Original Article

An Item Analysis Study on TIMSS 2015 Mathematics Items of Omani and Iranian Students Comparison IRT and CDM Approaches

Aliasghar Rahman Doust[1], Waseem Ahmad Khan[2], Mohammed Al-Ghafri[3]

[1]*Ph.D. Scholar of Education, Department of Teacher Training & Non-Formal Education (I.A.S.E) Faculty of Education, Jamia Millia Islamia, New Delhi, India*

[2]*Professor at the Department of Teacher Training & Non-Formal Education (I.A.S.E) Faculty of Education, Jamia Millia Islamia, New Delhi, India.*

[3] *Assistant Professor of Mathematics Education, Department of Curriculum & Instruction, College of Education, Sultan Qaboos University*

Abstract - *The present paper aims to analyze the psychological properties of the mathematics fourth-grade items of Omani and Iranian students by IRT and CDM item analysis. The statistical samples were selected from all Omani and Iranian fourth-grade students who took TIMSS 2015 mathematics test. The research methodology was a secondary analysis method. The results of the IRT showed, there are no same difficult items for both countries. However, 22 items were recognized as very easy items for Oman and Iran. Furthermore, IRT showed, there were just 5 of the same appropriate items for Omani and Iranian students. Besides, the CDM approach found the 9 most difficult items for both Omani and Iranian students. Consequently, CDM analysis analyzed better the psychological properties of the items.*

Keywords - Item Analysis, TIMSS, Mathematics, Assessment, CDM, IRT

I. INTRODUCTION

Educational assessment leads educational experts and policymakers more passionate to focus more on the outcomes of the educational and achievement tests which could have been designed inappropriately. In other words, it is more essential to acquire vital feedback about educational, psychological, and achievement tests which are being implemented to assess examinees' or test-takers ' abilities. Xu and Liu (2009) pointed out, the teachers usually use their past assessment experiences to form their current assessment practices.

Furthermore, the major concern of the tests outcomes and their reliabilities back to the achievements tests which are determinative for students' future carriers and paths. Therefore, these types of tests are being designed by teachers, educational test designers and most classroom assessments involve the tests that have been designed by teachers (Siri & Freddano, 2011). Through constructed tests, the teachers should be aware of the test items' performance that whether are they able to reflect the students' performance in the test or not (M. Patel, 2017).

The reliability of the educational assessment or evaluation of the education tests causes to show an increasing rate of the examinees' performance in a test. In other words, test development or test designing for achievement tests needs to take more accuracy and follow standard approaches. If the test-developer does not follow the standard methods to prepare achievement or ability tests, the outcomes of this type of test may lead participants to achieve unwanted results (Taib & Yusoff, 2014).

To resolve the probability issues of the educational tests' reliabilities and to receive more information on the structure of tests, applying item psychological properties analysis would release detailed evidence about the nature of the tests. Item analysis is an advanced approach to measure the quality of the examination by analyzing the properties of the test items. The item analysis applies a strong and specific statistical approach to illustrate statistical indices like item difficulty and item discrimination for examining the quality of the test items (Afraa et al, 2018).

Moreover, by item analysis, the items' psychological characters can be observed and the IA drives out, detailed information about test items. Furthermore, item analysis can identify the items with high difficulty and easiness index and offers effective and vital recommendations to change the inappropriate items (Kennedy & Ato Kwamina, 2017). Additionally, the item analysis method guides the test developers and teachers to take more thoughtfulness about developing test items in future tests for developing standard test items with a high level of accuracy from the test results. In other words, if the test items are designed appropriately, the test is able to assess educational goals properly. Furthermore, through the item analysis, teaching-learning and educational obstacles and difficulties can be identified better. In other words, vital feedbacks on the item analysis

outcomes would provide essentials suggestions and changes to avoid the iteration of unwanted mistakes in future exams (Afraa et al, 2018).

II. PSYCHOLOGICAL PROPERTIES OF THE ITEMS

The difficulty and psychological properties of test items can be analyzed through item analysis statistics, which is recognized as the most powerful method to analyze the test items in educational assessment and psychometrics subjects. The difficulty of the items indicates how well the examinees are able to respond to test items correctly or incorrectly (Sireci et al., 2006). A test item can make discriminate of students' abilities among individuals within the test-takers. If in a test, the nature of items shows high difficulty or easiness, therefore, these items cannot make a distinction amongst individuals (Revelle, 2017). In the present study, two different item analysis methods were applied for evaluating Omani and Iranian TIMSS 2015 fourth-grade mathematics items' psychological properties. The first method was based on the IRT item analysis approach and the second was based on the CDM item analysis approach (psych).

