample as a whole; identify general content areas or skills that need to be reviewed by the teacher, and
iv. To identify for individual students areas of weakness which might be in need of remediation

In order to get relatively accurate and reliable results, item-analysis was conducted as follows;

i. As the total score of the test is the most satisfactory criterion on which to base a ranking of individuals for an analysis of a test, the answer sheet of 385 students was arranged from high to low on their performance on the whole test.
ii. High and low scoring groups were identified by picking up answer sheets of 27% (103 students) from the top and 27 % (103 students) from the bottom.
iii. Destructors in each item, chosen by individuals in the high and low groups were also recorded.
iv. To get difficulty index (Facility index), sum of the number of correct answers to each item made by the combined high and low groups was divided by the total number of students in the combined high and low groups and Multiplied the result by 100.

$$F = \frac{N_h + N_l}{2n} \times 100$$

Where

$$F\ is\ facility\ index$$
$$N_h\ is\ the\ number\ of\ correct\ responses\ in\ the\ top\ 25\%$$
$$N_l\ is\ the\ number\ of\ correct\ responses\ in\ the\ bottom\ 25\%$$
$$N\ is\ the\ number\ of\ candidates\ or\ students$$

Items with the value of ranging from 15% to 75% were retained.

i. In order to calculate discrimination index (D) of an item, the following formula was used:

$$D = \frac{N_h - N_l}{n}$$

Where

$$D\ is\ the\ discriminating\ power$$
$$D\ is\ the\ discriminating\ power$$
$$N_h\ is\ the\ number\ of\ correct\ responses\ in\ the\ top\ 25\%$$
$$N_l\ is\ the\ number\ of\ correct\ responses\ in\ the\ bottom\ 25\%$$
$$n\ is\ the\ number\ of\ candidates\ or\ students\ in\ each\ group$$

ii. Items with discrimination of more than .20 were retained. Items with the value of D less than 20 were tossed out, as they were unable to discriminate between high achiever and low achievers.[11]

iii. Distracter was discarded and re-examined when:
   a. It was not at all more attractive to members of lower score group.
   b. It attracted high achievers more than low achievers.
   c. It attracted both groups almost equally.[12]

Traditional Item Analysis in Islamic Foundations of Education (Form A)

Item No.	Group	Alternatives	Correct Responses	Difficulty Index	Discrimination Index	Distracters
1.	High Low	11,73,6,3,6 44,45,7,5,2	B	57.82	0.27	
2.	High Low	12,29,57,5,1 15,41,33,8,6	C	43.60	0.23	
3.	High Low	7,26,52,19,0 10,33,35,24	C	34.46	0.28	

[11] C. Boopathiraj and K. Chellamani. (2013) *Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education,* International Journal of Social Science and Interdisciplinary Research, 2 (2).

[12] J. Qadir and I. Gillani. (2012), *Rasch Calibration of General Science Test at Grade-VIII in Pakistan,* Journal of Education and Vocational Research, 3(3), pp. 98:106.

4.	High Low	73,17,8,6,0 40,38,11,14,0	A	54.84	0.32	
5.	High Low	58,6,26,12,0 26,22,17,39,1	A	40.77	0.31	c
6.	High Low	12,67,14,11,0 29,38,26,10,1	B	51.54	0.25	
7.	High Low	12,50,28,10,4 19,58,16,10,1	B	52.81	-0.07	c
8.	High Low	10,21,19,53,1 29,24,22,25,3	D	37.57	0.28	
9.	High Low	59,23,9,11,2 14,25,28,33,3	A	38.43	0.49	
10.	High Low	17,16,34,34,3 13,21,19,46,5	C	25.72	0.15	
11.	High Low	13,34,35,21,1 22,24,36,20,1	B	28.16	0.09	
12.	High Low	29,15,22,34,4 18,22,40,18,5	D	24.75	0.16	a
13.	High Low	67,13,15,8,1 31,35,22,13,2	A	47.57	0.34	
14.	High Low	17,12,56,15,4 29,16,29,33,1	C	41.26	0.32	
15.	High Low	68,19,8,8,1 27,27,25,18,5	A	46.11	0.39	
16.	High Low	25,46,12,21,0 41,24,17,15,8	B	34.46	0.20	d
17.	High Low	11,14,8,70,1 38,12,19,34,0	D	50.97	0.33	
18.	High Low	18,16,20,49,1 33,29,11,23,7	D	34.95	0.25	
19.	High Low	18,25,9,52,0 20,32,31,20,	D	35.92	0.29	

