International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 14, Issue 10, 2020

Building a Standardized Achievement Test for Educational Psychology According to Classical Theory and Modern Theory

Abdullah Majeed Hamid Al-Atabi[a], Baidaa Nathir Tawffik[b], [a,b]Wasit University / College of Education for Humanities / Department of Educational and Psychological Sciences,

The aim of the present research is to build a standardized achievement test in the subject of educational psychology among the students of the second stage of the faculties of education in Iraq for the Department of Educational and Psychological Sciences according to the classic measurement theory and modern theory (Rush model).To achieve the objectives of this research, the researcher prepared test items that reached (80) items according to Bloom's cognitive levels (knowledge, understanding, application) as a tool for her research, and resorted to the method of multiple choice with four alternatives in building test paragraphs. The researcher chose to represent the sample size (50%) of the size of the adult population (1899) and thus the size of the sample (950) students distributed by university, college and sex, and all the necessary conditions have been applied with regard to the construction of the test as mentioned in the research procedures in Chapter III.

Key words: Building a Standardized Achievement, Test Educational Psychology, Classical Theory

Research Problem

University education in Iraq depends on the achievement tests to a large extent in evaluating students, and it contains many defects and errors, so it does not give the true picture of the capabilities and preparations of students to be evaluated (Al-Rubaie, 2: 2005)

Achieving objectivity using the current measurement tools prepared under the traditional measurement theory is a distant goal, as the outcome of the measurement is influenced by the

characteristics of individuals and vocabulary, and reliance on the raw degree in the estimation of the individual, and the absence of measurement units negatively affects the accuracy of the measurement.

(Kazem, 2000, 327), since traditional theory of measurement requires completely equal paradigms for a single test, and this is a very difficult requirement in practical application which reduces the value of the results based on that theory (Dowden; 2004, 34), so traditional theory assumes equal Measurement errors vary for all individuals, but we sometimes notice that some individuals have more consistent performance in testing than other individuals, and that this consistency varies with their level of ability. Therefore, it is expected that individuals with high ability will perform in parallel with test images more consistent than these individuals with medium ability. Thus, of course, the error will increase in the more difficult test applied to a group of individuals with low capabilities. (Allam: 2000, 675)

Therefore, the traditional theory of testing for years is the theoretical and practical pillar of measurement in behavioral sciences, based on the True Score Model, which cannot be measured directly, but rather depends on its calculation on the Observed Score, which contains some measure of error Measurement Error Crochar & Algeinia; 2009, 257)) As the literature of contemporary classic psychological measurement indicates that the degree that we obtain from the application of the test is called the raw degree Obtained Score, and this degree is a mixture of the true score and the error score and by subtracting the degree of error from the raw degree we get On the true degree, the Sax indicates that the measurement error or the degree of error is the result of the difference between the degree that the individual gets on the scale (the degree achieved or the degree observed) and the true degree. , (Sax, 1989; 260) Sachs agrees with Anastasy to emphasize the measurement error, as he sees that the individual's degree on the scale is the true degree of it, plus the degree of error. Where the true score means a hypothetical average of an unlimited number of applying the test when the accompanying effects are canceled when applying. Those effects are the result of accumulated experiences among those who re-apply the test to them (Anastiasi, 1976; 71) and since it was difficult to reach the real degree directly It can be reached by the degree of error. The standard deviation of the degree of error can be reached by the standard error of measurement. The true degree is indicated by the true contrast and the degree of error is the error of error (Sax, 1989; 261-262). The true degree of the individual is examined through a specific field, assuming that the degrees of error of individuals are Random and unrelated, for parallel test applications, these mean error scores are equal to zero. (Hambleton & Zeal, 1991: 69) Building tests, analyzing them, and interpreting their results according to the traditional theory suffers from some problems that reduce the accuracy and objectivity of the test, as the overall score of an individual in a test is influenced by his vocabulary: the individual’s score when he is tested with easy vocabulary is higher than Hard vocabulary, its

ability cannot be estimated with what these vocabulary measures accurately, so the result of the measurement varies according to the test used (Allam, 2000: 203)

From the foregoing, it is possible to determine the current research problem, as there is no achievement test prepared according to the classic and modern measurement theories (Rush Model) for the educational psychology course for the second stage in the Faculties of Education in Iraq, the Department of Educational and Psychological Sciences (within the limits of the researcher's knowledge) which is one of the important basic educational materials In that section.