The IRT approach would be able to drive the excellent inferences to detect the difficulty and discrimination indices of the test items. In other words, the IRT approach provides an opportunity for assessing and evaluating the examinees' and the items' characters respectively. There are two indices for measuring items' psychological properties in the IRT approach which their name could be referred like: 1. Difficulty index and 2. Discrimination index (Arifin, & Yusoff, 2017). The difficulty index indicates how many examinees have answered the items correctly, while the discrimination index referred to a function defines that how an item discriminates students' abilities in a test item (Albano, 2018). The interpretation of the difficulty and discrimination indices described that a low proportion of the difficulty index illustrates that the item is too difficult and only a few examinees would be able to answer it correctly. In other words, the high proportion of the difficulty index shows that most of the examinees could give a correct answer to an item regardless of their abilities. The criteria of the difficulty index are described between 0.20 and 0.80 and the criteria for the discrimination index are illustrated from 0.20 and 0.40. In other words, if the item's discrimination index is between 0.20 to 0.40, the item can discriminate the students' abilities from each other while if the item's discrimination index shows an index below 0.20, the item is not able to discriminate the students with high and low abilities from each other. Furthermore, a discrimination index above 0.40 illustrates that the items could be ideal. Siri and Freddano (2011) pointed out that the item analysis provides more fundamental details about test items for teachers and test designers to revise and make more changes on future tests.

III. OBJECTIVES

1. To study psychological properties of the fourth-grade mathematics test items of Omani and Iranian students.
2. To compare the IRT and the CDM item analysis approach for analyzing the psychological properties of the fourth-grade mathematics items.

IV. RELATED WORK

Hassan & Hod (2017) pointed out that if a test item in shows a high difficulty and poor discrimination index, the item is not considered an inappropriate item and should be revised. However, they stated that the major reason for the increasing number of non-functional distractors and low distractor efficiency reflects the teachers' low skills in developing plausible distractors for single best answer items. Toksöz & Ertunç (2017) showed, about 28% of the items had a low discrimination index, which indicates, the item could not make a distinction among students' abilities and the items should be changed or removed from the test. Furthermore, the study indicated that multiple-choice items were efficient which meant the items' difficulty indices were acceptable. Moreover, the results of the study suggested that the teachers or test-developers should participate in some relevant workshops to improve their test-developing skills. Suskie (2017) concluded that the item analysis is a strong approach to reveal the reliability, difficulty, and discrimination index of the test items that can guide teachers and instructors to revise inappropriate and inefficient items for further examination. Taib & Yusoff (2014) stated the optimal difficulty index could be suggested a range from 0.40 to 0.80 as they applied the criteria for their study's analysis. Furthermore, according to Taib & Yusoff, a difficulty index above 0.80 may indicate an easiness for an item. In other words, if the difficulty index is above 0.80, therefore, the item is known as a very easy item and the item should be revised. Kennedy & Ato Kwamina (2017) identified the item difficulty and discrimination values provide a technique for instructors to examine the constructed tests that whether are they able to assess educational goals in a specific subject or not. Khan (2015) pointed out, there are no completed and clear guidelines for the item analysis approach. They stated that applying item analysis can lead instructors and teachers to design and formulate formative questions for educational exams. The study also suggested that the results of the item analysis of the conducted studies should be implied in the test-designing process to avoid making unwanted mistakes. However, to determine the difficulty and discrimination criteria for analyzing the items, Afraa et al (2018) indicated that a difficulty index below 0.20 shows that the item is most difficult. Furthermore, Afraa showed a difficulty index from .025 to 0.70 would be defined that the items have an acceptable and good index and an index above 0.70 shows an easiness for the items. Taib & Yusoff (2014) suggested a criterion from 0.20 to 0.80 is an acceptable and good index. Karkal & Kundapur (2016) recommended a difficulty

index between 0.30-0.70 as a good index and an index above 0.50 to 0.70 as an optimal index. In other words, an item with a difficulty index below 0.30 is known as a difficult item and if the item's difficulty index is above 0.70, the item is very easy.

