ACTA SCIENTIAE

ISSN: 2178-7727

Developing the California test for Critical Thinking skills among university students According to Latent Traits Theory

Prof. Safa Tairq Habeeb Ph.D, Prof. Fatin Khairy Mohammed Saad Ph.D

Educational and Psychological Sciences Department, College of Education Ibn Rushd for Human Sciences, University of Baghdad, Iraq

ABSTRACT

The Present Study aims at developing The California Critical Thinking Skills Test for University Students According to the Latent Traits Theory. To achieve this aim, Scientific Steps are followed starting with the Translating the Items and its Instructions from English into Arabic, then re-translate the Arabic version into English, taking into consideration the accuracy and comprehensiveness of the whole ideas included in the test. Several modifications are required by translation experts who were considered later. Then the Arabic version of the test is exposed on an Arabic Language expert who recommended certain modification in the style of the Arabic translation, then were all taken into account.

The test has been applied on a stratified random sample of (500) male and female university students. Rasch's unit- dimension model, which is one of the Latent Traits Theory models, is applied for item analysis. To verify model's assumptions, the following steps have been: 1. uni-dimensional: In order to verify this assumption, factor analysis is done by using (the Principal components Method). Accordingly, One factor has been meaningful to the test. The same factor depended on the lower limits of Guttman who believes that a factor is considered statistically significant when the (Eigen value) equals or more than (1), and depending on a ratio of (0.30 – up) . 2. Matching the items of the model depending on the (Chi^2 value) according to the outputs of the program. This Procedure shows that four Items are statistically significant on a level of significance of (0.05). Therefore, (18) individuals of the sample of the statistical analysis were eliminated who have unmatched responses with the model's assumptions. 3. The discrimination coefficient value is found out to be near (1), where the exact value is found out to be (1. 03). 4. The independence of the scale to achieve objectivity of measurement is represented by Rasch's model, consequently, no items have been removed.5. In order to get rid of fractions and negative signals (i.e. minus), the logged unit is changed into the percentile WAT unit to find out the items difficulty and the individual's ability estimation. In the light of the statistical analysis above, suitable conclusions, recommendations, and suggestions for further studies have been put forward.

KEYWORDS: California test, critical thinking skills, latent trait theory .

INTRODUCTION

Research problem:

The problem of the current research is determined by the scarcity of tests that measure critical thinking skills among an important segment (university students) in Iraqi society, and the ones available are relatively old and based on traditional measurement theory, and because this theory has defects that are ultimately reflected in the tools derived from it,. Therefore, there is a need to build or develop an objective test to measure critical thinking skills that suits the Iraqi environment and achieves the assumptions of the Latent Trait Theory.

Therefore, the current study seeks to develop the California test for Critical Thinking skills among university students in accordance with the Latent Trait Theory. It was originally prepared according to the traditional theory, and there is no previous study, whether foreign, Arabic, or local - according to the researchers' knowledge - that addressed its development in light of modern theory.

ISSN: 2178-7727

The significance of the Research: The current research can be significant as follows:

- The research is an attempt to respond to the ongoing developments occurring in the field of psychological and educational measurement, and the need to keep pace with them.
- The research is the first study - according to the researchers' knowledge - aimed at developing the California test for Critical TRhinking skills in accordance with the Latent Traits Theory. This test was prepared in accordance with traditional theory, and is considered one of the modern and commonly used international tests in the field of psychological and educational research.
- This research derives its significance from the fact that it aims to provide a test with objective and accurate psychometric properties on an important sample of society, namely university students. This gives confidence in its future use for scientific research purposes.
- The development of the California Critical Thinking Skills Test is achieved by fulfilling the assumptions of the Latent Trait Theory by using (Rasch's model).

Aim of the Research: The current research aims to develop the California test for critical thinking skills among university students according to the Latent Trait Theory.

Limits of the Research: The current research is determined by:

- The California Test of Critical Thinking Skills (Formula - 2000), which is approved and applied until the present time.
- University students for morning studies in the city of Baghdad for the academic year (2022 - 2023).

Definition of Basic Terms:

1. Development:

• Theoretical definition of development:

"An organized and systematic process aimed at modifying and improving the negative aspects to what is better and developing and increasing the positive aspects of a specific topic."

• Procedural definition of development:

"A specific set of scientific procedures adopted by the researcher in order to make the test items appropriate to the Iraqi environment and compatible with the assumptions of the Latent Trait Theory."

2. California Test of Critical Thinking Skills:

This test was published by California Academic Press. It was built based on the concept of critical thinking agreed upon by (46) experts from various academic disciplines at the invitation of the American Philosophical Association, using the Delphi strategy over a period of two years. This test includes measuring five critical thinking skills: Analysis, Evaluation, Reasoning , Induction and Deduction (Facione, et al., 2002:3).

3. Skills:

Theoretical definition of skill:

In light of the above, the researchers define skill theoretically as "any motor or intellectual activity that an individual performs with the least possible effort and with an appropriate level of accuracy, ease, and speed."

4. Procedural definition of the skill:

"The performance performed by sample members, represented by correct responses to the test items, of the current research tool."

5. Critical Thinking:

The theoretical definition of critical thinking:

The current research adopts the definition contained in the American Philosophical Association's Delphi report, which was adopted by the California Academic Publishing Corporation in the test version of the current research instrument.

Acta Sci., 25(2), Mar./Apr. 2024
DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

Procedural definition of critical thinking:
"The total score achieved by sample members, which represents the sum of correct responses to the items of the California Test of Critical Thinking Skills."

6. The Latent Trait Theory:
The basic idea of the Latent Trait Theory, or response to the item, is determined by the existence of a relationship between the individual's latent trait (the individual's ability) and the probability of him responding correctly (Harris, 1989: 167). This theory assumes that an individual's performance on the test can be explained through a set of traits or latent ability, and the relationship between the latent traits and test items can be described by the item's characteristic function (Hambleton, 1990: 97).

This theory attempts to model the relationship between unobserved variables (the individual's ability) and the probability of a correct response to any test item. For this purpose, three logarithmic models are used (Harris, 1989: 157). These models aim to determine the relationship between the individual's performance in the test and the traits or abilities that are described through the mathematical function of each model, with the aim of estimating the individual's position in the latent trait or ability and then explaining his performance or predicting his subsequent performance (Gusstafsson, 1980:205).

Theoretical framework: It is divided into two main axes: the first axis includes psychometrics and its traditional and modern theories, with a focus on the Latent Trait Theory and the Rasch's model, and the second axis will address the critical thinking skills.