Research Importance

Measurement science is an important aspect in the study of various phenomena, it is concerned with measuring and appreciating the phenomenon being studied. The more accurate the objective measurement, the more accurate our understanding of the phenomenon will be, here the importance of studying in the field of measurement appears, with the aim of researching and exploring the methods, means and tools that achieve these goals (Hambleton & Swaminathan; 1985, 18 and the most important thing in the psychological and educational measurement in general is determining The amount of traits underlying individuals 'performance on different tests and making use of them in interpreting the results and forecasting their behavior in similar situations, and making specific decisions about them in light of this quantitative assessment, the primary goal of both the classic theory and the theory of underlying traits is to determine the relationship between individuals' responses to a particular test and the underlying trait Beyond these responses (Popham & Sullivan, 2005: 46), this means that the contemporary trend in psychological and educational measurement aims to reach the highest level of accuracy and objectivity in measurement so that it achieves the most accurate relationship between the measuring instrument and the individual's inherent trait components (Nunnally, 1978: 315). Therefore, contemporary measurement scientists have made unremitting, aggressive efforts since the 1970s to develop a contemporary psychometric theory, called the theory of the underlying features L atent Trait Theory or Item Response Theory, and this theory and its various models have evolved considerably from that time until the present. (Allam, 2005: 3) And the role of this theory in measuring, researching and evaluating within the broader networks In the research methodology and evaluation, it is necessary to consider research in educational, psychological and social sciences as a well-formed question or test (Crocker & Algeinia; 1986,27)), since the traditional theory of measurement leads to the construction of inflexible tests, as specialists have directed In measurement their efforts to have a more objective measurement system focused on better selection of test vocabulary, and allows the addition or deletion of vocabulary to the test without being affected by the test as a whole, and efforts have led to the emergence of so-called models of underlying features (Allam, 1987: 18).

The researcher extracts from the above interest in the modern theory and compared it with the traditional theory in order to evaluate this theory and diagnose its effectiveness in the field of measurements and tests, and studies that were conducted in this field, whether those that relate to the comparison between the two theories or between them and models in the other theory, confirm this importance. The current research, then it is possible to fill a gap in the field of psychological and educational measurement, as there is no similar study (according to the researcher's knowledge) that was conducted in the Iraqi environment, and it may be a scientific addition that opens the horizon for researchers to proceed to conduct such comparisons and studies.

Research objective

The current research aims to build a standardized achievement test in educational psychology for students of the second stage of the Faculties of Education in Iraq for the Department of Educational and Psychological Sciences according to the classical and modern theory (Rush model)

Search Limits

Objective limits: Building a standardized achievement test for educational psychology according to the traditional and modern measurement theories (Fundamentals of Educational Psychology book, edited by: Dr. Touq, Mohy El-Din and Adas, Abdel Rahman: 1984)

Spatial Limits: Iraqi public universities for morning studies

Human boundaries: Students of the second stage of Iraqi universities for colleges of education, Department of Educational and Psychological Sciences, for morning study

Time Limits: The school year 2018
Defining Terms

Below is a definition of the basic terms mentioned in the current and representative research, test, achievement test, codification, educational psychology, classical theory.

First: the achievement test:
1- Arafa Al-Zahir and others, 1999: an organized method aimed at measuring the amount of information that a student memorizes and remembers in a field of knowledge (0 Al-Zahir and others, 327.1999)
2- Al-Sayed, 2000: A measure of the knowledge levels of students and other individuals in every field of study subjects, and in every field of human knowledge (Al-Sayed, 2000: 327)

Second / technicians:
1- Anastase (1959An): It is the setting of uniform conditions for applying the test to all individuals and includes a standardized method for assessing anesthetic responses (74: 1959)
2-Abel, (Abel, 1972): It is the test that has been carefully developed through specialized expertise, and it usually passes experimental and analysis podcasts, and it has clear instructions for its management (i.e. its use and application) in a standardized way and has standardized tables of standards to explain the degrees that the individual gets in Application of the test to a specific sample (Abel, 1972; 465)

