Development and Validation of Computer Subject Test for Tenth Grade Jordanian Students, According to the Partial Credit Model

Anwar Bani Hani[1], Rohaya Talib[2]

[1] PhD Student Department of Educational Technology, Universiti Teknologi Malaysia

[2]Senior Lecturer, School of Education, Universiti Teknologi, Malaysia, Johor Bahru, Malaysia.

Corresponding Author: Anwar Bani Hani, School of Education, Universiti Teknologi Malaysia, Malaysia Email: anwar.hani91@yahoo.com

Abstract **-The study aimed at the development and validation of computer subject test for 10th –grade students according to the Rasch partial credit model (PCM) by using the descriptive approach as it is appropriate for the study aims. To achieve the study's objective, the test was constructed of the essay type, and it consisted of 25 items based on the (IRT) according to the Rasch PCM. The first administration of the test was conducted to verify the validity and reliability of the test. To verify the "face validity" of the test's objectives, they were presented to a group of 14 arbitrators who work as teachers and educational supervisors. They found that the contents are representative of the level of the goal, which is pursuing in theory. The empirical reliability was calculated for the test, where the value of person reliability reached 0.91. Moreover, the items reached 0.93. The study population consisted of all 10th-grade students at the schools belonging to the "Directorate of Education of Irbid District," whose numbers were 7365, represented by 3612 male students, and 3753 female students. According to the class regarding their sex (gender), a sample was chosen according to a cluster as the test unit was the class section. The sample size of the study was 1000 distributed to 490 male students and 510 female students. Besides, this study's findings have brought several issues concerning computer subject (CS) achievement by verifying the tests and reliability and accomplishing the IRT's suppositions according to the PCM.**

Keywords: **Computer Subject Test (CST), Item Response Theory (IRT) , Partial Credit Model (PCM).**

I. *INTRODUCTION*

Construction and validation of tests, especially academic achievement measures, contain complicated steps, procedures, the interrelationship of various ideas and latent variables. Subsequently, confirmed procedures must be followed to develop a test that is firmly identified with the expected outcomes. The two most essential steps in test development as spelt out by [1] are; (i) Item development, which includes content definition, preparation of test specifications, preparation of the item pool, content validation/experts judgment, pilot testing of the items, data analysis, and revision of test items. (ii) Item validation through item analysis. All these explained processes are closely linked with others. Additionally, these processes are carefully accomplished to ensure the validity and reliability of the instrument developed and used to estimate item and a person's ability. Validity is concerns about how assessment systems are built. Whether the assessment tool (Test) is standardized or locally-designed, the aim is to use an instrument that produces a true estimate of the examinee ability that could support valid inferences. The purpose of assessing student's learning includes licensing, certification, diagnosis, and placement.

The field of educational and psychological assessment and evaluation has received increased research attention by psychologists and educators. The primary objective of this field was to reveal individual differences of all kinds, whether inter-individual differences between groups or intra-individual differences. Measuring methods and instruments have been varied to achieve this goal in which the quality of the assessment depends on the quality of performance and the quality of the measurement process in Classical Test Theory (CTT). These efforts have led to the transition from the CTT used in the design of the tests, which have been used for a long time in the educational and psychological evaluation, to the modern approach, IRT, or the Latent Trait Theory (LTT) [2].

CTT being a traditional theory, still attract the measurement community in test development and analysis due to its theoretical and practical simplicity. The continuous application of CTT in item analysis is because of its "weak assumptions," which can easily be met by test data [3],[4]. Although, as a result of its continuous utilization, researchers have questioned it’s in the present-day measurement community [5]. The PCM could be a one-dimensional model for the analysis of responses recorded in two or more ordered classes, as well as Samejima's graded response model (GRM) [6]. The PCM differs from the GRM. However, therein it belongs to the Rasch family of models, then share the identifying characteristics of that family: separable person and item parameters, adequate statistics, hence, integrated additively. These characteristics modify "specifically objective" comparisons of persons and items [7] and permit every set of model parameters to be conditioned out of the estimation procedure for others.

Educators have been interested in teaching computers because of the goals that contribute to the education of adolescents and provide them with the knowledge side that forms part of their general culture [8], [9], [10]. The skills that qualify them to coexist in a technological environment [11] and develop their mental skills and the ability to learn through the use of computers as an educational tool. Additionally, given the importance of computers in developing societies, and to meet the needs of contemporary life, they came that the goals computer teaching in Jordanian schools is balanced with the requirements of the educational process.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

The researchers and computer teachers noticed the fear of the students of essay questions at the expense of the multiple-choice questions, as they prefer the multiple-choice questions much more of the essay/structural questions. Also, they have noticed the lack of studies that have addressed the computer curriculum in particular according to the PCM, which aims to determine the difficulty coefficient for each step, while answering the items of polytomous responses, which is considered as a generalization of the Rasch model in the dichotomous responses item [12]. As a consequence, this paper identified the problem in general in an attempt to select items from achievement test in the subject of computer, specifically for students in the 10th grade, because this stage is the transition from the primary stage to the secondary stage, also to demonstrate the importance and effectiveness of the (PCM) in achievement tests or other tests. The test has psychometric characteristics so that it can be applied and used in public and private schools. This study was designed to provide a (CST) for students in the 10th grade, according to the PCM.

The purpose of this study was to develop and validate (CST) for 10th-grade Jordanian students, according to the (PCM), and depend on the Rasch model (IRT). The two models were utilized to obtain valid and reliable test items relevant to measure the true ability of students from traditional and modern measurement perspectives. The analysis was conducted to determine the appropriate items that satisfied specific criteria for item quality. In light of these and many concerns, this study was conducted to investigate the nature of the IRT item parameters for CST in Jordan. Broadly speaking, IRT models can be divided into two families: unidimensional and multidimensional. Unidimensional models require a single trait (ability) dimension θ. Multidimensional IRT models model response data hypothesized to arise from multiple traits. However, because of the greatly increased complexity, the majority of IRT research and applications utilize a unidimensional model. IRT models can also be categorized based on the number of scored responses. The typical multiple-choice item is dichotomous; even though there maybe four or five options, it is still scored only as correct/incorrect (right/wrong).

A. Number of IRT parameters

Dichotomous IRT models are described by the number of parameters they make use of [13]. The 3PL is named so because it employs three item parameters. The two-parameter model (2PL) assumes that the data have no guessing, but that items can vary in terms of location bi and discrimination ai . The one-parameter model (1PL) assumes that guessing is a part of the ability and that all items that fit the model have equivalent discriminations, so that items are only described by a single parameter bi.