The Psychological Measurement: Measurement is the scientific method that transforms verbal descriptions into specific dimensions. It is the method that develops science and pushes it towards objectivity. Measurement means determining the degree to which a thing or person possesses one of the characteristics. According to this meaning, the individual needs measurement in all his daily behavior, and measurement is used in every case. It is possible to describe it with numbers (Khairy, 1997: 9).

The scientific study of behavior requires a quantitative determination of certain traits or characteristics based on measurement tools that provide us with numerical values or data that we analyze using statistical methods. Measurement aims to set numbers that represent the characteristics or characteristics of people, individuals, or events according to specific and clearly formulated rules (Allam, 2003: 19).

Psychological Measurement Theories : Knowledge of measurement theory is necessary for any researcher or evaluator whose experiments require developing a psychological test to measure the variable he is interested in studying. Measurement theory plays an important role in the overall processes of research methodology, in order to test the sensitivity and accuracy of the measurement methods that have been developed (Croker and Algina, 2009: 29).

Classical Test Theory: Classical measurement theory emerged at the beginning of the twentieth century. The basic components of three notable achievements of previous years matured in it: recognizing the existence of errors in measurement, recognizing those errors as random errors, and recognizing and noting the connections between variables (Traub, 1997:8).

This theory and its associated statistical models and methods related to the test item have dominated measurement methodologies throughout the past century, and Gulliksen presented in his comprehensive classic book (Test Theory) the details of this frame of reference and the related mathematical and statistical models. This theory is mainly concerned with measurement errors (Allam, 2005: 48). It is characterized by being based on weak hypotheses and being relatively easy to interpret (Wiberg, 2004: 1).

The score that an individual obtains on a particular scale is called the Obtained Score. However, this score is often tainted by measurement errors, the amount of which, if it is possible to determine them, then we obtain

ISSN: 2178-7727

the Error Score, and by subtracting the Error Score from the Obtained Score, we obtain the true score. True Score, meaning:

• Observed score = true score ± random error score
As such, the experimental (observed) score is the sum of the true score and the error score, so:

- Error score = observed score – true score
- True score = observed score - error score (Allam, 2000: 131).

Technological developments have contributed to the advancement of educational and psychological measurement and evaluation, to reconsidering classical psychometric theories and models, and to developing contemporary models and theories to overcome many of the problems related to constructing tests and measures in various aspects of human behavior, making them more flexible and appropriate to the nature of behavioral phenomena (Allam, 2000). : 678). Among these theories are:

The Latent Traits Theory (LTT) : During the past century, many people have contributed to the development of Item Response Theory, and one of these contributions is Lawley's paper D.N. Lawley in 1943, in which he stated that many of the structures of traditional theory can be expressed in terms of the characteristic curve of the item . This paper marks the beginning of the response theory of the item as a theory of measurement. Then came the work of Dr. Lord, Dr. Novick, Dr. Wright, and mathematician Dr. Rasch, as a driving force behind the development of the theory and its applications over the past years (Baker, 2001: ii). This theory was developed mainly to overcome the problems that traditional theory suffers from, and this theory is still in continuous development (Wiberg, 2004: 2). Its basic assumption is that an individual's ability is independent of the content of the test, and the relationship between the probability of correctly answering the item and the individual's ability can be shaped in different ways depending on the nature of the test (Hambleton, & Traub, 1971: 282).

Item Response Theory (IRT) is a psychometric model for constructing, estimating, and analyzing many types of tests and items. Because it is more developed, it advances several advantages over its predecessor, the traditional measurement theory (Thompson, 2009: 1). The basic concepts of Item Response Theory are clear, and its mathematical basis is more advanced. To some extent compared to traditional measurement theory, moreover, it is based on test items and not on total test scores (Baker, 2001: 1).

The title “Item Response Theory” refers to a diverse family of models designed to represent the relationship between an individual’s response and a latent trait (Fraley, et, al. 2000: 351), which includes three basic models that differ according to the number and nature of the features, which are the single-teacher Rasch model, Lord’s two-parameter model, and Birnbaum’s three-parameter model (Embretson, and Reise, 2000, 65-70).

Item response theory is based on basic assumptions mentioned by Hamilton & Swaminthan, 1991, which are:

1. Uni-dimensionality
2. Local independence.
3. Freedom from speededness.
4. Item Characteristic Curve.
5. Linearity of measurement.

(Murad and Suleiman, 2002: 425).

Item Response Theory Models: Latent Traits models are mathematical models or functions that determine the relationship between the individual in a given test item, which can be directly observed, and the traits or abilities that are assumed to underlie this performance. Therefore, these functions are considered probabilistic, not deterministic. This means that the relationship that... It is determined by the function according to probability theory (Allam, 1985: 103).

ISSN: 2178-7727

The following relationship gives the general picture of the logarithmic distribution, which indicates the probability of an individual (s) answering item (i):

$$P(X_{is} = 1IW_{is}) = \frac{EXP(Wis)}{1+EXP(Wis)}$$

Note that the symbol Wis is a conjunction of its scope, the individual's ability, and the features of the item (Al-Taqi, 2009: 14,16).

The Item Response Theory includes three basic models that differ according to the number and nature of the features:

- One-parameter logarithmic model: (Rasch Model).
- Two-parameter logarithmic model: (Lord Model).
- Three-parameter logarithmic model: (Birnbaum Model).

They are all of the dual-response type, meaning that they depend on the data of the items that are answered (zero - one), where the correct response = one, and the incorrect response = zero (Van-der Linden, and, Hambleton, 1997: 29-30).

Rasch's Model : This model is known by the name of mathematician George Rush. Rasch, who strongly opposed naming this model after him, because it is not considered one model but rather several models, called it Measurement Models (Allam, 2005: 69). Rasch proposed analyzing test data according to Probability Theory. Although he started from a very different frame of reference, the result of the item's characteristic curve model was a logarithmic model (Baker, 2001: 25).

The Rasch's model has desirable properties, such as: linear scaling and interval scaling. Therefore, Rasch's procedures are more faithful to mathematical operations, such as correlation and regression analysis, as well as change estimation. Thus, the Rasch's model fulfills the basic measurement requirements, instead of designing models (detailed tests) to fit the data, through linear interval measurement, examining data for individuals and items, and flagging the inappropriate ones (Conrad, et al, 2010: 5).

Theoretical Framework: Lord (1980) points out that Item Response Theory, in general, models the relationship between the level of an individual's latent trait measured by the test and the individual's responses to the test items. A distinction is made between an individual's performance on the test as an observed response to the test items and their latent characteristics. The probability of a correct response to the item decreases with increasing difficulty of the item, and increases with increasing ability.

