Asian People Journal (APJ)
eISSN : 2600-8971, Volume 1, Issue 1 (2018), PP 93-109
www.uniszajournals.com/apj

EVALUATING THE QUALITY OF ISLAMIC CIVILIZATION AND ASIAN CIVILIZATIONS EXAMINATION QUESTIONS

Ado Abdu Bichi [1*] & Rahimah Embong [2]

[1]*Faculty of Education, Northwest University, Kano-Nigeria*
[2] *INSPIRE, Universiti Sultan Zainal Abidin, Terengganu, Malaysia*

**Corresponding Author: Ado Abdu Bichi*
adospecial@gmail.com

Abstract: *Assessment of learning involves determining whether the content and objectives of education have been mastered by administering quality tests. Thus, the quality of the test items used in evaluating students' achievement should be a major area of concern in teaching and research in the field of education. This paper assesses the quality of Tamadun Islam dan Tamadun Asia (TITAS) or Islamic Civilization and Asian Civilizations Examination Questions by conducting item analysis. The developed one hundred (100) multiple choices questions were administered to N=36 degree students. The data obtained had been analyzed by conducting item analysis in order to determine the item difficulties, item discrimination indices and the distractors analysis. The finding of results indicates that, based on difficulty indices 18(18%) items were "problematic" or "faulty", 66(66%) of the items have poor discriminating power. Similarly, the result of the distractor analysis showed 59(59%) of the distractors been flawed, having failed to meet the set minimum standards. Based on the findings it could be assumed that the test used has not been validated during the item development processes. It is recommended that, TITAS test items used in measuring students' achievement should be made to pass through all the processes of standardization and validation by conducting item analysis to ensure its reliability as well as to minimize measurement errors.*

Key Words: *Item Analysis, Item Difficulty, Item Discrimination, Islamic Civilization and Asian Civilizations, Tamadun Islam dan Tamadun Asia (TITAS).*

Introduction

Assessment of students learning is very vital in education. The assessment of students' cognitive abilities, academic skills and intellectual development involves certain techniques employed to sample students' performance on a particular learning outcome targeted by the instructional objectives; one of those techniques is test (Bichi, et al. 2015).The test is expected to sample students' behaviours. Thus creating quality tests is very important in assessing the students' performance; many indices have been developed in order to construct valid and reliable items during test development. Test is regarded as the most popularly used technique for obtaining information in the school system (Olatunji and Owolabi, 2009). Abiri (2007) describes a test task presents a situation that makes it possible to elicit behaviour or performance of individuals and through it determine their knowledge, abilities, skills or feelings. According to Payne (1982), test enables teachers and other stake holders to systematically use its data for the purpose of making comparisons across individuals, classes,

schools, districts or countries. Whereas the teacher made tests developed and used in classroom instruction by the teacher are more popularly with the learners, standardized tests serve certification, quality control and benchmarking purposes (Okpala, Onacha, &Oyedeji, 1993; Ofo 1994).Now more than ever, it is critical that tests are efficient and effective at measuring ability and scores emanating from tests are reliable and precise measures of examinees ability.

Scores on assessments can affect the students, the teachers and the school administrators. Also, it is clear that quality assessment is likely to lead to improvements in student learning (Hamilton, Stecher and Klein, 2002). However, for assessment tool or test quality to be established certain criteria on the areas of test design, test analysis techniques and test score interpretation must be met (Bichi, 2015). Obviously, Quality test design is impacted by many elements including format, length, administration procedures, construction, validity and scoring schema (Kinsey, 2003). The nature and the quality of information gathered from the achievement test can control the educational development efforts and direct the instruction (Suruci and Rana, 2014). The most important characteristics of an achievement test used in assessing students' abilities are its reliability and content validity. Nunnaly, (1972) for a test to be reliable and valid, a systematic selection of test items with regard to subject content and degree of difficulty is necessary.

Sax (1989) is of the view that, making systematic and fair judgement of others performance can be a very challenging task. Evaluation cannot be made solely on the basis of intuition, harphazard guessing or custom. In education a variety of tools are used in evaluation of students' progress. Tests are tools frequently used to facilitate the evaluation process (Matlock-Hetzel, 1997). When tests are developed for instructional purposes, to assess the effects of educational programs, it is crucial to conduct item and test analysis.