Three / Classic Measurement Theory

1-: 1976 Zimmerma is the theory that relied on the philosophy of measuring differences between individuals based on a group criterion for individuals and includes a basic assumption that the degree observed on the test is the sum of the true degree and the degree of error and is expressed by the formula: X = T + E as it crosses (X) On the observed score, (T), the true score, and (E) error, the overall score for the individual on the test is compared to the relative position of the community members to which an individual belongs (Zimmerman, 1976: 89).
2- RANDALL (1998): It is a theory that focuses on the concept of true degree and degree of error, and it is assumed that if we could do the test several times on the individual with new elements and under different conditions, we get different observed degrees, average with the closest unbiased estimate of the individual's ability Or its real degree (Randall 1998: 85,)
Fourth / modern measurement theory:

It is a contemporary trend in psychological measurement that seeks to achieve objectivity and provide linear measurement, as it seeks to integrate the goals of measurement

And this theory assumes the presence and identification of one or more of the main unobserved features that determine the individual's responses to the test items (Kazem, 212: 1996).

Chapter II
Theoretical Framework

1- Classical (traditional) measurement theory:
The origins of traditional measurement theory go back to the English psychologist Spearman when he reached during the period (1904-1913) logical mathematical evidence about the degrees of scale and error to which he is exposed. Therefore, this theory is called from some the theory of true degree and degree of error (Theory of True and error score) because it explains the variance that occurs in the degree of the respondent when he answers the scale

more than once to the presence of an error created by irregular factors that interfere with the true degree. (Ghiselli et 1981: 195)

This theory must have the ability to interpret and analyze in order to be a theory suitable for use and application. The traditional theory of measurement is based on four main axioms:

A- The performance of the individual can be measured and estimated
B- The performance of an individual is a function of his characteristics
C- Characteristic, performance, and the relationship between them varies from individual to individual "individual differences"
D- The total virtual measurement consists of a real measure and another that is due to error (Saad Abdul Rahman, 1998: 75) 0

Psychometric Properties According to Classical Theory

Among these characteristics are related to the test items, including the test as a whole, and the following is a brief introduction to it:

Firstly, the psychometric properties of the paragraphs: The determination of the ability of standard reference measures to measure what is actually established for them depends on the extent of the psychometric properties that are represented by the honesty and consistency of their paragraphs (Holden et al., 1985, p. 386).
Thus, the most important psychometric properties of the reference paragraphs of standardized tests and metrics, which are based on what is assumed by the classic theory of measurement are:

A- Truthfulness of the paragraph: Classical theory considers the relation of the paragraph to an external or internal test, which is the most appropriate evidence for the validity of the paragraph, because the paragraph that is characterized by honesty must measure the characteristic or characteristic to be measured whether the respondent answers it by refusal or approval (Abdel Rahman, 1983: 414).

Given the lack of external evidence in most cases, the paragraph correlation is usually used for an internal test, which is the total score in the test. The correlation means that the paragraph measures the same concept as the total score. The paragraphs whose correlation coefficients are overall are thus statistically significant (Guilford, 1954: 417).

B - the stability of the paragraph.
Paragraph stability is the measure of the extent to which the paragraph contributes to the variance of the overall test and then its stability. This is done by examining whether or not the paragraph measures the same factor as other paragraphs. (Abdul Rahim, 2000: 48).
C- Difficulty of Paragraph:
One of the most important psychometric properties of vertebrae in standardized tests is the degree of difficulty of a single vertebra, calculated by the ratio of correct or incorrect responses to the vertebra for all individuals close to 50% to give the most accurate measure possible (Hetzel, 1997, P.1-6).

This is because both types of difficult or easy paragraphs that may not be answered correctly, or that all individuals answer, are deleted, because they do not allow us to identify individual differences. (Al-Zobaie et al., 1981: 77)
Given the difficulty of finding the average required for the difficulty factor, the difficulty of the paragraphs according to the traditional theory should extend between "30% to 70%" to give the maximum amount of test information about the differences between the examined (Allen & Yen, 1979, P: 121).