The Rasch's model assumes this type of response, as follows:

1. There is a high probability that the low-performing individual will answer correctly only the easy items.
2. There is a high probability that the individual with the highest performance will correctly answer all of the easy items, as well as some of the more difficult items.
3. In the middle area, there is a possibility that the individual will answer correctly about half of the items. In this field, half of the correct answers correspond to half of the incorrect answers. (Howie, et, al., 2008: 5)

Mathematical formula of the Rasch's model: The Rasch model is the only logarithmic formula that can give an estimate (θ_v) of an individual's ability, and (β_i) of the difficulty of the item ; So that each is independent of the other. Thus, it combines linearity of scaling and objectivity of measurement. The general equation of the model is:

$$P\ (X_{vi}- X|\theta_v - \beta_i) = \frac{\exp|X(\theta v-\beta i)}{1+\exp(\theta v-\beta i)} \quad ; x=0\text{-}1$$

ISSN: 2178-7727

This general picture is the most familiar among a group of measurement models belonging to Rasch (Kazim, 1996b: 443). This formula is called the single-parameter logarithmic weighting model (1PL), because it uses the exponential function to predict probabilities, and it includes only one parameter, which is the Item Difficulty Parameter, to represent the differences between paragraphs (Allam, 2005: 70).

Critical Thinking : Critical thinking is one of the most complex forms of thinking. Due to its connection to many behaviors, such as logic and problem solving, and its close connection to abstract thinking and contemplative thinking in terms of the similarity of many characteristics, critical thinking is a form of higher-order thinking that requires the use of advanced thinking skills similar to creative thinking, and some consider critical thinking to be a form of ability. To solve problems, the critical thinker can reach effective decisions and stable knowledge through his high ability to process information and judge it logically and with high effectiveness. Others considered critical thinking a scientific approach in dealing with the various information and situations that confront the critical thinker (Al-Atoum, et al., 2009 : 71).

Critical Thinking Skills: During the researchers' review of the literature and studies that dealt with the subject of critical thinking, he found that there are many classifications of critical thinking skills according to the multiple definitions and theoretical frameworks that explain them, and among these classifications is Facione's classification (Facione, 1998-2010): He based his definition of critical thinking skills on what It was reached in consensus by a panel of Delphi experts in 1992. The experts came up with a list that includes the main critical thinking skills, and the sub-skills that fall within each main skill, as follows:

1. Interpretation skill: It includes classification skills and extracting meaning or evidence.
2. Analysis skill: It includes the skills of examining ideas, identifying arguments, and analyzing arguments.
3. Evaluation skill: It includes evaluating claims and arguments.
4. Inference skill: It includes the skills of examining evidence and guessing alternatives, and reach conclusions.
5. Explanation skill: This includes the skills of announcing the results, justifying the results, and presenting the results.
6. Self-Regulation skill: It includes the skills of self-testing and self-correction.

(Facione, 1998: 6).

Attributes of Critical Thinker : Scholars and researchers interested in the subject of critical thinking have mentioned many of the characteristics of critical thinking (Abu Jado, and Nofal, 2007: 231). Scientists and specialists participating in the Delphi research (1990) point out that although critical thinking is a cognitive method, it is impossible to understand critical thinking without appreciating a glimpse into a person's life and his logical characteristics. Therefore, the consensus provided a detailed description of the ideal characteristics of a critical thinker, as follows: "The ideal critical thinker is curious, knowledgeable, confident, flexible, open-minded, fair in assessment, honest in the face of bias or personal tendencies, thoughtful in making judgments, willing to reconsider things, clear about issues, organized on complex matters, diligent in managing relevant information, be moderate in selecting criteria, focused on investigation, and diligent in seeking accurate findings of issues as circumstances permit" (APA, 1990: 3); (Facione, 1990: 2).

Critical Thinking Measurement: Many researchers have made multiple attempts aimed at measuring the critical thinking skills and abilities of individuals (Abu Jado, and Nofal, 279). There are many tests that attempt to measure critical thinking for different age stages, many of which have been linked to theoretical frameworks or training programs designed for critical thinking (Al-Atoum, et al., 2009: 94). Among the most common of these tests are:

1. New Jersey Test of Reasoning Skills, or New Jersey Test of Reasoning Skills
2. The Watson-Glaser for Critical Thinking Appraisal (WGCTA) test prepared by Ennis, & Millman, and issued by The Critical Thinking Company in 1985.

Acta Sci., 25(2), Mar./Apr. 2024
DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

3. The Cornell Critical Thinking Test: It was prepared by Ennis and Millman, and it was issued by The Critical Thinking Company in 1985. The test has two formats and two levels: The Cornell Critical Thinking Test, level (X) targets students Secondary stage students , and the Cornell Critical Thinking Test, level (Z), targets university students and adults (Ennis, 2009:2,3).
4. The Ennis-Weir Critical Thinking Test: Essay Test. The test was prepared by Ennis, & Weir in (1985), and is issued by the Critical Thinking Foundation. It targets high school and university students.
5. Logical Reasoning test: This test was developed by Hertzka & Gulford in 1995. It consists of two parts, and targets high school and university students (Al-Atoum, et al., 2009: 94).
6. The California Critical Thinking Skills Test Family: This set of tests is issued by the California Academic Press, and is based on the Delphi Expert Consensus Report, 1990, in defining the concept of critical thinking. Studies have shown the ability of these tools to predict By thinking critically through problem situations, and its success and stability in measuring the main thinking skills, it deals with the application of one of the thinking skills for the purpose of making a reflective judgment about what is believed or what works in a context or in a problem situation (Facione et al., 2002: 2). It targets students from the sixth grade of primary school to postgraduate students, various segments of society, and various fields and professions (Insight Assessment, 2010).

Research Methodology: The current research adopted the descriptive method to achieve its goals. The descriptive method is one of the forms of organized scientific analysis and interpretation. To describe a specific phenomenon or problem and depict it quantitatively by collecting standardized data and information about the phenomenon or problem, classifying it, analyzing it, and subjecting it to careful study (Melhem, 2000: 324).

The descriptive approach is appropriate for the current study, and according to this approach, the measures taken will be presented (some of which are applied and quasi-experimental), which are integrated to achieve the aim of the current research, which is to develop the California Test for Critical Thinking skills in accordance with the Latent Traits Theory . Here is a detailed description of these procedures:

Research Population : The population of the current study consists of all students in the (morning studies) of public universities (Baghdad, Al-Mustansiriya, Technology, Al-Nahrain) for the academic year (2022-2023).
The total number of the study population was (76,806) male and female students, distributed according to the gender variable: (35,934) male students, representing (47%) of the population, and (40,872) female students, representing (53%) of the population. Distributed according to the specialization variable into (28,127) students studying in scientific specializations, representing (37%) of society, while those studying in humanity specializations are (48,679) male and female students, representing (63%) of society.