Tamadun Islam dan Tamadun Asia (TITAS) or Islamic Civilization and Asian Civilizations course is a core course at the undergraduate levels in Malaysian universities. In Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia, TITAS is a core course, and some instruments or test items are usually developed from the content to assess the students' achievements in relation to the course objectives at a regular semester examinations. Multiple-choice and true/false items (questions) form the structure of the UniSZA semester TITAS examinations. Similarly, due to the relative importance of the course to the entire educational programs to meet the philosophy of Malaysian education systems; it is crucial to evaluate the quality of items used in assessing students' achievements

in this field. By conducting test and item analysis to evaluate the quality of UniSZA TITAS examination items, many statistics to be generated from the analyses can provide useful information for improving the quality and accuracy of the items used in UniSZA TITAS semester examinations. It will describe the statistical analyses which allow measurement of the effectiveness of individual test items. An understanding of the factors which govern effectiveness and a means of measuring them can enable us to create more effective test questions and also regulate and standardize existing tests.

The Islamic Civilization and Asian Civilisations

The Islamic Civilization and Asian Civilisations course *(Tamadun Islam dan Tamadun Asia)* or shortly known as TITAS or discusses civilizational studies which include introduction to civilizational studies, the interaction between the various Malay civilizations, China and India, Islam in Malay Civilization, issues in Islamic civilization and contemporary Asian civilizations, Islam Hadhari and the nation development process. The course strives to highlight the historical achievement of Islamic civilization, and how it compare and contrast with other civilization in essence and manifestation, and its critical role in dealing with fundamental problem of Muslims and non-Muslims in contemporary world. It reflects the universal understanding of Islamic and secular worldview and its intrinsic paradigm, and the crucial aspiration for the reconstruction of Islamic culture and civilization grounded on the essence of unity, rationalism, and tolerance, as strategic based in realizing the planning.

The objective of TITAS is to introduce students to the civilizational studies which include introduction to civilization, interaction between various civilizations, Issues-contemporary issues and implications to the country's development process as well as to produce students who have an attitude of respect, adopting the values and identity as a citizen.

Having taught the content of the Islamic and Asian Civilisations, the undergraduate students should be able to: (i) List the main concepts of Islamic civilization and Western civilization. (ii) Explain the importance and role of religion and their culture-each in life and (iii) Apply communication skills effectively in writing and orally at the individuals, groups and communities.

The Problem and Justification for the Item Analysis of TITAS

In UniSZA semester examinations are conducted for all students, at Sub-Degree (Diploma) and undergraduate level in order to assess their performance in relation to the content of the curriculum taught and to finally determine their suitability for the award of Diplomas and Bachelor Degrees of the University at the end of their programmes. Although the semester examinations are used as a criteria to make decision as to the suitability or otherwise of students to have mastered a particular curriculum contents; however, emphasis needs to be made on the quality and appropriateness of the items used in UniSZA TITAS semester examinations through a formal test and item analysis to select the appropriate items. To the point of this paper the non-standardisation of the UniSZA TITAS semester examinations items may grossly affects the validity of the entire examinations, the grades and the final awards of the Bachelor Degree as the case may be. So just how "good" are the Multiple-Choice UniSZA TITAS semester examinations items? How effective are the individual Multiple-Choice UniSZA IAC semester examinations items in predicting the students' overall performance in the whole area of study or field?

Purpose of this Evaluation

The purpose of this is to evaluate the UniSZA semester examination items to determine the validity of the items and whether the items are good enough to be used in assessing the students' abilities. Specifically this study aims to;

i. determine characteristics of UniSZA Islamic Civilization and Asian Civilization (IAC) Examinations Items
ii. identify IAC Examinations Item Difficulty indices
iii. identify IAC Examinations Item Discrimination indices
iv. determine the effectiveness of IAC Examinations item Distractors

Overview of The Item Analysis

The item analysis is an important phase in the development of an examinations, that are to determine the ability level of an examinee or student in a particular discipline or subject, the students ability is estimated using the total scores obtained on the response to the test items and contributes considerably in determining whether the examinee has passed or failed the subject.