D- The discriminatory force of the clause
It is the one that measures the trait or characteristic and distinguishes between two individuals who differ in the trait or characteristic of a behavioral difference and thus is considered an indicator of the ability of the test to detect individual differences. (Ebel, 1972, P: 399). Therefore it is necessary to exclude paragraphs that do not distinguish between them, as Gisli et al. (1981) pointed out the necessity to keep the paragraphs of a discriminatory force in the final image of the scale (Ghiselle, et al, 1981: 68).

It seems that there are two methods for calculating the discriminatory strength of the standard reference test items when identifying the two groups that should be distinguished between them:

The method of the two opposing groups in the trait or characteristic to be measured, used if there are external motors.

The method of the two extremes in the total scale of the scale, and the total degree adopts an internal criterion. (Gray, 1997, P: 7)

Osthof (1976) indicated that psychologists did not agree on which of the methods is most appropriate for calculating paragraph distinction, and the results of his study demonstrated that most of these indications give similar results, so the preference of any of them should be

based on statistical relevance and the ease of calculation and interpretation. (Osterhof, 1976, P: 145-147).

E- The efficacy of wrong alternatives:
It is one of the basic procedures in analyzing paragraphs of the multiple choice type, and the aim of this procedure is to find alternatives for an effective response to the respondents, although they are wrong alternatives. For the alternative to be attractive, it must be chosen by at least 2% of the number of respondents, and it should not be replaced. (Henrysoon, 1971, P: 138-143)

Modern Measurement Theory
Among These Theories

A- generalizability theory
This theory, also called field specimen theory, emerged in the mid-twentieth century when Tryon and Cronbach created a theory that dealt with variables and psychological measurements more realistically than the classic theory of reliability of stability (Ghiselli et al, 1981.209).

In light of this theory, the validity coefficient of the test equals the square root of the coefficient of persistence, and the concept of persistence means the ratio of the overall degree variance to the observed degree variance that can be obtained by using the method of variance analysis. This ratio is referred to as the generalization potential factor. (Allam, 1986: 212-215)

B- Theory of the underlying features
Lasersfeld is the first to use the term inherent trait in this sense, and the theory may be credited with developing this theory to Rasch (1952). (Anastasi & Urbina, 1997.73).
Thus, this theory has emerged, which represents the contemporary trend in psychological and educational measurement, and is also called the response theory of the paragraph, or the distinctive curve theory of the paragraph, when applied to the ability and achievement tests, as the distinct curve of the paragraph represents the probability of a correct response to the paragraph as a function of the level of the underlying feature That appear at each level of the attribute or ability. The theory of underlying features has determined the relationship between an individual's response to test items and an unobserved inherent characteristic that underlies or controls this performance.

Therefore, latent characteristic models mostly addressed the binary corrected response pattern that gives a score of (1) for the correct response and (zero) for the wrong response (Hambelton et al., 1978.475-476).

Interpretation of Degree through the Theory of Underlying Features

In this theory, the standard on which the degree is measured is the paragraphs that are included in the exam. In this theory, the answer to the paragraphs depends on two factors: the ability to respond to the student and the parameters in which the paragraph is presented, which is the degree of its difficulty, for example in the case of adopting a single-parameter logistical model (Al-Zamili, 2012: 213).

Singular Response Theory Models

In light of the number of parameters of vocabulary, it falls under the theory of underlying features, a group of psychometric models, the most important of which are:

1. Single parameter model
It is called the 1PL Rasch Model relative to the scientist who established his rules. This model assumes that the vocabulary of the test differs only in its difficulty and its distinction. Difficulty coefficient (b)] Is there a convergence or a convergence point between the two measurement theories? In the traditional theory we were talking about the real degree, and in the modern theory we talked about the ability, and the fact that the real degree and ability symbolize the same thing but with different scales except that the difference between them is that the true degree depends on the questions that make up the test, while the measure of "ability" It is independent of the test questions, as the model assumes equal coefficients of discrimination (ai) and lack of guesswork (ci), and it estimates the difficulty of paragraphs (bi) only, and the mathematical formula for this model is as follows: Hambleton & Jones; 1993; 521))

$$pi\ (\theta) = \frac{e^{D(\theta-\beta_i)}}{1+e^{D(\theta-\beta_i)}}$$

Where:: Pi (θ) is the probability that the examiner who randomly selected the power level will answer

(θ) Correct answer to the paragraph. (i): D represents the staging factor

(Scaling Factor): bi Difficulty Instructor.: القدرة Ability Instructor
Fischer, 1973 refers to the importance of this model, as it is distinguished from other models in terms of the number of assumptions that are available in the data in order to give the model accurate estimates as well as its ease in estimating teachers of the difficulty of the paragraph and the ability of individuals statistically, and that the raw score expresses the only information required From data for estimation of magnitude (362 (Fischer, 1973:.