V. METHODS

The present study is secondary analysis-based research that implies secondary data. The data have been taken from the TIMSS 2015 mathematics fourth-grade assessment. All Omani and Iranian fourth-grade students who participated in the TIMSS 2015 mathematics test have been selected as the statistical sample for this study. Descriptive information of the samples has been explained in table 1.

Table 1 descriptive information of the statistical sample

TIMSS 2015 fourth-grade students' information					
Oman			Iran		
Sex	Frequency	Percent	**Sex**	Frequency	Percent
Girls	4524	49.7	**Girls**	1863	48.7
Boys	4581	50.3	**Boys**	1960	51.3
Total	9105	100.0	**Total**	3823	100.0

Table1demonstrates the statistical sample's descriptive information. As indicated in table 1, 4524 (49.7) girls and 4581 (50.3) boys participated in the TIMSS 2015 mathematics fourth-grade test from Oman. However, there were 1863 (48.7) girls and 19680 (51.3) Iranian students in the TIMSS 2015 test. The descriptive information of the samples showed that in connection with the statistical population of the study, the Omani statistical population is significantly greater than Iran. To analyze the psychological properties of the items, 68 mathematics fourth-grade items were selected for analyzing item difficulty and item discrimination indices. The TIMSS 2015 assessment items have not been released for analysis purposes, therefore, the researcher decided to select 68 items that were released by the International Association for the Evaluation of Educational Achievement (IEA) for item analysis purposes, the IRT and the CDM approaches were applied. To analyze the data, the psych package (Revelle, 2017) through the r program has been applied.

Psych Package

For analyzing the psychometric, personality, and psychological related subjects, the "psych" package has been introduced by the Northwestern University in 2005. The "psych" package is including of some the following functioning which is more useful and listed as: "read.file, read.clipboard, describe, pairs.panels, scat-ter.hist, error.bars, multi.hist, bi.bars" (Ibid, 2017).
In the present work the following functions wee applied for item difficulty:

The related "functions" for running the item difficulty analysis based on the IRT approach:

```
#Load the package:
library (ltm)
#run the following functions:
data<-as.matrix(replicate(68,sample(0:1,1000,rep=TRUE)))
IRT2pl <-ltm(data~z1,IRT.param=TRUE)
coef(IRT2pl,prob=TRUE)
```

Load the "ltm" package on r programming. Where the "data" is your data set name. The number 68 represents the number of items that are being analyzed in the present study. The "coef" function will provide the difficulty and discrimination parameters for each item.

The related "functions" for running the item difficulty analysis based on the CDM approach:

```
#load the "psych" package:
library(psych)
#run the following functions:
difficulty.data<-colMeans(data,na.rm=TRUE)
difficulty.data
subset(difficulty.data,(difficulty.data <.2|difficulty.data >.8))
```

The "data" shows your data set name, the defined numbers of ".2" and ".8" indicate the difficulty cut off for the items. In other words, the ".2" illustrates the most difficult items whose indices are below ".2" and the cut-off ".8" defines the easiest items whose indices are above ".8".

VI. FINDINGS

To analyze the psychological properties of the Omani and Iranian students' TIMSS 2015 mathematics items, the CDM package in r programming, version 4.0.3, was applied.

Item Difficulty Analysis based on the IRT Approach

Table 2 indicates the criteria index for the IRT approach.

Table 2 the criteria for difficulty and discrimination indices based on the IRT approach

Difficulty	Discrimination
<0.20 : Difficult	< 0.20: Poor
0.20 to 0.80: Ideal	0.20 to .40: Acceptable
> 0.80: Easy	>0.40: Very Good

(Arifin & Yusoff, 2017).

Table 2 determines the criteria for the difficulty and discrimination indices as Arifin & Yusoff cited (2017). The difficulty index displays that an index below 0.20, indicates that the item is considered as too difficult items and should be revised or removed from the test. Moreover, if an item's difficulty index will appear above 0.80, the item's difficulty index shows that the item is identified by way of a very easy item and it should be deleted. Consequently, an item difficulty index "between" 0.20 to 0.80 represents that the item is ideal. The discrimination index of the test items specifies that the extent to which an item would be able to provide discrimination between the examinees with high scorers and the examinees who achieved low scorers in a test. The discrimination index is represented as a fraction and varies between -1 to 1. Optimally, an appropriate item should be included with a positive discrimination index of at least 0.20, which indicates, the test-takers with high scorers have a high probability fortuitous for answering to the test items correctly and the test-takers with low scorers have a low probability chance for giving correct response to the given items. However, the Items with negative indices should be considered to determine whether the item was flawed or miskeyed. Hence, the items with a discrimination index above 1, indicate that these items are known as good and appropriate items which can discriminate the examinees with low and high abilities from each other (George & Robitzsch, 2015). The outcomes of the psychological properties for mathematics fourth-grade items in the TIMSS 2015 for Omani and Iranian students are presented in table 3.