Item analysis is a process which examines student responses to individual test items in order to assess the quality of those items as well as the quality of the test as a whole (Shakil, 2008). Item analysis enables instructors to increase their test construction skills, identify specific areas of course content which need greater emphasis or clarity, and improve other classroom practices. According to Krishnan (2013)

> *Item analysis broadly refers to the specific methods used to evaluate items on a test, both qualitatively and quantitatively, for the purpose of evaluating the quality of individual items. The goal is to help its developers to improve the instrument by revising or discarding items that do not meet a minimally acceptable standard. The qualitative review is essential during item development and involves experts who have a mastery of relevant material. Test review boards and content experts cannot always be equipped with the knowledge they require to identify "bad" or "defective" items because of such factors as the multidisciplinary nature of the test content and the demographic characteristics of test takers. The statistical analysis could help to identify problematic items that may have slipped the experts' attention, one way or the other. Thus, the quantitative analysis is conducted after the test/tool has been administered to the test takers. The objectives of both the qualitative and quantitative assessments remain the same – to assess the quality of items (p.7)*

According to Thompson & Levitov, (1985, p. 163), "Item analysis investigates the performance of items considered individually either in relation to some external criterion or in relation to the remaining items on the test." For example, when norm-referenced tests (NRTs) are developed for instructional purposes, such as placement test, or to assess the effects of educational programs, or for educational research purposes, it can be very important to conduct item and test analyses. Similarly, criterion-referenced tests (CRTs) compare students' performance to some pre-established criteria or objectives (Shakil, 2008). Some of the researchers that have contributed immensely to the theory of test item analysis are Galton, Pearson, Spearman, and Thorndike.

Generally an item in a test may fail to meet the minimum quality standard, whatever the set standard is. It may be as a result of: (1) the flaws in the question and (2) the flaws in the instruction of the content (Krishnan, 2013).

This item analysis involves many statistics that can provide useful information for improving the quality and accuracy of the multiple-choice and true/false items (questions) used in UniSZA semester examinations. It will describe the statistical analyses which allow measurement of the effectiveness of individual test items. An understanding of the factors which govern effectiveness and a means of measuring them can enable us to create more effective test questions and also regulate and standardize existing tests.

The trait (or ability) of an examinee is defined in terms of a test, whereas the difficulty of a test item is defined in terms of the group of examinees. According to Hambleton, et al. (1991, p. 3), "Examinee characteristics and test item characteristics cannot be separated: each can be interpreted only in the context of the other."

Some important criteria which are employed in the determination of the validity and reliability of a multiple-choice examination are: (i) Item Difficulty (ii). Item Discrimination and (iii). Effectiveness of the Distractors

Item Difficulty

Item difficulty indicates the proportion of students who answered the item correctly/right in a test. A high percentage indicates an easy item/question and a low percentage indicates a difficult item, it is represented and called *p-values*. The difficulty value range is between 0.0 and 1.0. In general, an item should have values of difficulty no less than 30% correct and no greater than 70% (Adegoke, 2013; Zubairi & Kassim, 2006; Henning 1987). Items with the two extreme values (Very difficult or very easy) contribute little to the discriminating power of a test. The item difficulty index is one of the most useful, and most frequently reported, item analysis statistics

Item Discrimination

The item discrimination index is a measure of how well an item is able to distinguish between examinees who are knowledgeable and those who are not, or between higher ability and low ability students. There are actually several ways to compute item discrimination but the popular ones are Discrimination Index (D) and point-biserial correlation (r_{pbi}).The former is the difference between the proportion of the higher scorers and the bottom scorers who got

the item right (each consists of 27% of the total number of students who took the test and is based on the students' total score). The later looks at the relationship between an examinee's performance on the given item (correct or incorrect) and the examinee's score on the overall test. Discrimination index range is between -1 and +1. Ebel and Frisbe (1991), classified the items based on their discrimination values, 0.40 and above "Very good", 0.30 to 0.39 "reasonably good" but possibly subject to improve, 0.20 to 0.29 "Marginal" usually subjected to improvement and below 0.19 is "Poor" is to be rejected or completely revised. The closer the index is to +1, the more effectively the item distinguishes between the two different groups of students. For an item that is highly discriminating, in general the examinees who responded to the item correctly also did well on the overall test and vice versa. Sometimes an item will discriminate negatively. That is an indication that, the lower achieving students actually selected the key response more frequently than the higher performers.