2. Two-parameter model: Examples include the Word model (2PL Lord Model). These models assume that the vocabulary differs in its difficulty and distinguish it between different levels of ability. Coefficient of Discrimination (and Coefficient of Difficulty) Thus the mathematical formula for this model includes difficulty and discerning teachers to represent the properties of the paragraph as in the formula: (Al-Bazun, 2014: 181)

$P(\theta) = \frac{1}{1+e}$-L $= \frac{1}{1+e}$-a(θ-β)

Where: e = 2,718, paragraph difficulty = β.

Paragraph Marker = a , a(θ-β)) logarithmic deviation,

(θ) Capacity level (Baker, 2001: 22

.*3*Three-parameter model: An example is the 3PL Barenboim Model. These models assume that the vocabulary differs in its difficulty and distinction, as well as in Para estimation meter which represents the possibility of individuals with low ability to arrive at the correct answer for the individual, especially in multiple-choice vocabulary.) It denotes the coefficient of discrimination a (§ the difficulty coefficient (b) and the coefficient of conjecture (c) known as the Three-parameter logistic model (3PL), according to the following mathematical formula: Embreston & Reise; 2000; 631))

$$P(\theta) = c_i + (1- c_i)\frac{\exp\left(a_i(\theta-\beta_I)\right)}{1+\exp\left(a_i(\theta-\beta_I)\right)}$$

previous studies

A- Arabic Studies

-*1*Bani Yassin Study (2004)

The most important goals of the research: Identify the psychometric properties of the test reference test in chemistry for students of the first secondary scientific class, estimated according to classical and modern theories.

Research community: (2462) male and female students for the first secondary school scientific level in the schools of the Jordanian Irbid region.

The research sample: It consisted of (481) male and female students.

Research tool: Building an achievement test in chemistry for the first secondary grade, consisting of (52) items of choice type from four alternatives.

The most important results

-1The psychometric properties of the test (honesty and reliability) were estimated according to the classical theory, where the coefficient of honesty in terms of criterion (0.84) and coefficient of internal consistency (0.90) with the alpha-Cronbach equation,

-2The stability factor for the test was estimated according to the modern theory using the mono parameter model, where the stability factor for the test was (0.99) and the stability factor for individuals (0.88)

-3That the relative efficiency of the test whose paragraphs are more difficult is at the highest ability level, and that the relative efficiency of the test whose paragraphs are easy is greater than at the lower power level.

B- Foreign studies

-1Cook et al (1988)

The most important objectives of the study: To compare the estimates of the parameters of the paragraph calculated by traditional theory and theory

Paragraph response and clarifying the effect of modernity on education on the stability of the results.

Research community: university students.

Research sample: second year students from the college.

Research tool: Building an achievement test in biology.

The most important results

-1The difficulty of the joint paragraphs between the two pictures, numbered 58, was calculated

-2The two theories did not give conclusive results about the difficulty and calibration of (58) paragraphs

-3Thyssen and Zimowski study (1997)

The most important objectives of the research: To compare the accuracy of individual response theory with the accuracy of classic theory in field mark estimates for multiple choice

Research Community: Graduate students at the University of Kansas in the United States of America

Research Sample: The study sample consisted of (1000) male and female students from postgraduate studies

Research Tool: Building multiple test

The most important results: The results of the study indicated the superiority of the response theory of the individual to the classic theory of accuracy in estimating the behavioral field score (content domain)

Chapter III

First: Research Methodology

The researcher followed the descriptive analytical approach, as it prepares an investigation focused on a specific phenomenon or issue as it exists in order to diagnose and reveal its aspects and determine the relationships between its elements or between it and educational or psychological phenomena (Al-Zobaie and Al-Ghannam, 1981, 51)

Second: Research Society

The current research community * is determined by students of the second stage, for the departments of educational and psychological sciences in the colleges of education for the human sciences of Iraqi government universities for the morning study, for the year (2018-2019), which was represented in seventeen public universities that include colleges of education that contain the department of educational and psychological sciences by (1899 Male and female students, distributed by (668) males and (1231) females

Third / Research Sample

The researcher resorted to choosing a random stratified sample from the research community, and through reviewing the previous studies and the budget between them, the researcher chose that the sample size represent a percentage of (50%) of the adult community size (1899) male and female students, thus the sample size reached (950) male and female students distributed according to University, college, gender, and table (1) show the sample distributions in detail.