Table 3 Psychological properties of the Items for Omani and Iranian students Based on the IRT approach

Item	Omani Students Mathematics Items			Iranian Students' Mathematics Items		
	Difficulty	Discrimination	$P(x=1|z=0)$	Difficulty	Discrimination	$P(x=1|z=0)$
1	-0.676	1.289	0.705	-0.439	1.341	0.643
2	-0.092	1.391	0.532	-0.472	1.856	0.706
3	0.308	1.164	0.411	0.935	1.189	0.247
4	0.832	1.328	0.248	1.072	1.352	0.189
5	0.493	2.268	0.245	0.123	1.233	0.462
6	-0.038	1.227	0.511	-0.623	1.809	0.755
7	0.508	1.410	0.327	0.299	1.431	0.394
8	0.554	1.844	0.264	-0.251	1.663	0.602
9	0.077	1.059	0.479	1.593	0.824	0.211
10	1.003	1.030	0.262	1.825	0.696	0.219
11	1.132	0.831	0.280	0.678	0.733	0.378
12	-0.677	1.313	0.708	2.089	0.356	0.321
13	0.268	1.658	0.390	0.427	1.525	0.342
14	-0.918	1.033	0.721	-1.179	1.202	0.805
15	0.440	1.232	0.367	0.251	1.443	0.410
16	1.297	1.610	0.110	0.710	1.369	0.274
17	2.989	1.539	0.009	1.900	2.405	0.010
18	1.945	0.959	0.134	3.919	0.605	0.085

19	3.317	0.339	0.245	3.620	0.424	0.176
20	-1.503	0.175	0.565	1.325	0.680	0.288
21	-0.259	0.942	0.560	0.254	0.826	0.447
22	0.913	0.881	0.308	-0.053	1.043	0.513
23	-0.338	1.044	0.587	-0.131	1.269	0.541
24	1.283	0.878	0.245	2.435	0.439	0.255
25	1.770	2.385	0.014	1.463	2.635	0.020
26	1.429	1.032	0.186	1.278	1.317	0.156
27	-0.015	1.376	0.505	-0.152	1.406	0.553
28	1.011	0.346	0.413	0.722	0.388	0.430
29	2.665	0.688	0.137	3.962	0.762	0.046
30	1.004	0.827	0.303	0.517	1.172	0.353
31	1.639	1.489	0.079	1.576	1.779	0.057
32	1.960	0.832	0.163	1.686	0.868	0.187
33	0.838	1.050	0.293	1.087	0.729	0.311
34	-1.128	0.842	0.721	-1.529	0.807	0.774
35	1.198	1.159	0.199	0.822	1.481	0.228
36	2.529	0.386	0.273	2.290	0.524	0.231
37	0.369	0.987	0.409	-1.332	0.877	0.763
38	0.124	0.943	0.470	0.421	0.524	0.444
39	-0.653	1.858	0.771	-1.064	1.416	0.818
40	-0.546	1.854	0.733	6.650	0.243	0.165
41	0.730	1.019	0.322	0.730	0.575	0.396
42	1.791	0.351	0.347	1.359	0.459	0.348
43	2.008	0.865	0.149	4.274	0.832	0.027
44	2.095	1.376	0.052	3.220	0.857	0.059
45	1.978	0.519	0.263	1.436	0.740	0.256
46	1.556	2.190	0.032	1.995	2.283	0.010
47	0.506	0.922	0.385	0.860	0.755	0.343
48	3.935	0.406	0.168	2.046	0.616	0.220
49	0.429	1.388	0.355	-0.487	1.745	0.700
50	0.393	1.130	0.390	0.357	1.486	0.370
51	0.594	0.595	0.412	0.149	0.800	0.470
52	1.149	1.703	0.123	4.132	1.061	0.012
53	-0.620	1.619	0.731	-0.448	1.473	0.659
54	0.291	0.818	0.440	1.450	0.479	0.332
55	-0.615	1.498	0.715	-1.076	1.441	0.825
56	0.794	0.666	0.370	1.189	0.684	0.306
57	2.044	0.415	0.299	2.161	0.419	0.287
58	-0.040	1.965	0.519	-0.561	1.793	0.732
59	0.137	1.437	0.451	0.625	1.968	0.226
60	1.343	0.508	0.335	1.246	0.430	0.368
61	0.712	0.714	0.375	2.303	0.373	0.297
62	2.252	0.548	0.225	4.547	0.315	0.192
63	0.612	0.869	0.370	1.180	0.671	0.311
64	1.113	1.222	0.204	0.489	1.169	0.360
65	0.402	0.899	0.410	-0.163	1.263	0.551
66	0.049	1.417	0.482	0.523	1.136	0.355
67	1.359	1.731	0.086	1.073	1.627	0.148
68	2.931	0.793	0.089	1.948	1.489	0.052