Distractor evaluation

Distractor analysis is another important element in assessing the effectiveness of a test item. Distractor analysis assesses performance of the incorrect response options; all of the incorrect options should actually be distracting and plausible. According to Shih (2010) it is a procedure related to multiple choice formats and is conducted to see how distractors are functioning. In item analysis neither the item difficulty nor the item discrimination index considers the performance of the incorrect response options in a test. Preferably, each distracter should be reasonable to lower achieving examinee and be selected by a greater proportion of the lower scorers than of the higher scorers. Ideally for a distractor to be effective and acceptable it should attract at least one candidate or as maintained by Tarrant et al. (2009) that a functional distractor is one that was selected by at least 5% of examinees. If a distractor appears so unlikely that almost no examinee selected it, such an item is not contributing to the performance of the item, this particular distractor should completely be revise the option to make distractor a more plausible choice.

Materials and Method

Research Design

The Ex-Post- Factor research design was employed because no manipulation of any kind was made to the material used, administration and scoring of the students' responses i.e it has already occurred under the supervision of the course lecturer.

Participants

The participants comprises the entire students who sat for the examination in the semester, thirty six (36) undergraduate students, age (19-25) from the Faculty of Islamic Contemporary Studies participated in the study.

Table 1: *Demographic Information of the Respondents*

Variable	Level	Number	Percentage
Gender	Male	20	55.6%
	Female	16	44.4%
Age	19 - 22	17	47.2%
	23 - 25	19	52.8%

Instruments

UniSZA Islamic Civilization and Asian Civilisations semester examination items designed and constructed by the course lecturers was used. The examination comprises 100 items (i.e 70 multiple-choice items with four answer choices/options and 30 True or False questions).

Data Collection

The 100 items UniSZA Islamic and Asian Civilisations semester examination items was administered on the students after receiving specific instruction for the examination by the lecturer and the invigilators assigned by the university during first semester examinations of 2014/2015 academic session. The test items were scored by the course lecturer and the scores and responses was what is used as a data for this analysis.

Data Analysis

The item analysis to determine item difficulty, item discrimination indices and distractor effectiveness as well as the reliability of the items was carried out using the Statistical Package for the Social Sciences (SPSS 20v).

Results

Table 2: *Summary Item Statistics*

	Number of Items	Number of Examinees	Reliability ***(Alpha)***	Mean Scores	S.D	Mean ***P***	Mean r_{pbi}
Test items	100	36	0.70	57.83	7.41	0.578	0.120

The summary statistics on table 2, above shows that, for the 100 item test administered to 36 students, the overall reliability of the test as measured by the Cronbach's Alpha is 0.70, which is high in line with the Nunnally (1978), recommends acceptable value of 0.70. Similarly the items mean scores is 57.83 with standard deviation of 7.41. The mean item difficulty (p) is 0.578 and mean item discrimination of the test (r_{pbi}) is also 0.120 as presented.

Questions 1: What are item characteristics of the UniSZA Islamic Civilization and Asian Civilizations Examination Questions?

The item parameters of TITAS test generated [i.e item difficulty (***p***) and item discrimination (r_{pbi})] are presented in the table 2 below

Table 3: *Item characteristics (Item difficulty and Discrimination values)*

Item	Item Difficulty *(P)*	Item Discrimination *(rpbi)*	**Item**	Item Difficulty *(P)*	Item Discrimination *(rpbi)*
1.	.97	**.07**	51.	.89	.18
2.	.31	-.32	52.	.58	.25
3.	.81	.26	53.	.56	.38
4.	.72	.35	54.	**.28**	**-.06**
5.	**.22**	-.14	55.	.61	.44
6.	**.11**	.12	56.	.72	.31
7.	.36	.22	57.	.58	**.03**
8.	.42	.49	58.	**.19**	.12
9.	.86	.18	59.	.47	**-.09**
10.	**.28**	.11	60.	.39	.14
11.	.58	.15	61.	**.25**	-.20
12.	.69	**.02**	62.	**.19**	.39
13.	.94	.15	63.	.56	**-.28**
14.	.42	.13	64.	.56	.25