B. The Rasch model

The Rasch model is often considered to be the 1PL IRT model. However, proponents of Rasch modeling prefer to view it as a completely different approach to conceptualizing the relationship between data and theory [14] Like other statistical modeling approaches, IRT emphasizes the primacy of the fit of a model to observed data [10] while the Rasch model emphasizes the primacy of the requirements for fundamental measurement, with adequate data-model fit being an important but secondary requirement to be met before a test or research instrument can be claimed to measure a trait [15]. There are some models of the IRT with polytomous responses**:** many different models of the IRT appeared [16], [17]. Each of these models had a specific purpose. These models were mentioned as follows, with some supported studies for them:

C. Normal Ogive Model (NOM):

The NOM was the first IRT model for measuring psychological and educational latent traits [18], [19], [20], [21]. The NOM was refined later by [22]. In the model, an item characteristic curve (ICC) is derived from the cumulative density function (CDF) of a normal distribution. Besides, some studies that applied this model [23], [24].

D. Partial Credit Model (PCM):

The PCM is an extension of the 1PLM, Rasch model [25]. The study of [26] applied this model.

1. Generalized Partial Credit Model (GPCM):

The GPCM [27] is a generalization of the PCM with a parameter for item discrimination added to the model. The study of [28] used this model.

2. Rating Scale Model (RSM):

There are two different approaches to the RSM [29] proposed a response function, in which the values of the category scores are directly used as a part of the function. Another form of the RSM was proposed by [30], which can be seen as a modification of PCM. The recent studies that used this model were [31], [32].

3. Graded Response Model (GRM):

The GRM was introduced by [6] to handle ordered polytomous categories such as letter grading, A, B, C, D, and F, also polytomous responses to attitudinal statements such as a Likert scale .The study of [33] adopted this model.

4. Nominal Response Model (NRM):

The NRM, also called the Nominal Categories Model (NCM), was introduced by [34]. Unlike the other polytomous IRT models introduced above, polytomous responses in NRM are unordered or at least not assumed to be ordered. Even though responses are often coded numerically (for example, 0,1, 2…, m), the values of the responses do not represent some scores on items, but just nominal indications for response categories. There are some applications of the NRM found in uses with multiple-choice items. As for models of polytomous responses, it is used when the response consists of many scores, and each score has a Difficulty Coefficient (DC) according to the used model. One of these models is the (PCM) which identifies DC each step (k) while answering the item (i) of polytomous responses, as well as identifying the latent ability of the person and his performance. There is also the (GRM), each item has a Discrimination Index, and each section of the response has DC [35]. Some recent studies that adopted this model were [36] [37].

II. Research Questions

The following research questions were raised to guide the study:

i. Do the computer test data for tenth- grade students achieved the assumptions of Item Response Theory (IRT)?
ii. To what degree does the computer subject test's data conformity for the tenth-grade students with the Partial Credit Model (PCM)?
iii. What are the estimates of the values of the parameter of the items according to the Partial Credit Model (PCM)?
iv. What are the estimates of the values of the person's ability depending on the model used?
v. What are the criteria of the performance on the test items according to the Partial Credit Model (PCM)?
vi. What are values of test information functions (TIF) at different levels of ability?

III. Materials and Method

Research Design

The study aimed at the development and validation of (CST) for 10th-grade students according to, Rasch (PCM), by using a survey design as it is appropriate for the study aims.

1. *Participants*

Population of the Study

The population of the study consisted of all 10th-grade students at the schools belonging to the "Directorate of Education of Irbid District", whose numbers were 7365, distributed to 3612 male students, and 3753 female students.

2. Sample and Sampling Techniques

The sample in this study was drawn using stratified random sampling technique that chosen according to their gender and then the cluster, as the unit of choice was the classroom division, where 14 schools were divided into 7 male schools and 7 female schools, where the study sample size reached 1000 male and female students. Their answers to the statistical treatments, distributed to 490 male and 510 female students. The instrument is (CST) constructed for 10th-grade students and estimated the a Difficulty Coefficient (DC) following Rasch (PCM). The test consisted of (25 items) of essay type, and each item has multiple answers as each item needs. The test items covered the whole computer subject. The 25 items of the essay type constructed based on the (IRT), according to the Rasch PCM, was administered on the sample of 10th-grade students, which are mainly under the control of their respective schools. Each Item had four answers following the steps of the achievement test. After receiving specific instruction for the test by teachers under the supervision of the monitoring and evaluation unit, responsible for the regulation of primary education in the ministries, in the north area "Irbid district" in February 2020. After Coordinating with the school, management was set a date for a visit to set a date for the application of the study tool. They informed that the information obtained used for scientific research after applying the study tool in its final form on the targeted sample study. Then Collected the questionnaires, auditing and analyzing them statistically, to answer the questions of the study, and came up with appropriate recommendations in the light of the results**.** The data collected were analyzed, by using SPSS V 23. For factor analysis, estimated abilities of the 10th-grade students, mean, standard deviation (SD), standard error (SE), correlation coefficient bi-serial, and percentiles. Moreover, it was used (winsteps V 3.72.3) for the conformity ability of students on the test.

IV. Results

After analyzing the data obtained from the instrument, results were presented in table based on the research questions

1. First Research Question: Do the computer test data for 10th- grade students achieve the assumptions of (IRT)? To answer this question; the factor analysis was conducted using SPSS V23.0, to verify a uni-dimensional assumption of test items, as shown in table 1.

TABLE 1: THE RESULTS OF THE FACTOR ANALYSIS OF THE (CST) ITEMS FOR THE 10TH GRADE STUDENTS.

Some of Square of Saturation			*The Eigen values*			
Cumulative Explanatory variance %	*Explanatory variance*	*Total*	*Cumulative Explanatory variance %*	*Explanatory variance*	*Total*	*Component Number*
47.74	47.74	11.94	47.74	**47.74**	11.94	1
			51.38	**3.64**	0.91	2
			54.22	2.84	0.71	3

Table1 presents the results of the factor analysis of the (CST) items indicated to a uni-dimensional investigation of three indicators; as follow: the result of dividing the Eigenvalue of the first factor by the Eigenvalue of the second is greater than 2. Then, the result of dividing the quotient of the root of the second Eigenvalue from the first one on the quotient of the root of the third one from the second one has a high value, the value of the variance explained of the first component is higher than 20.0% [38].