The represented information of table 3, show that the Omani students' item analysis results revealed, the items 9, 38, 66, and 59 difficulty index was found below 0.20 which indicates these items were recognized as too difficult items and these items were not identified as suitable and appropriate items among the test items (left side). Items 4, 22, and 33 showed a difficulty index above 0.80 which were considered easy items. In other words, only 4 items were found as the most difficult items as well as 3 items were identified as the easiest items among Omani students. Furthermore, the items with difficulty indices above "1" were identified such: (items 10, 11, 16,17,18,19, 24, 25, 26, 28, 29, 30, 31, 32, 35, 36, 42, 43. 44, 45, 46, 48, 52, 57, 60, 62, 64,

67 and 68). However, the negative values illustrate that these types of items have essential issues such as miskeyed or other major flaws. Moreover, some items with a difficulty index ranged from 0.20 and 0.80 were identified which were: 3, 5, 7, 8, 13, 15, 37, 40, 41, 47, 49, 50, 51, 56, 61, 63, and 65. Besides, the items 11, 18, 21, 22, 24, 29, 30, 32, 34, 37, 38, 43, 45, 47, 48, 51, 54, 56, 57, 60, 61, 62, 63, 65 and 68 had a high discrimination values, which illustrated these items were able to discriminate low score and high score students' abilities from each other. Nonetheless, items 19, 28, 36, 42, had an acceptable discrimination index. Consequently, there was just an item with a discrimination value below 0.20, which was the "item number 20".

Conversely, the outcomes of the psychological properties of Iranian students' mathematics test's items have been presented in table 2. The results revealed that the items 5, 12, 51, 52, with a difficulty index below 0.20 were found as most difficult items among Iranian students. Additionally, the items 3, 35, 47, showed a difficulty index above 0.80 which indicating these items were the easiest items among Iranian students' responses to the Mathematics items. Consequently, for the difficulty index from 0.20 and 0.80 the items 7, 11, 13, 15, 16, 21, 28, 30, 38, 41, 50, 59, 64, 66, were recognized appropriate items. Though, the items: 4, 9, 10, 12, 17, 18, 19, 20, 24, 25, 26, 29, 31, 32, 33, 36, 40, 42, 43, 44, 45, 46, 48, 52, 54, 56, 57, 60, 61, 62, 63, 67 and 68 with a difficulty index above 1 were categorized as appropriate items. In other words, these items' difficulty index was acceptable. Nonetheless, the items 9, 10, 11, 18, 19, 20, 21, 24, 29, 32, 33, 34, 36, 37, 38, 41, 42, 43, 44, 45, 47, 48, 51, 54, 56, 57, 60, 63 were considered as good items. Furthermore these items showed very good discrimination indices. Consequently, the following items had a difficulty index above "**1**": 1, 2, 3, 4, 5, 6, 7, 8, 13, 14, 15, 16, 17, 22, 23, 25, 26, 27, 30, 31, 35, 39, 46, 49, 50, 52, 53, 55, 58, 59, 64, 65, 66, 67 and 68. Accordingly, the items 12, 28, 40, 61 and 62 had an acceptable difficulty index among Iranian students in the TIMSS 2015 fourth grade mathematics items.