Research Samples: In the current research, two samples were selected from the study population:
• The sample for clarity of instructions and understanding of the items was (50) male and female students, who were randomly selected from the sample for statistical analysis.
• Statistical analysis sample: Its purpose is to analyze data derived from its individuals' responses to the test items. The sample size was (500) male and female students. They were selected using a multi-stage random sampling method, with (225) males and (275) females, distributed proportionally among the variables of university/college and stage of study, according to the percentages of the distribution of the study population.

Research Instrument : The researcher used the California Critical Thinking Skills Test (CCTST-2000), issued in 2000 by the California Academic Press as part of a series of tests and scales to measure critical thinking that target different age groups and classes of society. It was prepared by Peter Facione , based on the American Philosophical Association's definition of critical thinking contained in the Delphi Expert Consensus Report, 1990. The test targets adults and university students.

Acta Sci., 25(2), Mar./Apr. 2024
DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

The test consists of (34) multiple-choice items with four or five alternatives, one correct alternative. Thus, the test is scored by giving one score for the correct answer and zero for the wrong answer, meaning that the test is double-scored (1-0). Thus, the test is designed to measure a group of the following skills: (analysis, evaluation, reasoning, deduction, and induction), where the first three skills were mentioned in the Delphi report as central skills, which together constitute a complete representation of critical thinking skills, followed by the skills of deduction and induction as more traditional methods of formulating Concepts of logical thinking, according to which logical thinking is divided into deductive reasoning and inductive reasoning (Facione et al. 2002: 2- 7).

Test development procedures: The process of developing the test (the current research Instrument) requires following some procedural steps that achieve this goal. Translating tests is one of the important issues in measurement, because the validity of the scores of any translated test depends partly on the accuracy of translating and adapting the test to the new environment. (Gierl, 2000: 281) The goal of translating the test when trying to develop or codify it from one culture to another is to make Its vocabulary and items are consistent with what is common in the culture of that country (Butcher & Rouse, 1996: 4).

To achieve this, as a first step, the test items and instructions were translated from English to Arabic, then re-translated from Arabic to English, taking into account the accuracy and comprehensiveness of all ideas in the test. The amendments made to the translation were taken into account when they were presented to the translation experts. Then the test instructions and items were presented to an expert specializing in the Arabic language, to ensure the integrity of the Arabic language.

Logical analysis of the test items : To verify the clarity of the instructions and the validity of the items of the California Test of Critical Thinking Skills to measure what was designed to be measured in their apparent form and their suitability to the Iraqi environment, they were presented to arbitrators in the educational and psychological sciences, to ensure the availability of appropriate characteristics for it and its alternatives in terms of form and content, and an agreement rate of (95%) was adopted. to demonstrate the extent to which the arbitrators agreed on the validity of the items and their suitability to the Iraqi environment, and it became clear that all the items received the approval of the experts in accordance with the approved standard, except for some amendments, including unifying the number of alternatives to become (four) alternatives.

Pilot Study: Clarity of Items and Instructions: The test was applied to a sample of item clarity, and the researchers recorded all the notes received during the experiment, in addition to determining the time taken to answer all the items for each subject. It was found that the instructions and most of the test items were clear and appropriate. The range of time taken is from (15) minutes to (85) minutes, with a mean of (50) minutes, and a standard deviation of (19.42) minutes.

Administering the Test : The researchers prepared a timetable for applying the test in the colleges where the test will be applied, distributed over a period of time extending from 13. Febuary.2022 to 3.August.2022. Within this timetable, the test was applied by the researchers on the statistical analysis sample of (500) students. The application was completed in one session, within the allotted time (50) minutes.

Analyzing items statistically according to the Rasch model:

1. Investigating of Unidimensionality : In order for the test to achieve objectivity of measurement according to the Rasch's model, some important steps must be verified, including achieving the assumption of unidimensionality (Hulin, et al., 1983: 79). This assumption was verified by factor analysis using the Principle Component method, and one factor was obtained with a latent root with a value of (30,442), and an explained meaning within (89,534%) of the total variance. The results were identical to the results of the direct analysis before rotation, where the interpretation of the factor was based on Guttman's lower bound, which is considered a statistically significant factor when the latent root (Eigensvalue) that can be interpreted is equal to or greater than (one) integer (Abdul Khaleq , 1983: 148).

ISSN: 2178-7727

After obtaining the correlation coefficients between the items using factor analysis, the saturation coefficients of the items are examined using the general factor (Abu El-Nil, 1987: 372). It turned out that all test items had a saturation of the general factor higher than (0.30) or higher according to the Guilford test, and therefore no item was excluded.

There is another indication that the items measure one trait, as the Point-Biserial Correlations coefficient is used. In this regard, Anastasi (1988) indicates that the most important thing in the process of constructing the test is the correlation coefficient of the item score with the total score.

Thus, the Items measure a single trait (Anastasi, 1988: 209).
The correlation coefficient of the score of each item with the total score of the test was calculated using the statistical analysis sample, and the values of the correlation coefficient ranged from (0.261-(0.533), and by comparing the values of the calculated correlation coefficients with the critical value of the correlation coefficients of (0.128) at a significance level of (0.01), and a degree of freedom (498). It was found that the relationship of all items to the total score of the test is statistically significant, which is an indication that the test items are consistent with each other in measuring one trait. Thus, the assumption of (unidimensionality) has been fulfilled.

2. Estimation of Test Parameters : The computerized program of Rasch's Model Item Calibration Program- Rascal was used to estimate the parameters of the model: the difficulty of the items and the abilities of the subjects, and to calibrate them, that is, grading the test items according to the responses of the sample members for statistical analysis. The program displays the output after analyzing the data in the form of (7) pages. Including the fifth page (variable map for items and individuals), as in Figure (1).