Figure 1: Showing the Eigenvalues for the factors that make up the test was used with emphasis on a uni-dimensional assumption.

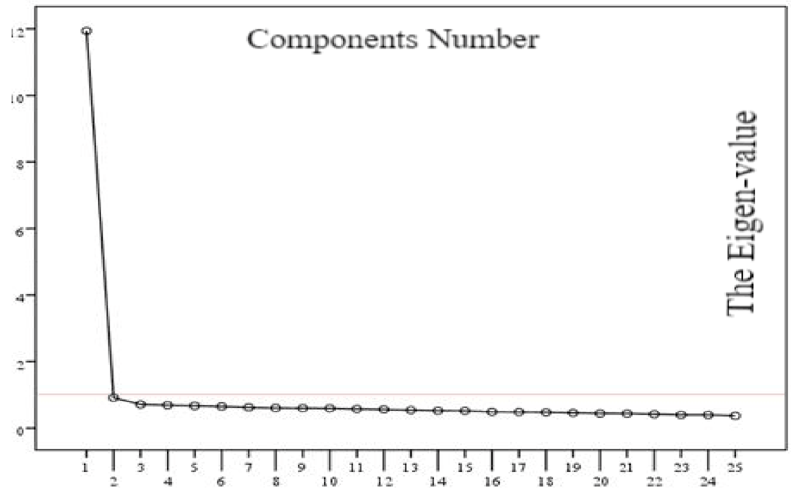

Figure 1: The Plot for Sorting Test of the Values of the Eigenvalues of the Factors of the Test.

TABLE 2: THE FREQUENCIES AND PERCENTAGES OF (LI) OF THE ITEMS OF TEST.

Percent %	*Frequency of Correlational Pairs*	*Status of local independence*
8.33	25	Dependent
91.67	275	Independent
100.00	**300**	**Total**

Table 2: Highlighted the assumption of (LI) for the items test was verified by calculating the standard value ($\chi 2$) of the standardized form of the LI (Standardized LD $\chi 2$), each pair of test items (300) has a correlational pair that is calculated by multiplying (25) items by (24) and then dividing by (2), using the (IrtPro V3.1.21505.4001) software. The frequencies and percentages of both LI cases were then monitored provided that the standard LI value is greater than (10) indicating that the LI of a certain number of correlational pair has not been achieved and vice versa if 10 is lower, indicating that the LI of a certain number of correlational pair has been achieved. Moreover, the table 2 also shows that LI is achieved in 275 a correlational pair of 300 a correlational pair to items of a test in percentage 91.67%.

 https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 2: Highlighted the hypothesis of monotonicity had been verified for the item characteristics curve (ICC) of the test by the construction of a graph showing the Test Score Response Function (TSRF) at each raw score of the (CST) items.

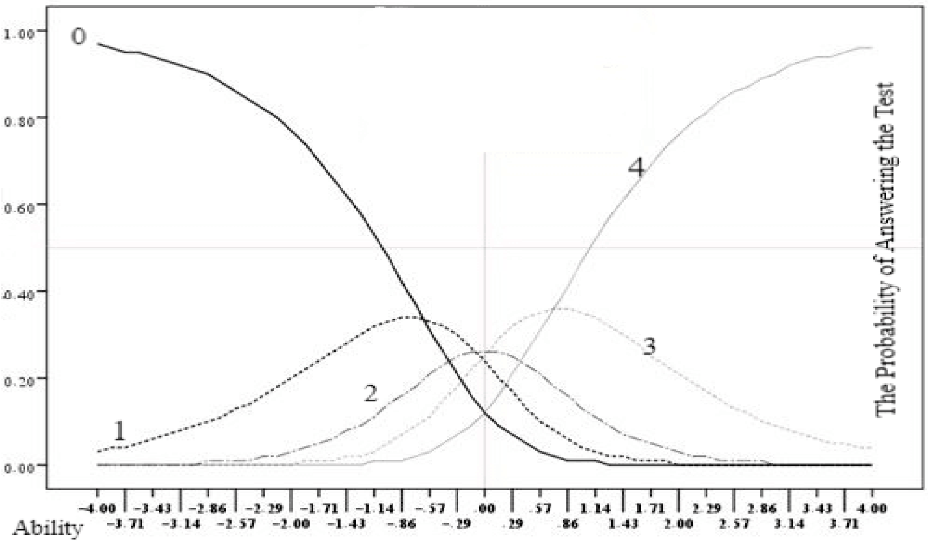

Figure 2: A graph of the Response Function at each Raw Score of a (CAT) Item

Figure 3: To verify the assume of monotonicity, the characteristics of the (ICC) of the test, a graph showing the Test Boundary Response Function (TBRF) has been created after the raw scores of test items have been dismissed.

The primary 10th-grade students' results on the (CST) were verified to match the (IRT) assumptions with the (PCM) by using the "win steps" program, where 69 students 32 and 37 students were found to be out of conformity according to information weighted fit statistic (INFIT), and outlier–sensitive fit statistic (OUTFIT), that is measured according to the mean square residuals which observed frequency of expected frequencies, that are distributed according to the approximate distribution of $\chi 2$, the values of which must range from 0.8 to 1.2 and according to which they are converted to (standard mean –square residuals) with mean =0 and standard deviation=1, its values should range from -1.99 to +1.99.

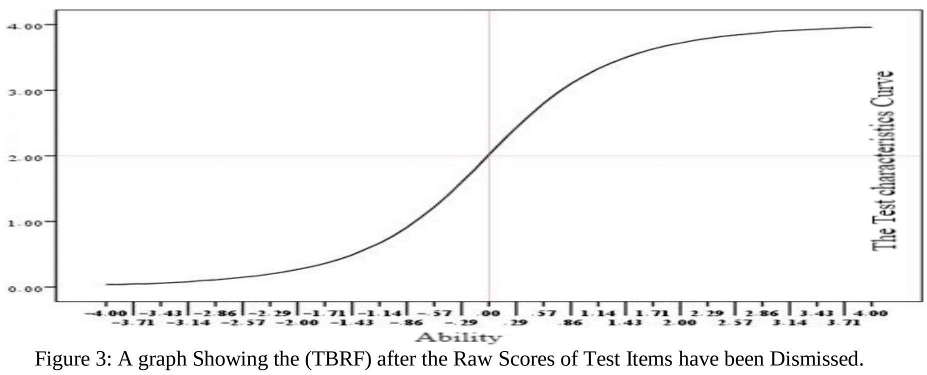

Figure 3: A graph Showing the (TBRF) after the Raw Scores of Test Items have been Dismissed.