15.	.75	.18	65.	.44	.35
16.	.64	.18	66.	.44	-.09
17.	.39	**.02**	67.	.47	.24
18.	**.06**	**-.06**	68.	**.14**	**-.05**
19	**1.00**	**0.00**	69.	.50	.11
20.	.89	.11	70.	.42	**.04**
21.	.81	**.09**	71.	.39	.20
22.	.83	**.07**	72.	.33	-.17
23.	.69	**-.00**	73.	.67	.43
24.	.31	.11	74.	**.14**	**.04**
25.	.39	.33	75.	.81	.55
26.	.97	.30	76.	.75	.38
27.	.78	.28	77.	.67	**.00**
28.	.92	.16	78.	.94	-.14
29.	.69	-.22	79.	.67	.35
30.	**.14**	**-.10**	80.	.94	.33
31.	.58	.12	81.	.39	**.03**
32.	.33	.19	82.	**.22**	**.11**
33.	.97	.25	83.	.67	**-.02**
34.	**.22**	-.32	84.	.86	**.17**
35.	.94	**.11**	85.	**.14**	-.47
36.	.75	.31	86.	.72	.26
37.	.39	**-.06**	87.	.64	.49
38.	.61	.28	88	**1.00**	**0.00**
39.	.81	.16	89.	**.06**	.19
40.	.72	.30	90.	.92	.33
41.	**.28**	**-.42**	91.	.78	**-.19**
42.	.89	**.12**	92.	**.11**	**-.30**
43.	.50	.30	93.	.83	**.07**
44.	**.22**	.28	94.	.78	**.05**
45.	.39	.39	95.	.78	-.18
46.	.61	.18	96.	.81	**.07**
47.	.97	**.09**	97.	.58	.39
48.	.67	**.11**	98.	.81	**.12**
49.	.39	.20	99.	.61	.15
50.	.61	**-.01**	100.	.78	.44

The result of analysis presented on table 3 shows item characteristics of UniSZA IAC. The item characteristics; difficulty **(*p*)** and discrimination indices **(r_{pbi})**] were generated using SPSS 20v. The value of each characteristics is stated and the items of special interest are in

bold in both difficulty and discrimination parameters. The item characteristics presented are used for answering the research questions 2 and 3 based on the predetermined standards or guidelines for determining test item quality.

Question 2: Are the UniSZA IAC test items of the acceptable difficulty level?

Table 4: *Distribution of Items based on Difficulty Indices*

Difficulty index	Items	Total Items
Easy (P> 0.70)	1, 3, 4, 9, 13, 15, 19, 20, 21, 22, 26, 27, 33, 35, 36, 39, 40, 42, 47, 51, 56, 75, 76, 78, 80, 84, 86, 88, 90, 91, 93, 94, 95, 96, 98, 100	37 (37%)
Moderately ($0.31 \leq 0.70$)	2, 7, 8, 11, 12, 14, 16, 17, 23, 25, 29, 31, 32, 37, 38, 43, 45, 56, 48, 49, 50, 52, 53, 55, 57, 59, 60, 63, 64, 65, 66, 67, 69, 70, 71, 72, 73, 77, 79, 81, 83, 87, 97, 99	45 (45%)
Difficult ($P \leq 0.30$)	5, 6, 10, 18, 30, 34, 41, 44, 54, 58, 61, 62, 68, 74, 82, 85, 89, 92.	18 (18%)

Based on the set standards for interpreting difficulty indices 45 (45%) of the Items were of moderate difficulty, 37(37%) were easy, and 18(18%) were considered difficult. With this rule, 18 items are difficult and can be considered 'poor' or 'faulty' items. In conformity with the rule, 45 out of the 100 items are "good" (moderately difficult) and 37 items can be seen as "fair" (easy). On the basis of the item selection criteria of difficulty indices of (*0.30>P>.070*), 55 items that failed to satisfy the condition are considered 'poor' items. These poor items should be completely revise specially items 6, 18, 30, 58, 62, 68, 74, 85, 92 which have less than 20% of the examinees getting it right, and the items 19 and 88 which have all the examinees getting it right (100%)

Question 3: Are the UniSZA IAC items discriminate between higher and low achieving examinees?