RASCAL for Windows (tm) Version 3.50 Page 5
Copyright (c) 1982 - 1995 by Assessment Systems Corporation
Rasch Model Item Calibration Program
Date: 02 Jun 2011 Time: 21,03

ITEM BY PERSON DISTRIBUTION MAP

ITEMS	Numbers of PERSONS	Items / People
+ -4.0 +	0 / 0	
| -3.8 |	0 / 0	
| -3.6 |	0 / 0	
| -3.4 |	0 / 0	
| -3.2 |	0 / 0	
+ -3.0 +	0 / 0	
| -2.8 |	0 / 1	
| -2.6 |##	0 / 11	
| -2.4 |	0 / 0	
###| -2.2 |###	1 / 15	
###+ -2.0 +###	1 / 15	
###| -1.8 |#####	1 / 23	
#######| -1.6 |#####	2 / 25	
| -1.4 |########	0 / 34	
###| -1.2 |########	1 / 36	
##########| -1.0 |########	3 / 35	
##########| -0.8 |#######	3 / 31	
#######| -0.6 |########	2 / 36	
#######| -0.4 >|##########	2 / 45	
###| -0.2 |####	1 / 21	
###+< 0.0 +##	1 / 11	

Acta Sci., 25(2), Mar./Apr. 2024
DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

```
       ######|  0.2 |###              2 /  16
       ######|  0.4 |######           2 /  31
        ###|  0.6 |#######           1 /  35
        ###|  0.8 |#####             1 /  24
      #########+  1.0 +##              3 /  11
        ###|  1.2 |###               1 /  17
       ######|  1.4 |##               2 /  12
         |  1.6 |##                 0 /  9
        ###|  1.8 |#                 1 /  4
        ###+  2.0 +                  1 /  2
       ######|  2.2 |                2 /  0
         |  2.4 |                   0 /  0
         |  2.6 |                   0 /  0
         |  2.8 |                   0 /  0
         +  3.0 +                   0 /  0
         |  3.2 |                   0 /  0
         |  3.4 |                   0 /  0
         |  3.6 |                   0 /  0
         |  3.8 |                   0 /  0
         +  4.0 +                   0 /  0
+---+---+---+---+-----+---+---+---+---+
20 15  10  5         5  10  15  20
  Percent of                Percent of
  Items                     Examinees
```

Summary Information:

	Average Difficulty	S.D. Difficulty	Average ability	S.D. ability
(Theta Metric)	-0.00	1.24	-0.48	1.04
(Scaled Score Metric)	100.	11.3	95.7	9.5

Figure (1), page five, variable map for items and individuals

Figure (1) shows the variable map for items and individuals, and at the bottom of the page the average difficulty of the test appears (-0.00) in logit units and (100.0) for weighted scores, with a standard deviation of (1.24) in logit units and (11.3) for weighted scores. As for the average ability of the sample members, it was (-0.48) in logit units, (95.7) for weighted scores, and a standard deviation of (1.04) in logit units (9.5) for weighted scores.

On the seventh page, the program presents the test information curve. As in Figure (2).

```
          RASCAL for Windows (tm) Version 3.50          Page  7
      Copyright (c) 1982 - 1995 by Assessment Systems Corporation
              Rasch Model Item Calibration Program
Date: 02 Jun 2011                          Time: 20,52
   Test characteristics:     Estimated      Expected      Average
                          reliability    Information   Information
                            0.850          5.785         4.715


                    Test Information Curve


  8.0  I
```

ISSN: 2178-7727

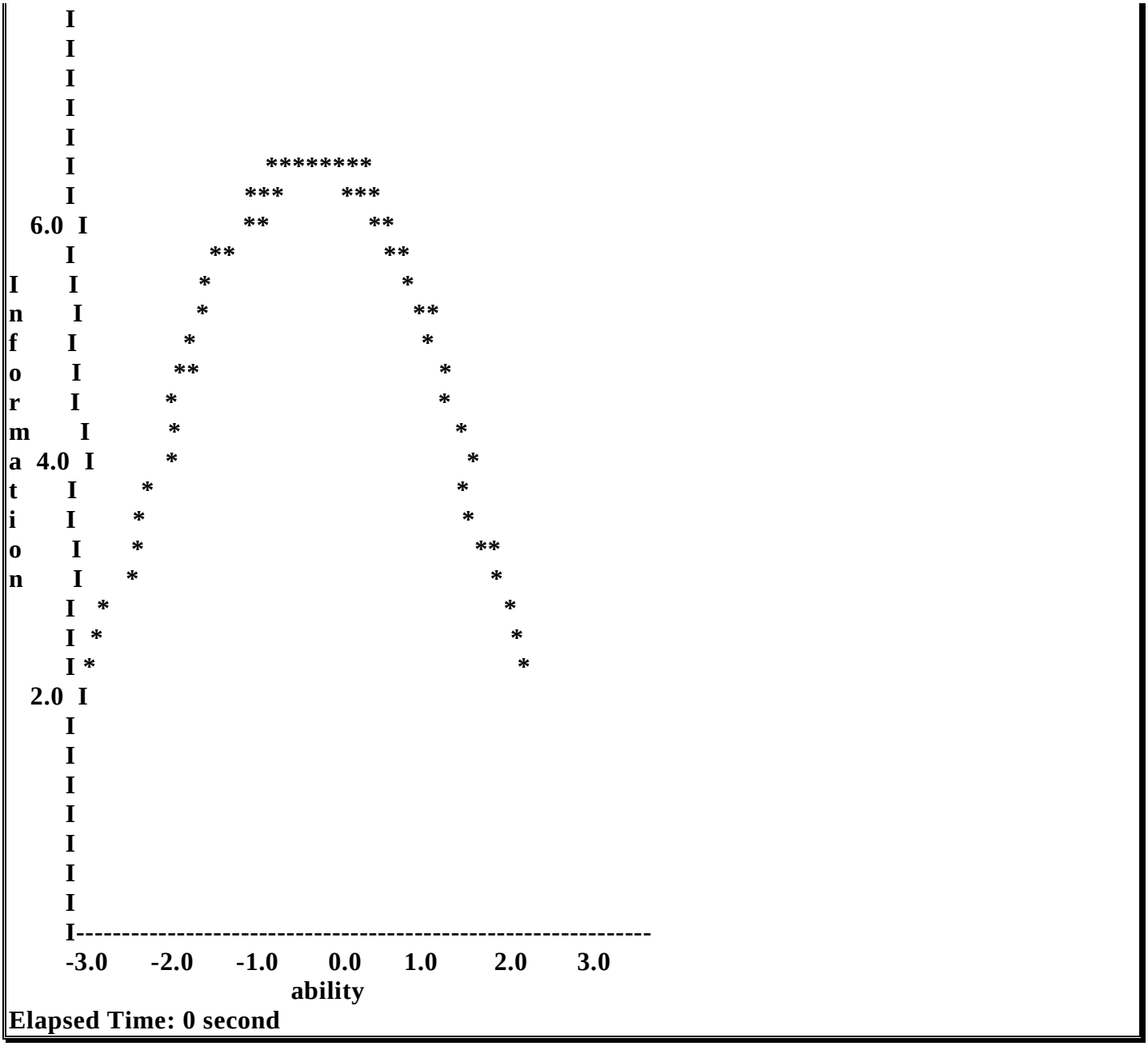

Figure (2) The test information function as displayed by the program on page seven.