2. Second Research Question: To what degree does the (CST) data conformity for the 10th -grade students with the (PCM)?

TABLE 3: THE DESCRIPTIVE STATISTICS OF NON-MATCHING INDIVIDUALS INDICATORS BASED ON (INFIT) AND (OUTFIT).

Point biserial Correlation Coefficient	*Standardized Residual ZSTD*		*Mean Square MSQ*		*Standard Error*	*Ability*	*Grade*	*Statistical*
	OUTFIT	*INFIT*	*OUTFIT*	*INFIT*				
0.53	**-0.23**	**-0.48**	**1.08**	**1.00**	**0.23**	**0.04**	**51.12**	**Arithmetic Mean**
0.23	**2.15**	**2.24**	**0.70**	**0.65**	**0.02**	**1.05**	**20.96**	**Standard Deviation**
-0.21	**-2.75**	**-3.28**	**0.36**	**0.32**	**0.21**	**-1.71**	**17.00**	**Minimum Value**
0.89	**3.90**	**3.19**	**2.54**	**2.25**	**0.29**	**1.91**	**85.00**	**Maximum Value**

Table 3 indicates that the raw score for non-match students ranged from 17 to 85 out of 100 scores, that their abilities ranged from -1.71 to 1.91, and that the standard errors SE of their abilities ranged from 0.21 to 0.29.

TABLE 4: THE DESCRIPTIVE STATISTICS OF THE MATCHED OF INDIVIDUALS INDICATORS BASED ON THE ZSTD AND MSQ

Point biserial Correlation Coefficient	*ZSTD*		*MSQ*		*Statistical*
	OUTFIT	*INFIT*	*OUTFIT*	*INFIT*	
0.06	**-1.93**	**-1.98**	**0.47**	**0.51**	**Minimum value**
0.535	**-0.089**	**-0.087**	**0.992**	**0.994**	**Arithmetic mean**
0.152	**0.792**	**0.842**	**0.254**	**0.242**	**Standard deviation**
0.85	**1.92**	**1.88**	**2.04**	**1.77**	**Maximum value**
-0.515	**0.132**	**0.050**	**0.533**	**0.393**	**Skewness**
-0.202	**-0.501**	**-0.574**	**-0.041**	**-0.352**	**Kurtosis**

TABLE 5: SHOWS THE DESCRIPTIVE STATISTICS OF INDICATORS MATCHING OF ITEMS.

ZSTD		***MSQ***		
OUTFIT	***INFIT***	***OUTFIT***	***INFIT***	***Item Number***
-0.57	-0.41	0.97	0.98	**1**
-0.61	-0.46	0.97	0.98	**2**
-0.16	-0.41	0.99	0.98	**3**
-1.24	-0.62	0.90	0.97	**4**
-0.88	-0.56	0.96	0.97	**5**
-1.47	-1.13	0.93	0.95	**6**
-0.26	-0.35	0.99	0.98	**7**
0.35	0.33	1.02	1.01	**8**
-1.05	-0.53	0.95	0.97	**9**
0.38	-0.21	1.02	0.99	**10**
1.01	0.96	1.05	1.04	**11**
0.39	0.38	1.02	1.02	**12**
0.71	0.80	1.04	1.04	**13**
-0.45	-0.45	0.98	0.98	**14**
-1.51	-1.18	0.92	0.95	**15**
0.81	1.16	1.04	1.05	**16**
0.75	0.11	1.04	1.01	**17**
-0.79	-1.24	0.96	0.94	**18**
0.62	0.43	1.03	1.02	**19**
0.44	0.72	1.04	1.04	**20**
-0.10	0.14	0.99	1.01	**21**
0.14	0.06	1.01	1.00	**22**
-0.46	-0.66	0.98	0.97	**23**
0.45	0.82	1.02	1.04	**24**
0.57	0.77	1.03	1.03	**25**
-0.12	-0.06	0.99	1.00	**Arithmetic mean**
0.75	0.70	0.04	0.03	**Standard deviation**
-1.51	-1.24	0.90	0.94	**Minimum value**
1.01	1.16	1.05	1.05	**Maximum value**

Table 5 highlighted that the (INFIT), according to Mean-Square Residuals (MSR) for observed frequencies on expected frequency range from -1.24 to 1.16. Moreover, the (OUTFIT), according to Standardized Mean-Square Residuals (SMSR) for observed frequencies on expected frequencies ranges from -1.51 to -1.01.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

3. Third Research Question: What are the estimates of the values of the parameter of the items according to the Partial Credit Model (PCM)? To answer this question, the descriptive statistics conducted to each raw score and the 10th-grade students' ability who's matching the test, the (SE) for ability according to the Rasch model, and (PCM).

TABLE 6: SHOWS THE DESCRIPTIVE STATISTICS EACH RAW SCORE AND THE ABILITY OF STUDENTS AND (SE) FOR ABILITY.

The Standard error for ability	***Ability***	***Raw score***	***Statistical***
0.21	-2.01	13	**Minimum value**
0.236	0.226	54.632	**Arithmetic mean**
0.027	1.088	21.066	**Standard deviation**
0.46	3.13	95	**Maximum value**
2.046	0.025	-0.140	**Skewness**
7.639	-0.967	-1.171	**Kurtosis**

Table 6 presents raw score values for students in the 10^{th} grade of the (CST) ranged from 13 to 95 with an (AM) of 54.632, an (SD) of 21.066, a [-0.140] Skewness and a Kurtosis of -1.171. The values of the abilities of the 10th-grade students to test the rated CST range from -2.01 to 3.13 with AM of 0.226, an SD of 1,088, a Skewness of 0.025, and a Kurtosis of -0.967, it ranged from 0.21 to 0.46 with an arithmetic mean of 0.236, a standard deviation of 0.027, a Skewness of 2.046, and a Kurtosis of 7.639.

4. Fourth Research Question: What are the estimates of the values of the person's ability depending on the model used? To answer this question; the difficulty parameter values for the estimated (CST), SE, and the Point biserial Correlation Coefficient for 10th-grade students were calculated according to the (PCM).

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE7: THE VALUES OF THE DIFFICULTY PARAMETER FOR THE ESTIMATED (CST) ITEMS, STANDARD ERRORS, AND THE POINT BISERIAL CORRELATION COEFFICIENT.