20.	High Low	12,26,15,50,1 44,25,19,14,1	D	31.08	0.34	
21.	High Low	28,28,20,26,2 36,19,28,20	B	22.53	0.09	
22.	High Low	39,22,37,6,0 29,29,34,11,1	A	34.49	0.07	
23.	High Low	24,29,36,13,2 29,34,27,10,3	A	26.14	-0.03	c
24.	High Low	35,29,22,15,2 16,44,30,10,3	A	24.75	0.18	d
25.	High Low	31,15,47,10,1 9,22,36,36,0	C	46.77	0.09	a
26.	High Low	53,14,22,14,1 46,14,25,16,2	A	40.05	0.06	
27.	High Low	22,27,45,9,1 38,34,24,5,2	C	33.06	0.19	
28.	High Low	11,9,21,53,0 6,34,37,24,2	D	37.37	0.29	
29.	High Low	31,15,47,10,1 64,25,7,5,2	A	58.73	-0.08	
30.	High Low	63,22,11,8,0 37,22,28,13,3	A	48.02	0.24	

On the basis of facility index item no 21 has the minimum value (22.53).While item no 29 has the maximum value (58.73).On the basis of discrimination index item no 7,11,21,22,23,25,26,29 were needed modification as they could not discriminate between the high and low group.

For the investigation of the suitability of Multiple-choice items in the test, the Behavior of each of the distracter was clearly examined. Alternatives A (12, 25, 28), C (5, 7, 18, 23), D (16, 24), were rejected as they were more attractive to the members of high achievers than to the low achievers.

Traditional Item Analysis: Islamic Foundations of Education (Form B)

Item No.	Group	Alternatives	Correct Responses	Difficulty Index	Discrimination Index	Distracters
1.	High Low	2,53,25,21,2 7,37,35,24,1	B	45.69	0.15	
2.	High Low	12,56,26,11,0 19,33,37,14,1	B	43.69	0.21	
3.	High Low	4,3,17,79,1 13,25,48,16,2	D	46.6	0.60	

4.	High Low	24,6,63,11,0 37,25,27,13	C	46.15	0.32	
5.	High Low	57,19,12,13,3 29,20,19,24,1	A	47.09	0.16	
6.	High Low	5,79,13,5,2 14,64,12,13,1	B	70.39	0.12	
7.	High Low	6,57,26,12,3 16,32,40,15,1	B	43.02	0.24	
8.	High Low	55,10,19,19,1 23,20,28,33,0	A	37.86	0.31	
9.	High Low	18,58,15,12,1 41,17,29,14,2	B	35.41	0.39	
10.	High Low	17,13,21,53,0 28,16,25,34,1	D	42.23	0.19	
11.	High Low	8,28,45,20,3 23,27,30,23,1	C	36.41	0.14	
12.	High Low	57,20,15,9,3 15,40,28,17,4	A	34.95	0.40	
13.	High Low	60,13,13,16,2 18,37,22,25,2	A	37.65	0.40	
14.	High Low	12,15,16,60,1 24,16,27,32,4	D	38.32	0.37	
15.	High Low	19,67,6,12,0 40,13,29,20,2	B	38.83	0.52	
16.	High Low	18,41,12,32,1 19,35,18,31,1	B	37.38	0.04	
17.	High Low	56,19,14,12,3 46,18,17,20,3	A	50	0.08	
18.	High Low	12,62,12,18,0 15,21,30,38,0	B	40.29	0.38	
19.	High Low	37,24,26,16,1 33,16,16,36,3	A	33.98	0.03	b
20.	High Low	16,69,14,6,0 27,24,27,9,2	B	44.65	0.42	
21.	High Low	5,20,8,72,0 21,29,16,38,0	D	52.91	0.32	
22.	High Low	16,65,14,10,10 28,25,31,17,3	B	43.69	0.36	
23.	High Low	57,8,22,16,1 41,21,25,14,3	A	47.57	0.15	
24.	High Low	52,31,9,11,1 26,39,19,17,3	A	37.96	0.25	
25.	High Low	11,55,30,7,1 15,27,47,12,3	B	38.8	0.27	
26.	High Low	13,14,65,11,1 46,13,22,10,4	C	46.15	0.41	