Table 5: *Distribution of Items based on Discrimination Indices*

Discrimination index	Items	Total Items
Very Good ($D \geq 0.40$)	8, 55, 73, 75, 87, 100	6 (7%)
Reasonably Good (0.30 – 0.39)	4, 25, 26, 36, 40, 43, 45, 53, 56, 62, 65, 76, 79, 80, 90, 97	16(16%)

Marginal (0.20-0.29)	3, 7, 27, 33, 38, 44, 49, 52, 64, 67, 71, 86	12(12%)
Poor (D ≤ 0.19)	1, 2, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 28, 29, 30, 31, 32, 34, 35, 37, 39, 41, 42, 46, 47, 48, 50, 51, 54, 57, 58, 59, 60, 61, 63, 66, 68, 69, 70, 72, 74, 77, 78, 81, 82, 83, 84, 85, 88, 89, 91, 92, 93, 94, 95, 96, 98, 99.	66(66%)

On the basis of discriminating index criteria set, the results indicates that 66 (66%) of the items failed to differentiate between students of different abilities, 12 (12%) items are marginal need to be reviewed, 16 (16%) of the items are satisfactory and 6(6%) of the items functions very well. based on the selection criteria of discriminating index (i.e. $r_{pbi} \leq 0.20$), 66 items are 'poor' and failed to satisfy the condition the items can be eliminated or completely revise.

Question 4: Are the UniSZA IAC items distractors effective?

Table 6: *Distractor Analysis*

Item No.	A	B	C	D	Key		Item No.	A	B	C	D	Key
1.	**1**	35	**0**	**0**	B		36.	27	2	**0**	7	A
2.	22	3	11	**0**	A		37.	5	14	13	4	B
3.	**1**	**1**	29	5	C		38.	9	22	**0**	5	B
4.	26	2	3	5	A		39.	7	**0**	29	**0**	C
5.	8	4	18	6	C		40.	26	**1**	2	7	A
6.	4	5	2	25	D		41.	10	13	8	5	B
7.	**0**	13	11	12	B		42.	32	**1**	2	**1**	A
8.	6	11	4	15	D		43.	2	13	3	18	D
9.	3	31	**0**	2	B		44.	8	11	5	12	D
10.	10	8	10	8	C		45.	3	11	8	14	D
11.	7	4	4	21	D		46.	5	22	3	6	B
12.	**0**	7	4	25	D		47.	35	**0**	**1**	**0**	A
13.	34	**1**	**1**	**0**	A		48.	24	6	6	**0**	A
14.	12	15	7	2	B		49.	5	14	3	14	B
15.	**1**	**1**	27	7	C		50.	8	22	**0**	6	B
16.	5	7	**1**	23	D		51.	**1**	2	32	**1**	C
17.	14	17	3	2	B		52.	12	2	**1**	21	D
18.	2	15	**0**	19	D		53.	7	20	2	7	B
19	36	**0**	**0**	**0**	A		54.	14	5	10	7	A

20.	**1**	32	2	**1**	B		55.	4	22	8	2	B
21.	29	**0**	**1**	6	A		56.	**1**	5	26	4	C
22.	30	5	**0**	**1**	A		57.	4	9	21	2	C
23.	9	**0**	25	2	C		58.	2	18	9	7	B
24.	**0**	**1**	11	24	D		59.	7	3	9	17	D
25.	1	14	17	4	C		60.	2	14	5	15	D
26.	**0**	**0**	**1**	35	D		61.	9	13	2	12	B
27.	**1**	2	5	28	D		62.	2	**1**	7	26	D
28.	**0**	**1**	2	33	D		63.	2	2	20	12	C
29.	3	2	6	25	D		64.	20	6	7	3	A
30.	7	5	19	5	C		65.	11	2	16	7	C
31.	3	5	7	21	D		66.	4	13	3	16	D
32.	7	14	3	12	B		67.	5	2	17	12	C
33.	**0**	**1**	**0**	35	D		68.	16	13	5	2	A
34.	8	7	15	6	C		69.	2	4	12	18	D
35.	**0**	34	**0**	2	B		70.	**1**	2	18	15	C

Table 6 above gives a more detailed look at how the distractors functions in the test. The 70 multiple choice items part of the test has 280 responses/options out of which seventy (70) are the key or correct options and the remaining 210 are the wrong options/distractors in the data set and many were flawed.30 (14.3%) of the distractors were flawed because they were not chosen by any of the examinees. In addition, 29 (13.8%) of the distractors were also flawed because only 1 examinee choose the options which is less than 5% of examinees as suggested by Tarrant et al. (2009). By this definition, only 151 (72%) of 210 distractors on the examination functioned properly.