Point biserial Correlation Coefficient	*The standard error of the parameter*	*Difficulty parameter*	*Score*	*Item number*
0.70	0.04	-0.55	2512	**1**
0.62	0.04	0.73	1512	**2**
0.59	0.04	1.14	1161	**3**
0.64	0.04	-1.16	2942	**4**
0.66	0.04	0.56	1574	**5**
0.74	0.04	0.02	2030	**6**
0.70	0.04	-0.41	2360	**7**
0.63	0.04	-0.03	2035	**8**
0.63	0.04	1.28	1035	**9**
0.64	0.04	0.67	1475	**10**
0.66	0.04	0.47	1635	**11**
0.60	0.04	-0.60	2496	**12**
0.68	0.04	0.77	1430	**13**
0.67	0.04	-0.82	2589	**14**
0.74	0.03	0.07	1945	**15**
0.66	0.04	0.42	1686	**16**
0.69	0.04	0.71	1510	**17**
0.65	0.04	-0.84	2685	**18**
0.70	0.04	-0.32	2322	**19**
0.58	0.04	-1.33	3054	**20**
0.68	0.04	-0.59	2531	**21**
0.67	0.04	0.24	1875	**22**
0.69	0.04	-0.15	2166	**23**
0.63	0.04	-0.28	2264	**24**
0.62	0.04	0.03	2053	**25**

TABLE 8: THE DESCRIPTIVE STATISTICS FOR RAW SCORES AND THE DIFFICULTY PARAMETER FOR THE ESTIMATED (CST) ITEMS AND (SE).

The standard error of the parameter	*Difficulty parameter*	*Score*	*Statistical*
0.03	-1.33	1035	**Minimum value**
0.0396	0.0012	2035.08	**Arithmetic mean**
0.002	0.697	542.184	**Standard deviation**
0.04	1.28	3054	**Maximum value**
-5,000	-0.049	0.024	**Skewness**
25,000	-0.714	-0.777	**Kurtosis**

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 8 shows the values of raw scores for the estimated (CST) items of the 10th-grade students ranged from 1035 to 3054 in an (AM) of 2035.08, an (SD) of 542.184, Skewness of 0.024, and Kurtosis of -0.777. The values of the difficulty parameter for the estimated CST items in the 10th-grade students ranged from -1.33 to 1.28 with calculation AM of 0.0012, a SD of 0.697, Skewness of -0.049, and kurtosis of -0.714. The (SE) values of the difficulty parameter are noted for the estimated CST items in the 10th-grade students. It ranged from 0.03 to 0.04 with an AM of 0.0396, an SD of 0.002, skewness of -5,000 and kurtosis of 25,000.

5. Fifth Research Question: What are the criteria of the performance on the test items according to the Partial Credit Model (PCM)? To answer this question, the percentiles were calculated using SPSS V23.0 for the10th-grade students' abilities estimated according to the IRT depend on PCM and their raw grades as standards for testing CST to give a clear view of the relative position of the individual in their group and to compare student performance levels on a test their CST according to the sample study and their gender (male, female).

Table 9 shows that 10th-grade students who have the ability 1.71 or the raw score 17 Percentile 3 this means they are equivalent. Moreover, they were with the ability 1.61 or the raw score 81, percentiles 87, which means that the male students were equivalent to female students with the ability 1.55 or the raw score 80. This explained that the CST had good psychometric properties.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

Total		*Gender*				*Performance*	
		Female		*Male*			
Percentile	*Frequency*	*Percentile*	*Frequency*	*Percentile*	*Frequency*	*Ability*	*Raw Score*
47	12	47	8	47	4	**0.14**	**54**
48	6	49	4	48	2	**0.18**	**55**
49	12	50	8	48	4	**0.23**	**56**
51	15	51	6	50	9	**0.28**	**57**
52	7	52	3	51	4	**0.33**	**58**
53	16	53	8	52	8	**0.38**	**59**
54	10	55	7	54	3	**0.42**	**60**
55	9	56	1	55	8	**0.47**	**61**
57	15	57	7	56	8	**0.52**	**62**
59	24	59	15	58	9	**0.57**	**63**
61	14	61	6	60	8	**0.62**	**64**
62	10	62	6	61	4	**0.68**	**65**
64	21	64	13	63	8	**0.73**	**66**
66	15	66	5	65	10	**0.78**	**67**
67	14	68	7	67	7	**0.83**	**68**
69	17	69	9	68	8	**0.89**	**69**
70	16	71	5	70	11	**0.94**	**70**
72	14	72	8	72	6	**1.00**	**71**
73	11	74	8	73	3	**1.05**	**72**
75	12	75	8	74	4	**1.11**	**73**
76	23	77	10	76	13	**1.17**	**74**
78	15	79	8	78	7	**1.23**	**75**
80	9	81	7	79	2	**1.29**	**76**
81	17	82	9	80	8	**1.35**	**77**
83	17	84	8	82	9	**1.41**	**78**
85	15	86	5	84	10	**1.48**	**79**
86	11	87	6	85	5	**1.55**	**80**
88	18	88	8	87	10	**1.61**	**81**
89	11	90	6	89	5	**1.69**	**82**
90	14	91	6	90	8	**1.76**	**83**
92	20	93	11	92	9	**1.84**	**84**
94	16	94	6	94	10	**1.92**	**85**
96	11	96	5	96	6	**2.00**	**86**
97	12	97	5	97	7	**2.09**	**87**
98	6	97	3	98	3	**2.18**	**88**