It is clear from Figure (2) that the value of the reliability coefficient reached (0.850), and this indicates that the capabilities of individuals follow a normal distribution. As for expected information, which is the average amount of information provided by the test about a sample of subjects drawn from a population in which ability is normally distributed, it reached (5.785).

Average Information Provided: Average Information, which is the average of the highest height of the test information curve, and represents the average amount of information provided by the test regarding the abilities of a sample of examinees. These abilities are distributed orthogonally, and it reached (4.715).

3. Goodness of fit test : It means matching the data of the item with the expectations of the model. In fact, estimating the parameters of the difficulty of the item and the capabilities of individuals are only the first steps. It may appear upon practical application, and unexpectedly, that the model assumptions are not met in some way. This appears in the observed results not matching the model's expectations. It may be due to two sources: poor fit of items, poor fit of individuals, or both (Kazim, 1996a: 336).). Kazem points out that there

DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

are basic criteria upon which the testing of items that conform to the assumptions of the Rasch's model and the exclusion of non-conforming items is based:

The first test: That the item agree in its definition of the variable with the rest of the test items: After conducting the preliminary analysis, it was found that there were four items that were statistically significant, meaning that they did not conform to the model assumptions. The item should be deleted, because it does not express the characteristic expressed by the rest of the items. These statistics are also used to exclude subjects with responses that do not conform to the model, as the relative difficulty of the items for these subjects differs in the results of conformity to the model before deleting the items, then re-analyzing and then deleting the items that do not conform to the model, as they may be responsible for some of the items not conforming to the model. (Wright, & Stone, 1979: 82).

Because the RASCAL program does not calculate the values of statistics for individuals not matching the model , Therefore, the researcher manually deleted the responses of the subjects who did not conform to the model's assumptions by dividing the test items into (easy and difficult) tests according to the Rascal program's grading. After deleting the responses of these (18) individuals, the test data was re-analyzed, and it appeared that all test items conformed to the assumptions. The model, and this indicates that these responses are responsible for the mismatch of the items. Thus, the sample size became (482) individuals.

The second test: That the item be independent of the sample: This is achieved when the test items fit the model, as there is a general shape or curvature to the distinctive curves of the items, meaning that these curves are parallel and then have the same power to distinguish between individuals on the trait continuum (Kazim, 1996b: 453). Through the outputs of the statistical analysis of the RASCAL program, it was found that the test items had one discriminatory power, reaching (1.040), and thus this test was verified.

The third test: that the items have appropriate discriminatory power: Items with medium discriminatory power are considered the best items in terms of the power of discrimination and its effectiveness, and when its angle of inclination is (45o) on the power axis, then the slope of this possible curve of the model around the value one (Kazim, 1996a: 354). The discriminatory power reached (1.040), which indicates that the discrimination coefficient is within the acceptable range, around the value (1).

Verifying assumptions of objectivity of measurement: Verifying the extent to which the requirements of objectivity are met in the measurement tool that was divided (developed) according to the Rasch model is, in essence, a test of the validity of this model in achieving objectivity of measurement, and this verification is carried out in the following aspects:

First: Uni-measurement (uni-dimensional): that is, the test items define one variable, and this requires verifying whether the test items are graded in a way that shows a meaningful interconnected trend, and this was verified as mentioned previously.

Second: Independence of measurement: The independence of measurement is verified through the independence of the measurement from the ability of the sample that performs the test (Sample Free), and the independence of the measurement from the items that individuals answer (Items Free). All test items were equivalent to their corresponding statistical estimates, and all differences were less than the sum of the standard error of the two estimates. This means the equality of the corresponding estimates in the analysis of the total sample as (reference estimates) and those derived from the performance of the high and low-level samples, and this indicates (liberal The difficulty of the items depends on the ability of the sample to which the test was applied).

Test Validity : The use of the Rasch's model in developing the California Test for Critical Thinking Skills implies the availability of validity and reliability conditions for estimates of both the difficulties of the test items and the individuals' abilities. The validity of the test in the Rasch's model appears when the first

Acta Sci., 25(2), Mar./Apr. 2024
DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

requirement of objectivity is met in the measurement tool prepared using this model, which is that the items know one variable among themselves. (Kazim, 1996a: 366).

Likewise, deleting the screened individuals whose responses do not match the model's assumptions and retaining the screened individuals whose responses match the model's assumptions, that is, those whose observed responses to the item and their probability of success in it are consistent. This means that there is consistency between their responses to the item and their overall grades on the test, which is An indicator of the validity of the scale (Awadallah, 2000: 159-160).

By excluding non-conforming individuals, all items became consistent with the model's assumptions. When the test items do not match the model's assumptions and expectations as a result of the non-conforming individuals' responses, this is an indication of the test items' sensitivity and validity in measuring the target trait.

Test Reliability : Contemporary measurement theory asserts that quizzes that include good quality items can be more stable if we assume that the discrimination of their items is consistent (Allam, 2005: 56). This applies to the current research tool, as it consists of (34) items, which have equal discriminatory power, reaching (1.040) in the initial analysis, and reaching (1.030) in the second analysis.

The reliability of the measurement is achieved when the remaining requirements of objectivity are met, that is, when the independence of the measurement is achieved from the items of the test applied to the sample, and its independence from the sample to which the test is applied (Kazim, 1996b: 367). The independence and freedom of measurement provided by the Rasch's model provides the opportunity for measurement stability, as the measurement (whether it is the ability of the individual or the difficulty of the item) does not vary depending on the rating sample or the test used to measure ability. The reliability coefficient of the test, as calculated by the Rascal computer program, was (0.850).

There are other indicators that the researchers adopted to indicate the reliability of the test, as follows:

1. The value of the Item Separation Index in the current test was (10.957), and since this value exceeds (2), the test items are considered sufficient to define the connection of the measured trait (critical thinking skills), and the coefficient is considered an indicator of the validity of the items in Measure this attribute.
2. The value of the reliability coefficient related to the sample was (0.99), and indicates the adequacy of the sample of individuals in separating the items.
3. The value of the reliability coefficient related to the test items was (0.802), and indicates the adequacy of the sample of items in separating between the sample members and thus in distinguishing between the level of (critical thinking skills) of these individuals.