27.	High Low	27,44,22,10,1 26,30,31,13,4	B	38.41	0.12	
28.	High Low	19,7,61,17,0 39,25,16,22	C	37.38	0.43	
29.	High Low	14,63,19,7,1 36,40,20,4,4	B	50.49	0.21	
30.	High Low	47,11,26,16,3 30,18,26,27,3	A	37.86	0.17	

On the basis of facility index item no 6 has the minimum value (70.39).While item no 19 has the maximum value (32.88).On the basis of discrimination index item no 6,11,16,17,19,27 were needed modification as they could not discriminate between the high and low group.

For the investigation of the suitability of Multiple-choice items in the test, the behavior of each of the distracter was clearly examined. Alternatives B (19) were needed modification as they were more attractive to the members of high achievers than to the low achievers.

Validity and Reliability of the Instrument

Face validity was determined through expert opinion available to the researcher while content validity was ascertained by preparing table of specification (separate for each test).

Test reliability was computed through Chronbach Alpha

Sr. no.	Reliability	Percentage
1	Introduction to Education (A)	0.71
2	Introduction to Education (B)	0.82

Findings and Conclusions

In the subject of Islamic Foundations of Education (Form A) the value of facility index was calculated. Item no 21 has the minimum value (22.53).While item no 29 has the maximum value (58.73).On the basis of discrimination index item no 7,11,21,22,23,25,26,29 were needed modification as they could not discriminate between the high and low group.[13] In the subject of Islamic Foundations of Education (Form B) the value of facility index was calculated. Item no 6 has the minimum value (70.39).While item no 19 has the maximum value (32.88). On

[13] C. Boopathiraj and K. Chellamani. (2013). op.cit.

the basis of discrimination index item no 6, 11, 16, 17, 27 were needed modification as they could not discriminate between the high and low group.[14] Behavior of distracters was also calculated. Distracters a, c and d in Introduction to Education (Form A) and b in Introduction to Education (Form B) were identified. As a result of this study seven items in Form A and six in Form B were observed to be less than the acceptable percentage (20 Percent), whereas no single item in both forms was found more than the acceptable percentage (80 percent) which shows that the items beyond normal range should either be discarded or de-modified before further administration of the test.

3.2.1 Recommendations for future study

This is a teacher made test which is in the process of standardization. It has been attempted to follow test protocols of standardization. It is recommended to follow those including rules and methods of administration and purpose of the study should clearly be communicated to the concerned persons before the administration of the test.[15] It is also recommended to conduct the study on a large size of sample. Furthermore, number of items should also be increased in order to attain more content validity. The meager items should be improved as it is necessary part of test standardization. The study can be repeated in other subjects to develop an item bank for the students, teachers and the test developers. It is also recommended that item analysis for other developed tests may be performed by using other methods of item analysis; for example, item response theory, as it is an important phase in the development of test items.

[14] ibid and J. Qadir and I. Gillani. (2012), op.cit.

[15] J. C. Witt, et. al. (1994). *Assessment of Children: Fundamental Methods and Practices*, Madison: WCB Brown and Benchmark Publisher.