Table 1: The research sample is distributed by university, college and gender

total	Type		collage	The University	s
	female	male			
91	58	33	Education	Kirkuk	2
100	53	47	Education	Mosul	3
100	50	50	Education	Tikrit	4
37	23	14	Education	Al-Mustansiriya	5
62	38	24	Education	Baghdad / Ibn Rushd	6
100	50	50	Education	Karbala	7

100	**50**	**50**	**Education**	**Babylon**	**8**
100	**57**	**43**	**Education**	**Wasit**	**9**
133	**71**	**62**	**Education**	**Dhi Qar**	**10**
70	**51**	**19**	**Education**	**Maysan**	**11**
57	**42**	**15**	**Education**	**Basra**	**12**
950	**543**	**407**	**total**		

Fourth / Research Tools: There are specific scientific steps for building achievement tests, and the researcher will rely on steps to build an achievement test in the current research on the steps presented in the theoretical framework as follows:

1- Determine the content to be measured
The content was determined by identifying the basic components to be achieved in teaching the subject of educational psychology for students of the second stage of the departments of educational and psychological sciences in colleges of education, and this was achieved by restricting the vocabulary of the study subject which was mentioned in the book of the sectoral body for the departments of specialization in addition to the presence of a systematic textbook By the Ministry of Higher Education and Scientific Research (Fundamentals of Educational Psychology written by Dr. Touq, Mohy El-Din and Adas, Abdel-Rahman. 1984), which the researcher found through her contact with the course teachers that he is accredited by all the colleges of education included in the research community.

Clarity of Instructions and Understanding of Paragraphs

It is necessary to verify the extent to which the target sample understands the scale instructions and the clarity of its paragraphs they have (Faraj, 1980: 159), for the purpose of ascertaining the extent of clarity and understanding of the instructions and test paragraphs, the method of response among respondents and calculating the time taken to answer the test was applied to a sample of (40) students And a student of (20) males and (20) females who were randomly chosen from the research community who are students of the second grade table (5). Discussion with the sample members after completing the completion of the answer to the test, it became clear that the instructions are clear, and the time taken to answer ranged between (30-60) minutes, with an average of (45) minutes.

Statistical Analysis of Paragraphs
The researcher resorted to statistical analysis of the test items for several justifications

1- The difference in the size and sample of the current research represented by university students from the sample of tests that the test was applied to in the studies the researcher examined.
2- The nature of the educational environment for the sample to which the test was applied in the literature and previous studies.

The Length of Time That Passed the Construction of the Test
Psychometric (Standard) Properties of the Test

Stability: The stability factor is one of the indicators of the accuracy of the test and the consistency of its paragraphs in measuring what must be measured 0 (Odeh, 1998, 340) and the constant test is the one that gives the same results after applying it twice in two different times to the same individuals (Weiten, et, al, 1999: 57 The researcher verified the stability in the following methods:

a. Halftone method
This method is concerned with calculating the internal consistency between the two parts of the test, as the test is divided into two parts in a suitable way to represent two equal images and the correlation is calculated between them. For the purpose of calculating the stability coefficient of the test in this way, the test paragraphs were divided into two parts (even - odd) and then calculating the Pearson correlation coefficient between the two sections, and for the two sections to be equivalent because the test consists of (75) items, and this is not divisible. The number of paragraphs was not equal to the other half and added as a regular paragraph value. For the purpose of ensuring parity of two halves, the value of (f) for large variance and small variation of (0.87) was calculated which is less than the tabular value of (1.96) at the level of significance (0.05) Then the correlation coefficient value was calculated, so the stability coefficient of the test was (85%), and then the Spearman Brown equation after the correction was 88%.