Statistical indicators of the test in its final form: After completing the test analysis procedures and grading it according to the Rasch's model, all (34) test items appeared to be consistent with the model's assumptions, and were graded according to the values of the difficulty parameters of the items, and the statistical indicators were as follows:

1. The average test difficulty is (-0.00) in logit units and (100.0) in weighted scores, with a standard deviation of (1.26) in logit units and (11.5) in weighted scores. The average ability of the sample of examinees was (-0.44) in the logit unit and (96.0) in the weighted scores, with a standard deviation of (1.03) in the logit unit and (9.4) in the weighted scores.
2. The difficulty of the items ranged from (-2.320) to (2.839) Logit.
3. The ability of the sample members ranged from (-2.53) to (1.99) Logit.
4. The angle of inclination of the characteristic curve of the test on the power axis fluctuated between (23-27) to (57-51).
5. The item discrimination factor was (1.03), the average information was (4.701), and the expected information was (5.754).

Acta Sci., 25(2), Mar./Apr. 2024
DOI: 10.2563/acta.sci.2024.2.11

ISSN: 2178-7727

CONCLUSIONS : In light of the procedures undertaken by the researchers to develop the California Critical Thinking Skills Test (CCTST-2000) in accordance with the Latent Trait Theory, and after verifying the assumptions of the Rasch's model. The researcher reached some conclusions as follows:

1. The effectiveness of the Rasch's model in developing the test, the current research tool, by matching the test data to the model's assumptions.
2. The test is suitable for the level of the sample members (university students), due to the consistency of the sample members' abilities and the difficulty of the items in achieving the test aim, and through the convergence of the average ability of the individuals and the average difficulty of the items. Thus, applying the test does not cause boredom or guessing about the response.
3. Suitability of test items to the Iraqi environment. was achieved through one of the model's assumptions of independence of measurement, which is that the difficulty of the items is free from the ability of individuals. The difficulty of the test items does not differ depending on the individuals examined, and this leads to the use of the test according to the Rasch's model on different individuals.
4. The presence of inappropriate individuals in the sample leads to some test items not conforming to the program's expectations.
5. The availability of several assumptions related to the adopted model or the requirements of objectivity in the measurement tool is, in essence, a test of the tool's validity and reliability.

RECOMMENDATIONS: In light of the results reached by the researcher, he recommend the following:

1. Using Latent Traits Theory models, including the Rasch's model, in constructing and developing tests for different types of thinking, because these models achieve the requirements of objectivity in measurement, and to overcome the criticisms directed at traditional measurement methods and the problems they suffer from.
2. It is necessary for Iraqi universities and research centers to provide some international tests and standards in the field of psychological and educational measurement so that researchers can develop and adapt them to the Iraqi environment.
3. Providing computerized statistical programs to facilitate the process of analyzing test data according to the Latent Trait Theory.
4. Training researchers and graduate students on the use of available computerized statistical programs.
5. Adopting the current test in various educational and psychological fields.

SUGGESTIONS: In light of the results reached by the researchers , some suggestion have been put for further studies :

1. Conducting a study similar to the current study aimed at building or developing a test for critical thinking according to the theory of the Latent Traits for other academic stages.
2. Conducting a study similar to the current study aimed at constructing or developing a test for critical thinking using teacher response models for the item (two- and three-part).
3. Conducting other studies aimed at developing tests designed for other environments and other types of thinking and adapting them to the Iraqi environment.
4. Conducting a study to build a thinking battery that includes different thinking skills, and includes the current test items within this battery.

REFERENCES

1. Abu Jado, Saleh Muhammad Ali, and Nawfal, Muhammad Bakr (2007): **Teaching Thinking, Theory and Application**, 1st edition, Amman, Dar Al-Masirah for Publishing and Distribution.
2. Abu Saada, Faisal Hussein Muhammad (2008): **The effectiveness of a proposed discussion-style program for developing some story writing skills among ninth graders**, Gaza, Islamic University, College of Education (unpublished master's thesis).
3. Abu Al-Azm, Abdel-Ghani (D.D.): **Al-Ghani's Dictionary, Sakhr website - Arabic Dictionaries**, the International World Wide Web www.Lexicons.ageeb.com.

4. Abu Al-Nil, Mahmoud Al-Sayed (1987): **Psychological, Social and Educational Statistics**, Beirut, Arab Renaissance House for Printing and Publishing.
5. Al-Bustani, Fouad Ephrem (1956): **Munjid Al-Talib**, 23rd edition, Beirut, Dar Al-Mashreq.
6. Al-Taqi, Ahmed Muhammad (2009): **The Modern Theory of Measurement**, 1st edition, Amman, Dar Al-Masirah for Publishing and Distribution.
7. Khairy, Al-Sayyid Muhammad (1997): **Psychological Statistics**, Cairo, Dar Al-Fikr Al-Arabi.
8. Abdel-Khaleq, Ahmed Mohamed (1983): **Basic Dimensions of Personality, Alexandria**, University Knowledge House.
9. Abdul Majeed, Muhammad Muhyiddin, and Al-Sabki, Muhammad Abdul Latif (d.): **Al-Mukhtar from Sihah Al-Lughah**, 4th edition, Cairo, Al-Istiqama Press.
10. Abdul Hadi, Nabil, Abu Hashish, Abdul Aziz, and Basandi, Khaled Abdul Karim (2003): **Skills in Language and Thinking**, 2nd edition, Amman, Dar Al Masirah for Publishing and Distribution.
11. Al-Atoum, Adnan Yousef, Al-Jarrah, Abdel Nasser Dhiyab, and Bishara, Muwafaq (2009): **Developing thinking skills, theoretical models and practical applications**, 2nd edition, Amman, Dar Al-Masirah for Publishing and Distribution.
12. Allam, Salah al-Din Mahmoud (1985): **Analysis of mental test data using the Rasch log-probit model - an experimental study**, Kuwait, Arab Journal for the Human Sciences, Volume 17, Issue 5.
13. **Contemporary Developments in Psychological Measurement**, Kuwait University.
14. **Educational and psychological measurement and evaluation, its basics, applications and contemporary trends**, 1st edition, Cairo: Dar Al-Fikr Al-Arabi.
15. **Analysis of psychological, educational and social research data**, 3rd edition, Cairo, Dar Al-Fikr Al-Arabi.
16. **Single-dimensional and multi-dimensional optional item response models and their applications in psychological and educational measurement**, Cairo: Dar Al-Fikr Al-Arabi.
17. American Philosophical Association [APA] (1990). **Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction**, Recommendations Prepared for the Committee on Pre-College Philosophy. ERIC Doc. No. ED 315-423.
18. Anastasi, A. (1988) : **Psychological Testing** , 6th ed. New York, Macmillan Puplishing.
19. Arabic Language Academy (D.T.): **Al-Wajeez Dictionary**, Beirut, Arab Center for Culture and Arts.
20. Awad, Abbas Mahmoud (1998): **Psychological measurement between theory and practice**, Alexandria, University Knowledge House.
21. Awadallah, Muhammad Abdul Rahim (2000): **A comparison between the two methods of the Rasch model and the traditional method of constructing intelligence tests using a criterion for predicting academic achievement**, unpublished doctoral thesis, University of Baghdad, College of Education, Ibn Rushd.
22. Baker, Frank, B. (2001): **The Basics of item responses Theory**, 2ed, ERIC Clearinghouse on Assessment and Evaluation.
23. Brennan, L.R. (1997): **A perspective on the History of Generalizability Theory**, Educational Measurement Issues and practice ,16, 4, 14-20.
24. Butcher , J.N. & Rouse, S.V. (1996): **Personality: Individual Differences and Clinical Assessment**, Annual Review of Psychology, Vol. 47.
25. Conrad, K.J., Conrad, K.M., Riley, B.B., Funk, R., Dennis, M.L. (2010): **Validation of the Current Withdrawal Scale (CWS) to the Rasch Measurement Model**, GAIN Methods Report 1.0. Chicago, IL: Chestnut Health Systems. Retrieved from TBA.
26. Crocker, Linda, and Al-Jina, James (2009): **An Introduction to Traditional and Contemporary Measurement Theory**, translated by Zeenat Yousef Da'na, 1st edition, Dar Al-Fikr for Publishing and Distribution.
27. Embretson .S.E. & Reise ,S.P.(2000): **Item response Theory for psychologists.** New jersey: Lawrence Eribaum Associates, Inc, publishers.
28. Ennis, R.H. (2009): **An Annotated List of Critical Thinking Tests**, University of Illinois UC. www.criticalthinking. Net/critthink.gif.