Discussion

The focus of this study is to evaluate the quality of the UniSZA Islamic Civilization and Asian Civilizations test used in assessing students' abilities.

The findings reveals that based on the established standards 45 (450%) of the Items were of acceptable difficulty level, i.e ($0.31 \leq 0.70$). 37 (37%) were easy, and 18 (18%) were difficult. On the basis of the item selection criteria of difficulty indices of (*0.30>P>.070*), 55 items that failed to satisfy the condition are considered 'poor' items and were to be rejected. This finding disagreed with the findings of Pande et al. (2013) and Suruci and Rana (2014) whose findings revealed that, majority (75%) and (78%) of the items respectively, were of acceptable level as far as difficulty was concern.

On the basis of item discrimination indices, the results indicates that 66 (66%) of the items failed to differentiate between students of different abilities, 12 (12%) items are marginal need to be reviewed,16 (16%) of the items are satisfactory and 6 (6%) of the items functions very well. Considering the Ebel and Frisbie, (1991) set criteria of item selection based on its discriminating index (i.e. rpbi ≤ 0.20), 66 items are "poor" and failed to satisfy the condition the items can be reviewed, replaced or eliminated completely from the test. This denotes that 36% of the test items are in the range of good and very good acceptable discrimination level. This study is also in consistent with the findings of Bichi (2015) and that of Pande et al. (2013) whose study on evaluating the quality of multiple choice questions (MCQs) in Chemistry and formative examination in Physiology revealed having 80% and 75% of the items within acceptable to excellent discrimination.

Similarly, the results of the options or distractors analysis gives a more detailed information on the performances of the options/distractors, out of the 280 options with 70 key or correct options; in the remaining 210 wrong options/distractors, 59 (28%) of the distractors were flawed because they were not chosen by any of the examinees or were only chosen by one examinee. the flawed options in the data set should be revised to make it more plausible choice because it is not contributing to the performance of the items. This can affect the entire test score reliability and the validity of the results, as according to Tarrant et al. (2009) a functional distractor is one that was selected by at least 5% of examinees. If a distractor appears so unlikely that almost no examinee selected it, such an item is not contributing to the performance of the item.

Conclusion and Recommendations

Findings of this paper emphasises a significant role of item analysis to educators and test developers in determining the quality achievement test items especially during item development. Looking at the item parameters of difficulty and discrimination will assist a test developer in detecting the defective and good individual items. Here this evaluation has been able to establish that an individual item in a test with moderate difficulty and a good positive discrimination power are ideal for a good test. However, an items having zero or negative discrimination power with very low or high difficulty estimates should be completely revise, improve or out rightly rejected. Item analysis results that are generated may be influenced by many factors which include students' perception of the test, students having poor understanding of difficult topics, ambiguity in wordings of the questions or even

inappropriate key, instructional procedure applied, it may also be due to personal variations in students‟ intelligence level.

Going by the significant role played by item analysis in evaluating and improving test items, it is recommended that;

Item analysis should be maintained in UniSZA TITAS test development to develop a good item bank for assessing the real students' abilities.

Secondly, UniSZA TITAS examination items should adequately be improved by removing or revising the poorly constructed items to make its results more reliable and valid in assessing the undergraduate capacities

Thirdly, the flawed distractors should be made to be more plausible & that would improve their performance. Additionally, new items should be develop and reviewed to verify their clarity, accuracy, content and structure by employing the services of subject matter experts and test or measurement specialists.

Finally, more other verifiable model of item and test development (i.e Item Response Theory Principles) should be applied in the development of UniSZA TITAS examination items to ensure that valid and reliable test items are use in making a correct or right decision on students' progress

References

Abiri, J.O.O. (2007). *Element OF Evaluation and Measurement Techniques in Education*. Ilorin: Library and publication committee University of Ilorin Nigeria.

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*, Vol.4, (22).