Total		*Gender*				*Performance*	
		Female		*Male*			
Percentile	*Frequency*	*Percentile*	*Frequency*	*Percentile*	*Frequency*	*Ability*	*Raw Score*
1	1	1	1			***-2.02***	***13***
1	1	1	1			***-1.93***	***14***
1	6	1	2	1	*4*	***-1.86***	***15***
2	7	2	4	2	*3*	***-1.78***	***16***
3	10	3	5	3	*5*	***-1.71***	***17***
4	9	4	3	4	*6*	***-1.65***	***18***
5	7	4	5	5	*2*	***-1.58***	***19***
5	11	5	4	6	*7*	***-1.52***	***20***
6	7	6	4	7	*3*	***-1.46***	***21***
7	11	7	5	8	*6*	***-1.40***	***22***
8	8	8	4	9	*4*	***-1.34***	***23***
9	10	9	7	10	*3*	***-1.29***	***24***
10	10	11	6	10	*4*	***-1.24***	***25***
11	7	12	5	11	*2*	***-1.18***	***26***
12	14	13	7	12	*7*	***-1.13***	***27***
14	13	15	9	13	*4*	***-1.08***	***28***
15	11	16	5	14	*6*	***-1.03***	***29***
16	14	17	5	16	*9*	***-0.98***	***30***
18	7	18	2	17	*5*	***-0.93***	***31***
19	13	19	8	18	*5*	***-0.88***	***32***
20	17	21	8	20	*9*	***-0.83***	***33***
22	14	22	9	21	*5*	***-0.79***	***34***
23	12	24	4	23	*8*	***-0.74***	***35***
25	12	25	5	24	*7*	***-0.69***	***36***
26	7	26	5	25	*2*	***-0.65***	***37***
27	17	27	8	27	*9*	***-0.60***	***38***
28	13	29	6	28	*7*	***-0.55***	***39***
30	12	30	6	30	*6*	***-0.51***	***40***
31	14	32	11	31	*3*	***-0.46***	***41***
33	16	34	8	32	*8*	***-0.42***	***42***
34	12	35	6	33	*6*	***-0.37***	***43***
35	10	36	7	34	*3*	***-0.32***	***44***
37	12	37	3	36	*9*	***-0.28***	***45***
38	14	39	7	37	*7*	***-0.23***	***46***
39	8	40	3	39	*5*	***-0.19***	***47***

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

98	4	98	3	98	1	**2.28**	**89**	40	17	41	9	40	*8*	***-0.14***	***48***
99	3	98	1	99	2	**2.39**	**90**	42	11	42	6	41	*5*	***-0.10***	***49***
99	3	99	3			**2.50**	**91**	43	14	44	8	43	*6*	***-0.05***	***50***
99	2	99	1	99	1	**2.77**	**93**	45	11	45	4	44	*7*	***0.00***	***51***
99	1			99	1	**3.13**	**95**	45	5	46	3	45	*2*	***0.04***	***52***
								46	8	46	2	46	*6*	***0.09***	***53***

TABLE 9: THE VALUES OF 10TH-GRADE STUDENTS' ABILITIES ON (CST) ITEMS.

6. Sixth Research Question: What are the values of test information functions at different levels of ability?

To answer this question, the descriptive statistics for the 10th-grade students' Performance Information function were calculated on the computer subject test items according to the (IRT) depend on (PCM) at different ability levels using SPSS V23.0.

TABLE 10: THE DESCRIPTIVE STATISTICS TO INFORMATION FUNCTION FOR10TH–GRADE STUDENTS' PERFORMANCE ON (CST) ITEMS.

Kurtosis	*Skewness*	*maximum value*	*Standard deviation*	*Total*	*Arithmetic mean*	*Minimum value*	*%*	*Number*	*Level of ability*
-0.23	-0.49	16.00	1.03	5373.10	14.85	11.89	5.59	52	-2
-0.28	-0.24	22.68	1.45		20.09	17.36	24.60	229	-1
-2.01	0.09	22.68	1.01	5167.68	21.62	20.66	25.67	239	0
-0.88	-0.11	20.66	1.67	6672.85	17.99	14.79	29.97	279	1
-0.75	-0.32	14.79	1.56		12.78	9.18	13.53	126	2
0.30	-1.01	8.16	1.38		7.06	4.73	0.64	6	3
0.22	-0.85	22.68	3.35	17213.63	18.49	4.73	100.00	931	**Total**

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

International Journal of Computer Science and Information Security (IJCSIS),
Vol. 19, No. 2, February 2021

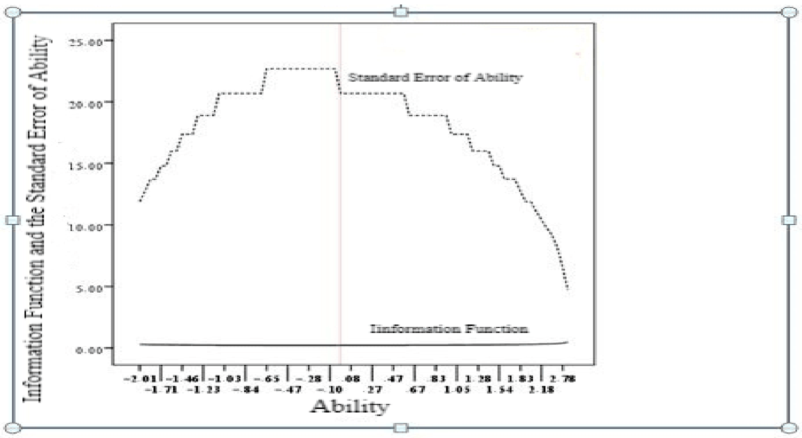

Figure 4: A scree Plot of the Student's Information Function in the Test at the Different Ability Levels

Figure 4 shows the amount of information function of students' performance on items tests at students who have high ability is considered higher than the students who have the low ability.

V. Discussions of Findings

The result of the first study question showed, achieving the results of the 10th-grade students on the CST for an un-dimensionality assumption according to the IRT depending on the PCM with three indicators, which means that the performance of the students examined on the test can be attributed to a dominant trait or only one ability, as some (latent- trait models) assume the existence of a single-trait that lies behind the interpretation of the performance of the students examined on the test. Likewise, which means that the test items were homogenous among themselves and measure the same trait and that the items, despite their different difficulties, did not differ among themselves in terms of measuring the same trait.

The results for the first study question also showed achieving the second assumption of IRT. It is a local independence LI, which means that the examined students' responses to the test's various were statistically independent. In other words, the examined students' performance doesn’t affect either negatively or positively on the items on the test on (his/her) response to any other items of the test. This means that there is reliability in assessing students’ abilities and the difficulty and reliability of the items, despite the difference in the sample of individuals used in the measurement scale as long as it is an appropriate sample. Besides, there is reliability in estimating both the individual's ability and the item difficulty and their reliability, despite the difference in the group of items used in the measurement, as long as it is an appropriate item. Moreover, the result of the first question reveals achieving the monotonicity assumption, which means that the probability of responding correctly to the terms should increase with increasing ability. Besides, for this explanation, this means that the speed factor doesn’t play a role in the response of the examined student to the test items, meaning that the reason for the examined student’s failure to answer the test items correctly is due to his limited ability, and not

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

because he was unable to reach all the test items because of the speed factor [39]. The findings of this question were consistent with the findings of [40], [41], [42], [43], [44], [45], [46], [47].