29. Facione, P.A. (1990): **Critical Thinking: A statement of expert consensus for purposes of educational assessment and instruction**. ("The Delphi Report"). published by the California Academic Press: Milbrae, CA as ERIC Document Reproduction Service
30. Facione, P.A, (2010): **Critical Thinking: What It Is and Why It Counts**. Millbrae, CA: California Academic Press.
31. Facione, N, Blohm, S. & Giancarlo, C. (2002): **The California critical thinking skills test**: Test manual, 2002 updated edition. Millbrae, CA: California Academic Press.
32. Fisher K. (2001)Curioser and curioser :**the virtue of wonder jornal of Education** ,182(2).
33. Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000): An Item Response Theory Analysis of Self-Report Measuresof Adult Attachment, **Journal of Personality and Social Psychology**, Vol. 78, No. 2, 350-365.
34. Gustafsson ,J. (1980) : Testing and Obtaining fit of Data to the Rasch model , British , **Journal of mathematical and Statistical psychology** ,33 .
35. Gierl, M. J. (2000). Construct equivalence of translated achievement tests. **Canadian Journal of Education**, 25, 280- 296.
36. Habeeb , Safa' Tariq , Fatin Khairy M.S. , and Balqees Hmood Kadhum (2021) : **Evaluation and Total Quality of the Higher Education Institutions** , Ist edition , Dar AL-Sadiq Cultural Institution , Iraq , Babylon.
37. Zahraa R. Abdul Rasul, Using Dual Criterion Model in the Construction of Teaching Thinking Test among Students of the Departments of Educational and Psychological Sciences . **Journal for ReAttach Therapy and Developmental Diversities** , Vol : 6 , No. 9s(2) (2023).
38. Hambleton ,R.K. (1990): Item Response theory: Introduction And Bibliography, **Psicothema**, vol,2,No1, pp. 97-107.
39. Traub, R. E. (1971): Information curves and efficiency of the three parameter logistic test models, **British Journal of Mathematical and Statistical Psychology**, 24, 273- 281.
40. Harris , D. (1989): Comparison of 1-2-3 parameter I.R.T. models Education measurement : **Issues and practice** ,8 .
41. Howie, S.J., Long, c, Sherman, V., & Venter, E. (2008): **The Role of IRT in Selected Examination Systems**, An Umalusi Research Report. Published by UMALUSI.
42. Hullin, C.L. & Drasgow, F. & Parson, C.k. (1983): **Item Response Theory : Application to Psychological Measurement**, Ilinios, Dow Jones, Irwin, Homewood, USA.
43. Kazem, Amina Muhammad (1996a): **A critical theoretical study on the psychological measurement of behavior according to the Rasch model**, in Anwar Al-Sharqawi et al., Contemporary Trends in Psychological and Educational Measurement and Evaluation, Cairo, Anglo-Egyptian Library.
44. Kazem, Amina Muhammad (1996b): **Using the Rasch model in constructing an achievement test in psychology and achieving objective interpretation of the results**, in Anwar Al-Sharqawi et al., Contemporary Trends in Psychological and Educational Measurement and Evaluation, Cairo, Anglo-Egyptian Library.
45. Melhem, Sami Muhammad (2000): **Research Methods in Education and Psychology**, Amman, Dar Al-Masirah for Publishing and Distribution.
46. Murad, Salah Ahmed, and Suleiman, Amin Ali (2002): **Tests and standards in the psychological and educational sciences**, steps for their preparation and characteristics, Cairo, Dar Al-Kitab Al-Hadith.
47. Insight Assessment (2006): **CCTST 2000 Interpretation Document**, a division of California Academic Press, 217 La Cruz Ave. Millbrae, CA 94030-2106 USA.
48. Odeh, Ahmed, and Malkawi, Fathi (1987): **Fundamentals of Scientific Research in Education and Human Sciences**, Zarqa, Al-Manar Library for Publishing and Distribution.
49. Paul, R.W., & Elder (2006): **Critical Thinking, Concepts and Tools**, 4th ,The Foundation for Critical Thinking.
50. Thompson, N.A. (2009): **Ability Estimation with Item Response Theory**, Assessment Systems Corporation.

ISSN: 2178-7727

51. Traub, E, R (1997): **Classical Test Theory in Historical Perspective**, Educational Measurement Issues and Practice, 16, 4, 8-14.
52. Van der Linden, W.J, & Hambleton, R. K. (1997): **Handbook of modern item response theory**. New York: Springer.
53. Wiberg, M. (2004): **Classical Test Theory vs. Item Response Theory An Evaluation of The Theory Test in the Swedish Driving- License Test**, Umea University, Department of Educational measurement.
54. Wright, B.D, & Stone, M. H. (1979): **Best test design: Reach measurement,** Chicago, Mesa press.