Item Difficulty Analysis Based on CDM Approach

Table 4 represents the item difficulty analysis results based on the CDM item analysis approach. The CDM-based item analysis approach's criteria for determination of the difficulty index has defined a range from 0.20 and 0.80. The defined criteria lead to interpreting the difficulty index such as if the difficulty index shows an index below 0.20, the item is categorized as a most difficult item. Hence, if an item's difficulty index present above 0.80, therefore the item would be considered as an easy item.

Table 4 Item Difficulty Based on CDM

Item Difficulty Based on CDM			
Oman		Iran	
Item	Difficulty	Item	Difficulty
1	0.658	1	0.608
2	0.521	2	0.636
3	0.428	3	0.295
4	0.302	4	0.252
5	0.344	5	0.472
6	0.507	6	0.675
7	0.370	7	0.425
8	0.339	8	0.572
9	0.482	9	0.239
10	0.298	10	0.239
11	0.305	11	0.391
12	0.659	12	0.326
13	0.422	13	0.389
14	0.685	14	0.752
15	0.396	15	0.434
16	0.187	16	0.326
17	0.027	17	0.064
18	0.168	18	0.097
19	0.250	19	0.184
20	0.565	20	0.305
21	0.550	21	0.454
22	0.333	22	0.510
23	0.571	23	0.531
24	0.274	24	0.263
25	0.078	25	0.111
26	0.227	26	0.216
27	0.490	27	0.515

28	0.412	**28**	0.424
29	0.154	**29**	0.055
30	0.319	**30**	0.363
31	0.137	**31**	0.118
32	0.187	**32**	0.207
33	0.318	**33**	0.317
34	0.690	**34**	0.736
35	0.240	**35**	0.274
36	0.277	**36**	0.235
37	0.416	**37**	0.719
38	0.466	**38**	0.437
39	0.671	**39**	0.734
40	0.642	**40**	0.165
41	0.347	**41**	0.402
42	0.351	**42**	0.354
43	0.178	**43**	0.036
44	0.097	**44**	0.077
45	0.274	**45**	0.277
46	0.113	**46**	0.057
47	0.399	**47**	0.358
48	0.175	**48**	0.236
49	0.385	**49**	0.635
50	0.407	**50**	0.404
51	0.417	**51**	0.473
52	0.205	**52**	0.020
53	0.661	**53**	0.615
54	0.446	**54**	0.340
55	0.664	**55**	0.773
56	0.386	**56**	0.334
57	0.309	**57**	0.301
58	0.516	**58**	0.680
59	0.469	**59**	0.341
60	0.347	**60**	0.382
61	0.392	**61**	0.308
62	0.241	**62**	0.201
63	0.392	**63**	0.338
64	0.261	**64**	0.407
65	0.428	**65**	0.559
66	0.492	**66**	0.402
67	0.174	**67**	0.246
68	0.112	**68**	0.111

Table 4 illustrates, no easy items were found based on the CDM analysis outcomes for both Omani and Iranian students. It would be concluded that the CDM items analysis method explains better the item's psychological properties than the IRT approach. Furthermore, to find out the most difficult items, a difficulty index of 0.20 was between Oman and Iranian students was applied. Table 5 illustrates the items with a difficulty index below 0.20.

Table 5 Items with difficulty index below 0.20 based on CDM

Items with difficulty index below 0.20 based on CDM			
Oman		**Iran**	
Items with High Difficulty	**Difficulty**	**Items with High Difficulty**	**Difficulty**
16	0.187	**17**	0.064
17	0.027	**18**	0.097
18	0.168	**19**	0.184
25	0.078	**25**	0.112
29	0.154	**29**	0.055
31	0.137	**31**	0.118
32	0.187	**40**	0.165

43	0.178	**43**	0.036
44	0.097	**44**	0.077
46	0.113	**46**	0.057
48	0.175	**52**	0.020
67	0.174	**68**	0.111
68	0.113		

The outcomes of the CDM analysis to find out the most difficult items with a difficulty index below 0.20 showed that items 17, 18, 25, 29, 31, 43, 44, 46, and 68 were the most difficult items for both Omani and Iranian students. However, the Omani students' performance was greater in item 18(DI:Oman=0.168; Iran=0.097), item 29(DI:Oman=0.154; Iran: 0.055), item 31(DI:Oman=0.137; Iran=0.118), item 43(DI:Oman=0.178; Iran=0.036), item 46(DI:Oman=0.113; Iran=0.057). Moreover, Iranian students' performance showed that they achieved a higher rank than Omani students on item 25 (DI: Iran=0.112; Oman=0.078).