The second study question results showed matching results of the 10th-grade students in terms of ability parameter for them on the CST of the assumptions of the IRT, PCM, where only 69 students were deleted. Their answers did not match the expectations of the PCM. Where a non-matching student means that his/her observed responses deviate from the model’s expectations, such as he/she may answer about the items incorrectly despite its difficulty level below his/her ability level, or he/she is answering about the items correctly, despite its difficulty level above his /her ability. Moreover, the results of the same question also showed matching the findings of the 10th-grade students in terms of the difficulty parameter of the items on the CST, where none of the test items did match the expectations of the PCM; where is meant by the non-matching item is that the probability of answering the items is high for students with low abilities and low for students with high abilities. The findings of this question in terms of matching the items to the PCM are in agreement with the findings of a study [27], [48], [49], [50] and disagreed with the findings [51]. The third study question results showed that the values of the abilities of the 10th-grade students matched on the CST were ranged from -2.01 to 3.13 with an AM=0.226, SD=1.088, skewness=0.025, and kurtosis=-0.967. This means that the student's abilities on the test are not distributed using the normal distribution (ND) as they are supposed to. Since, hence their abilities skewness is positive; this means that their abilities are located on the left side of the normal distribution more than they are on the right side. Since their abilities kurtosis is negative, this means the spread and distribution of their abilities within a wide range under the ND and not being concentrated in a specific place. Moreover, the results showed that the value of the arithmetic mean AM of the student's abilities becomes clear that their abilities are higher than the test level, meaning that it is easy for them to test despite the positive skewness abilities. The results of this question agree with the result of the study [52] and are not in agreement with the result of the study [53]. The results of the fourth study question showed, the values of the difficulty parameter for the CST ranged from -1.33 to 1.28, within AM= 0.0012, SD= 0.697, the skewness = -0.049 and kurtosis = -0.714, which means mediating the difficulty of the CST items, as it is not extreme in its difficulty, in the sense that the distribution of its difficulties has been negatively skewness, which indicates its ease and in terms of its negative kurtosis, which means that its difficulties are not concentrated in a specific place under the ND and in terms of the narrow range of the difficulties of the items, which means that they are close within a narrow range of their values. This question's results agreed with the results of the study [54], [55]. The fifth study question results showed; the convergence of the percentile values of different raw grades and students 'abilities at the level of the study sample and the level of the two categories of the students' gender variable. This means that the CST results for tenth-grade students are stable, which leads to the possibility of generalizing the use of the test to all the community of 10th-grade students. The fifth question results regarding performance criteria are in agreement with the study [56]. The results of the sixth study question showed, the highest value of the AM of the students' performance information function on the test is at the intermediate ability level and that the lowest value of the AM of the student performance information function on the test is at the high ability level, and that the sum of what was submitted by the 10th-grade students with low abilities is less than the sum of what has submitted students with high abilities information, which means that the test is suitable for students with high ability than it is suitable for

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

students with low ability. The amount of students' performance information on the test for high-ability students is higher than for low-ability students. The result is the PCM's suitability in calculating the test information function because it corrects the SE value corresponding to each ability level.

VI. *Conclusion*

This study was considered unique in choosing the primary stage in public schools in Irbid government as well as it is unique in its approach to construct an achievement test in CS, in particular. Despite this, it is a study in common with previous studies in its general field of achievement tests and its attempt to identify the degree of effectiveness of applying the PCM in achievement tests. The findings of this study have brought several issues that concerning (CS) achievement through verifying the and reliability of the test, and it’s the accomplishment of the suppositions of the (IRT) according to the (PCM). (i) The foreign studies that dealt with the current topic are a lot and various. In other words, there is great interest in the topic by foreign research, whereas it is limited in Arab societies and Arab studies in terms of using and employing it. (ii) All the samples were university students or secondary stage students in these studies. As a result, we will apply it to the primary stage students to demonstrate its effectiveness at this stage, compared to the older age stages. (iii) The majority of the previous research focused on using statistical methods in the light of the Classical Theory to verify from psychometric characteristics, a few of them used the modern forms in measurement, so it has been applied according to Rasch Model PCM and demonstrated its effectiveness with the achievement tests.

VII. *Recommendations*

Based on the findings of this study and considering the significant place of computer in our educational system, the study made the following recommendations:

1. Teachers and other stakeholders should pay special attention to encourage and motivate students to develop a good study habit to improve their academic achievement in CS.
2. Further studies should be adopted the PCM to see the contribution of this model in measuring the achievements of students.
3. Adoption of the current CST by10th-grade teachers.
4. Teachers and other stakeholders should endeavor to encourage and motivate students to learn (CS).
5. Teachers may need to be more sensitive to the different needs of male and female students. Hence, care has to be placed when teaching both genders.
6. Curriculum developers should develop instructions that would improve students’ knowledge by laying more emphasis on the perceived difficulty areas in CS.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