Bichi, A. A. (2015). Item Analysis using a Derived Science Achievement Test Data.*International Journal of Science and Research (IJSR), Volume 4 Issue 5, 1656-1662*

Bichi, A. A.,Embong, R., Mamat, M. & Maiwada, D. A. (2015).Comparison of Classical Test Theory and Item Response Theory: A Review of Empirical Studies. *Australian Journal of Basic and Applied Sciences 9(7) pp. 549-566*

Ebel, R.L. and Frisbie,D.A. (1991).*Essentials of Educational Measurement*. 5th Edn., Prentice Hall, Engelwood Cliffs, New Jersey.

Gurski, L. F. (2008). *Secondary Teachers' Assessment and Grading Practices in Inclusive Classrooms*. A Thesis Submitted to the College of Graduate Studies and Research in Partial Fulfilment of the Requirements for the Degree of Master of Education, University of Saskatchewan.

Henning, G. (1987). *A Guide to Language Testing: Development, Evaluation, Research*. Newberry House Publisher, Cambridge Mass.

Kinsey, T. L . (2003). *A Comparison of IRT and RASCH Procedures in a Mixed-Item Format Test*. Unpublished Doctoral Thesis, University of North Texas.

Klein, S.P., & Hamilton, L.S. (1999).Large-scale testing: Current practices and new directions (IP-182). Santa Monica, CA: RAND

Krishnan, V. (2013).*The early child Development Instruments (EDI): An Item Analysis using Classical Test Theory (CTT) on Alberta"s Data*. Early Child Development Mapping (ECMap) Project Alberta, Community- University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta.

Matlock-Hetzel,S. (1997).*Basic Concepts in Item and Test Analysis.*(Online) available at files.eric.ed.gov/fulltext/ED406441.pdf [accessed on June 24, 2014.

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.

Ofo,J. E. (1994). *Research Methods and Statistics in Education and Social Sciences*.Joja Educational Research and Publishers, Lagos.

Okpala, P.N., Onocha, C. O. and Oyedeji, O. A. (1993).*Measurement and Evaluation in Education*.Jattu- Uzairue Stirling- Horden Publishers Nigeria.

Olatunji,D.andOwolabi, H. O. (2009). Difficulty and Discrimination of Economics Test Items with Various Option Formats among Secondary Schools in Ilorin, Nigeria.*Ilorin Journal of Education,* Vol. 28 pp.49-63.

Pande, S.S., Pande, S.R., Parate,V.R., Nikam, A.P., and Agrekar, S.H. (2013). Correlation between Difficulty and Discrimination Indices of Multiple Choice Questions in Formative Exam in Physiology.*South East Asian Journal of Medical Education*, 7: pp.45-50.

Payne,J. (1982). Contingent Decision Behaviour.*Psychological bulletin*, 92, Pp.382-402.

Pope,G. (2009). *Item analysis analytics part 1: What is Classical Test Theory?* (Online) available at http://blog.questionmark.com/item-analysis-analytics-part-1-what-is-classical-test-theory[accessed on July 5, 2014]

Sax, G. (1989).Principles of educational and psychological measurement and evaluation (3rd ed). Wadsworth, Belmont, CA.

Shakil,M. (2008). *Assessing Student Performance Using Test Item Analysis and its Relevance to the State Exit Final Exams of MAT0024 Classes - An Action Research Project*. A Paper presented on MDC Conference Day, March 6th, 2008 at MDC, Kendall Campus.

Shih, Y. (2010).*An Item Analysis of an English Achievement Test Taken by EFL College Students in Taiwan.*Online available at https://www.researchgate.net/publication/265025312_An_Item_Analysis_of_an_English_Achievement_Test_Taken_by_EFL_College_Students_in_Taiwan_An_Item_Analysis_of_an_English_Achievement_Test_Taken_by_EFL_College_Students_in_Taiwan. Accessed on 23rd April, 2015.

Suruchi and Rana, S. R. (2014). Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology. *Paripex - Indian Journal of Research, Vol. 3(6) 56-58*

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

Thompson,B. and Levitov, J. E. (1985). Using microcomputers to score and evaluate test items. *Collegiate Microcomputer*, 3, pp.163-168.

Varma, S. (2008*).Preliminary item statistics using point-biserial correlation and p-values,* (Online) available at http://www.eddata.com/resources/publications/EDS_point_Biserial.pdf[accessed on October 7, 2014]

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, pp.1-20.