VII. CONCLUSION

Item analysis stands as an essential assessment approach that can drive additional significant information about the psychological properties of the test. In the present study, two different item analysis approaches have been applied to make a comparison outlook to find out which approach is able to analyze the Omani and Iranian students' mathematics items' psychological properties better. The IRT analysis outcome revealed, there are just 4 difficult items with a difficulty index below 0.20 for Omani and Iranian students' mathematics test in the TIMSS 2015 fourth grade. In other words, just 4 items were recognized as the most difficult items in Omani and Iranian students' test items. However, the items were not the same as each other. Though, the IRT analysis indicated that the following items had a difficulty index above 1 for both country: 10, 18, 19, 24, 25, 26, 29, 31, 32, 36, 42, 43, 44, 45, 46, 48, 52, 57, 60, 62, 67 and 68. Likewise, the IRT approach could find the same items which had a difficulty index between 0.20 and 0.80 for both countries which the items are as follows: 7, 13, 15, 41, and 50. Moreover, the CDM approach analysis outcomes showed that the items: 17, 18, 25, 29, 31, 43, 44, 46, and 68 were the most difficult items for Omani and Iranian students mutually. Furthermore, the IRT analysis included two difficulty and discrimination indices which the difficulty index, which indicates an easiness and difficulty index of the items, where the discrimination index reveals a separation between the students with low abilities and high abilities in an item. Therefore, the CDM item analysis approach outcomes discovered that the Omani students' difficult items were greater than Iranian students. Besides, the IRT item analysis approach drove out a major number of items with high discrimination index that the items were able to discriminate students' abilities from each other. In other words, these items could discriminate the students with low ability and the students with high ability from each other (Oman: 25 items and Iran 28 items). However, 22 items out of 68 items were recognized as very easy items for Oman and Iran (Difficulty index >0.80). In other words, the IRT item analysis approach showed, just 5 items were recognized as appropriate items for Omani and Iranian students which the items were the same. Additionally, the CDM item analysis approach found 9 items as the most difficult items for both Omani and Iranian students. Nevertheless, there were 22 same items with an index above 1 which are known as very easy items for both countries. Consequently, the findings of the IRT approach revealed the same most difficult items for Omani and Iranian, while the CDM approach showed that the most difficult items of Omani students were greater than Iranian. In other words, Omani students had more issues with the mathematics items than Iranian students. Therefore, the CDM analysis approach could analyze the items' psychological properties better than the IRT approach.

The IRT item analysis approach revealed that according to the Omani students' test's item difficulty results, among the most difficult items, the topic area of item No. 5 was "Whole Numbers" and its "cognitive domain' was identified as "applying". Besides, item number "5" was presented in the 4th cycle of the TIMSS in the year 2007. item No. 9 "cognitive domain" was illustrated as "knowing" and the item was offered in the TIMSS 2011 for the first time. The item's topic area was "Geometric Shapes and Measures". item No. 68 "cognitive domain" was "reasoning" and it was presented in the TIMSS 2007 before presenting in the TIMSS 2015. The topic area of the item was found "Reading, Interpreting, and Representing".

Moreover, the Iranian students' mathematics test's items' item difficulty results showed that among the most difficult items, the item No. 5 topic area was "Whole numbers" and its cognitive domain was identified "Applying". The items No. 51 and 52 cognitive domain was recognized "Knowing" and the topic area of the items were found "Points, Lines, and Angles". However, item No. 51 was presented for the first time in the TIMSS 2007 assessment, and item No. 52 appeared in the TIMSS 2011.

Moreover, The CDM item analysis method identified that the item 17, 18, and 25 cognitive domain was "knowing" and the item's topic area was "Fractions and Decimals". Hence, all three items were offered in the TIMSS 2007 for the first time. Furthermore, the "cognitive domain" of item 29 was "applying and the topic area of the item was determined "Expressions, Simple Equations, and Relationships". Besides, the item appeared initially in the TIMSS 2007. The items 43, 44, and 46 were presented in the TIMSS 2007. Hence, the topic area of the items was "Two- and Three-dimensional Shapes", and the

"cognitive domain" of the item were identified "reasoning". The cognitive domain" of item 68 was reasoning and the item appeared in 2007 in the TIMSS assessment. The topic area of the item was "reading, interpreting, and representing".