VIII. References

[1] Haladyna, T. M., & Downing, S. M. (2011). Twelve steps for effective test development. In *Handbook of test development* (pp. 17-40): Routledge.
[2] Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*: Sage.
[3] De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical education, 44*(1), 109-117.
[4] Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational measurement: issues and practice, 12*(3), 38-47.
[5] Zaman, A., Kashmiri, A.-U.-R., Mubarak, M., & Ali, A. (2008). Students ranking, based on their abilities on objective type test: Comparison of CTT and IRT.
[6] Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.
[7] Rasch, G. (1977). On specific objectivity. An attempt at formalizing the request for generality and validity of scientific statements in symposium on scientific objectivity, Vedbaek, Mau 14-16, 1976. *Danish Year-Book of Philosophy Kobenhavn, 14*, 58-94.
[8] Black, J., Brodie, J., Curzon, P., Myketiak, C., McOwan, P. W., & Meagher, L. R. (2013). *Making computing interesting to school students: teachers' perspectives.* Paper presented at the Proceedings of the 18th ACM conference on Innovation and technology in computer science education.
[9] Hsu, T.-C., Chang, S.-C., & Hung, Y.-T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296-310.
[10] Steinberg, J. (2000). Frederic lord, who devised testing yardstick, dies at 87. *New York Times*.
[11] Van Horn, R. W. (1991). *Advanced Technology in Education: An Introduction to Videodiscs, Robotics, Optical Memory, Peripherals, New Software Tools, and High-Tech Staff Devel*: Brooks/Cole Publishing Company.
[12] Tuckman, B. W. (1993). The essay test: A look at the advantages and disadvantages. *Nassp Bulletin, 77*(555), 20-26.
[13] Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories.
[14] Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. *Mathematical and theoretical systems, 4*, 7-16.
[15] Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical care*, I7-I16.
[16] Nering, M. L., & Ostini, R. (2011). *Handbook of polytomous item response theory models*: Taylor & Francis.
[17] Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models*: Sage.
[18] Ferguson, L. W. (1942). The isolation and measurement of nationalism. *The Journal of Social Psychology, 16*(2), 215-228.
[19] Lawley, D. N. (1943). XXIII.—On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics, 61*(3), 273-287.
[20] Mosier, C. I. (1940). Psychophysics and mental test theory: Fundamental postulates and elementary theorems. *Psychological review, 47*(4), 355.
[21] Richardson, M. W. (1936). The relation between the difficulty and the differential validity of a test. *psychometrika, 1*(2), 33-49.
[22] Lord, F., & Novick, M. (1968). Statistical theories of mental test scores. Edited by. In: Reading, MA. Addison-Wesley Publishing Company.
[23] Tran, U. S., & Formann, A. K. (2009). Performance of parallel analysis in retrieving unidimensionality in the presence of binary data. *Educational and Psychological Measurement, 69*(1), 50-61.
[24] Zwitser, R. J., & Maris, G. (2016). Ordering individuals with sum scores: the introduction of the nonparametric Rasch model. *psychometrika, 81*(1), 39-59.

[25] Masters, G. N. (1982). A Rasch model for partial credit scoring. *psychometrika, 47*(2), 149-174.

[26] Choi, S. W., & Swartz, R. J. (2009). Comparison of CAT item selection criteria for polytomous items. *Applied Psychological Measurement, 33*(6), 419-440.
[27] Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series, 1992*(1), i-30.
[28] Chen, L.-Y. (2010). *An investigation of the optimal test design for multi-stage test using the generalized partial credit model.*
[29] Andersen, E. B. (1997). The rating scale model. In *Handbook of modern item response theory* (pp. 67-84): Springer.
[30] Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594.
[31] Dehqan, A., Yadegari, F., Asgari, A., Scherer, R. C., & Dabirmoghadam, P. (2017). Development and validation of an Iranian Voice Quality of Life Profile (IVQLP) based on a classic and Rasch Rating Scale Model (RSM). *Journal of Voice, 31*(1), 113. e119-113. e129.
[32] Gómez, L. E., Arias, B., Verdugo, M. Á., & Navas, P. (2012). Application of the Rasch Rating Scale Model to the assessment of quality of life of persons with intellectual disability. *Journal of Intellectual and Developmental Disability, 37*(2), 141-150.
[33] LaHuis, D. M., Clark, P., & O'Brien, E. (2011). An examination of item response theory item fit indices for the graded response model. *Organizational research methods, 14*(1), 10-23.
[34] Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *psychometrika, 37*(1), 29-51.
[35] Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *psychometrika, 55*(2), 195-215.
[36] Huggins-Manley, A. C., & Algina, J. (2015). The partial credit model and generalized partial credit model as constrained nominal response models, with applications in M Plus. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(2), 308-318.
[37] Zu, J., & Kyllonen, P. C. (2020). Nominal response model is useful for scoring multiple-choice situational judgment tests. *Organizational Research Methods, 23*(2), 342-366.

[38] Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and ltenls. *Applied Psychological Measurement, 9*(2), 139-164.

[39] Hambleton, R. (1985). Item response theory: principles and applications/Ronald K. Hambleton, Hariharan Swaminathan. *Evaluation in education and human services. Kluwer-Nijhoff Pub, Distributors for North America, Kluwer Boston*.
[40] Becker, D. F., & Forsyth, R. A. (1992). An empirical investigation of Thurstone and IRT methods of scaling achievement tests. *Journal of Educational Measurement, 29*(4), 341-354.
[41] Craig, S. B., & Kaiser, R. B. (2003). Applying item response theory to multisource performance ratings: What are the consequences of violating the independent observations assumption? *Organizational Research Methods, 6*(1), 44-60.
[42] Douglass, J. B. (1981). A Comparison of Item Response Theory Models for Use in a Classroom Examination System. Promising Applications of Latent Trait Models and Evidence for Their Validity.
[43] Forster, F., & Karr, C. (1980). Using the Rasch Model to Increase the Power of Item Analysis.
[44] Henning, G., Hudson, T., & Turner, J. (1985). Item response theory and the assumption of unidimensionality for language tests. *Language testing, 2*(2), 141-154.
[45] Orr, C. S. (1983). THE EFFECTS OF TEST SPEEDEDNESS AND CONTEXT ON RASCH MODEL PARAMETERS.
[46] Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *psychometrika, 52*(4), 589-617.
[47] Wilson, T. M., & MacGillivray, H. L. (2007). Counting on the basics: mathematical skills among tertiary entrants. *International Journal of mathematical education in science and technology, 38*(1), 19-41.
[48] Afrassa, T. M., & Keeves, J. P. (1999). Changes in students' mathematics achievement in Australian lower secondary schools over time. *International Education Journal, 1*(1), 1-21.
[49] Hashway, R. M. (1977). *A comparison of tests derived using Rasch and traditional psychometric paradigms.* ProQuest Information & Learning,
[50] Selfhout, M. H., Branje, S. J., Delsing, M., ter Bogt, T. F., & Meeus, W. H. (2009). Different types of Internet use, depression, and social anxiety: The role of perceived friendship quality. *Journal of adolescence, 32*(4), 819-833.
[51] Glas, C., & Verhelst, N. (1989). Extensions of the partial credit model. *psychometrika, 54*(4), 635-659.
[52] Yuan, R. (2001). Changes in students' achievement in learning the Chinese language across grades and over time.

[53] Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. *Applied Psychological Measurement, 16*(1), 1-16.

[54] Gershon, R. C. (1994). Analyzing Multiple Choice Tests with the Rasch Model: Improving Item Calibrations by Deleting Person-Item Mismatches.
[55] Kelkar, V., Wightman, L. F., & Luecht, R. M. (2000). Evaluation of the IRT Parameter Invariance Property for the MCAT.
[56] Waugh, R. F. (2002). Measuring self-reported studying and learning for university students: linking attitudes and behaviours on the same scale. *British Journal of Educational Psychology, 72*(4), 573-604.