The item analysis approach should be considered as a strong method to analyze the items and make any probability changes or revisions of the test. Teachers and test developers should get more familiarity with item analysis methods to lead students for taking a reliable test. Since this test has been designed by the IEA which is the TIMSS test provider across the words, the items' psychological properties seemed good. Consequently, many items could make high discrimination for both Omani and Iranian students' responses.

VIII. SUGGESTIONS

This study focused on Omani and Iranian fourth-grade students' mathematics test items who participated TIMSS 2015 test. Therefore, the generalization of this study may be limited. However, the results of this study could lead instructors to conduct the same study to increase the test accuracy. The major concern can be bold to make a judgment on students' abilities based on test results which they maybe have been designed inappropriately.

REFERENCES

[1] Arifin, W. N., & Yusoff, M. S. B. Item Response Theory for Medical Educationists. Education in Medicine Journal, 9(3) (2017) 69–81. https://doi.org/10.21315/eimj.

[2] George, A. C., & Robitzsch, A. Cognitive Diagnosis Models in R: A didactic. The Quantitative Methods for Psychology, 11(3) (2015) 189-205. doi: 10.20982/tqmp.11.3.p189.

[3] Hassan, S., & Hod, R. Use of Item Analysis to Improve the Quality of Single Best Answer Multiple Choice Question in Summative Assessment of Undergraduate Medical Students in Malaysia. Education in Medicine Journal, 9(3) (2017) 33–43. https://doi.org/10.21315/eimj2017.9.3.4

[4] Karkal, Y., & Kundapur, G. Item analysis of multiple choice questions of undergraduate pharmacology examinations in an International Medical School in India. Journal of Dr. NTR University of Health Sciences, 5(3) (2016) 183. https://doi.org/10.4103/2277-8632.191842.

[5] Khan, G. Using Item analysis on essay types questions given in summative examination of medical college students: Facility value, Discrimination index. International Journal of Research in Medical Sciences, 1(2015). https://doi.org/10.5455/2320-6012.ijrms20150131.

[6] M. Patel, R. Use of Item analysis to improve quality of Multiple Choice Questions in II MBBS. Original Research Article, 4(1) (2017) 22–29. https://doi.org/10.18231/2393-8005.2017.0007

[7] Musa, A., Shaheen, S., Elmardi, A., & Ahmed, A. Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. Khartoum University Medical Journal, 11(12) (2018) 1477–1486. https://www.researchgate.net/publication/328583573.

[8] Quaigrain, K., & Arhin, A. K. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education, 4(1) (2017) 1301013. https://doi.org/10.1080/2331186x.2017.1301013.

[9] Revelle, W. Psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version=1.7.5 (2017).

[10] Sireci, S. G., & Parker, P. Validity on Trial: Psychometric and Legal Conceptualizations of Validity. Educational Measurement: Issues and Practice, 25(3) (2006) 27–34. https://doi.org/10.1111/j.1745-3992.2006.00065.x

[11] Siri, A., & Freddano, M. The Use of Item Analysis For The Improvement Of Objective Examinations. Procedia - Social and Behavioral Sciences, 29 (2011) 188–197. https://doi.org/10.1016/j.sbspro.2011.11.224(2011) .

[12] Taib, F., & Yusoff, M. S. B. Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. Journal of Taibah University Medical Sciences, 9(2) (2014) 110–114. https://doi.org/10.1016/j.jtumed.2013.12.002.

[13] Tobin, Mary Ann. A guide to item analysis. PhD, Assistant Research Professor and Instructional Consultant Schreyer Institute for Teaching Excellence. The Pennsylvania State University. https://www.schreyerinstitute.psu.edu(2018).

[14] Toksöz, S., & Ertunç, A. (2017). Item Analysis of a Multiple-Choice Exam. Advances in Language and Literary Studies, 8(6), 141. https://doi.org/10.7575/aiac.alls.v.8n.6p.141.

[15] Tony D. A. Introduction to Educational and Psychological Measurement Course Notes, Course notes. University of Nebraska-Lincoln(2018).

[16] XU, Y., & LIU, Y. Teacher Assessment Knowledge and Practice: A Narrative Inquiry of a Chinese College EFL Teacher's Experience. TESOL Quarterly, 43(3) (2009) 492–513. https://doi.org/10.1002/j.1545-7249.2009.tb00246.x.