B. The Koder Richardson Equation 20 (K.R.20)
This method is based on the variation in the performance of the individual from one paragraph to another provided that the alternatives are only two, and the Richardson Coder formula (20) represents the average coefficients resulting from splitting the test over parts in different ways. (Abdul Rahman, 1983, 201) Using this method the coefficient was Reliability of test items (0,90).

C. Hurst equation
This equation is used to calculate the stability coefficient when the two halves of the scale or test are not equal in paragraph numbers. This equation also requires extracting the percentage for each part of the test (Faraj, 1989: 317). The value of the coefficient of stability in this

method was (0.91) because the correlation coefficient It reached (0.90) and the percentage of paragraphs for the largest part (51%) and the smallest part (49%)

Statistical analysis of the test according to the theory of underlying features
Verify the one-dimensional assumption and suitability of the Rush model for test data

The test items measure one ability or trait, and this assumption is considered one of the most important assumptions of response theory for the individual, which is to measure the vocabulary of the test only one variable, and for this hypothesis to be achieved, a global analysis of the data resulting from the application of the vocabulary should be used (Warm, 1978: 365). In order to know the level of suitability (Rush Model), one-dimensional for test data, one must verify the assumption of one-dimensional whether the observational data for the responses of the research sample on the test paragraphs, each of them measures only one inherent characteristic and not more, as this assumption is one of the most prominent assumptions of the theory of inherent features The dimension of the two-response type (1, 0) for each individual, in addition to the possibility of this assumption by describing the ability of students through one variable and not more, which represents the attribute or the inherent ability, as this feature must be taken into account in building the test according to the theory of the underlying features (Rush Model). (Kolen & Brennan, 2004: 154)

Estimate parameters of the test items:
The monolithic Rush model was used in this study and using the computerized program Rascal, to scale the test items through the responses of the sample's 500 individuals, and to estimate the difficulty and ability teachers, and the program displays the output (after Put), after the analysis of the data and Figure (7) pages as follows:

After one-dimensional validation, the researcher subjected the test paragraphs of (80) items to statistical analysis according to the Rush model using the computerized program "Raskal", where this program was implemented to scale the test paragraphs, and for each sub-test of the seven tests by analyzing the responses of sample individuals The analysis of (500) individuals, to estimate the teachers of the difficulty of paragraphs and the ability of individuals, where the program displays the output (after put) after analyzing the data in the form of (7) pages, as follows:

1- The first page: as the program displays it when the gradient is zero based on the difficulty of the paragraphs and the ability of individuals, see Tables (10 / a) (10 / b) with a zero gradient.
2- The second page: as the program displays it when the gradient is zero on the basis of difficulty, and the zero is on the power basis, see Tables (11 / A) (11 / B).

3- The third page: as presented by the program, on the basis of a teacher's ability estimate corresponding to each grade, see Table No. (9).
4- The fourth page: as the program displays it when the gradient is zero on the basis of difficulty, and the zero is based on the ability in the case of arranging the paragraphs in ascending order according to its difficulty estimates, see tables (10 / a) (10 / b).
5- The Fifth Page: As shown by the program, which is a map of distributing the difficulties of the paragraphs and the capabilities of individuals and is called the "variable map for individuals and paragraphs" (Figure Byperson Distribution Map). Figure (2) shows the image of this page.
6- The sixth page: as presented by the program, which is the distinct curve of the test (Test characteristic Curve), by linking this curve between the capabilities of the examiners (in the logit unit) and the estimated correct response rate, Figure (3) shows the image of this page.
7- The seventh page: As shown by the program, which is the direction of the test information, it is noted from Figure (4) that the calculated stability of the test is (0.942), the expected information is (16.963), and the calculated information is (13.229). In general, a moderate curve can be considered a good example of information from the test.

The Fourth Chapter

In order to achieve this goal, all necessary steps and conditions have been applied in relation to building the test and calculating the apparent honesty and sincerity of construction through two indicators: the discriminatory force in the two groups method and finding the correlation using the Point by Cyrill correlation coefficient available in the Amos program in the statistical bag of social sciences as well The consistency of the test was calculated by the aforementioned methods in separating the research methodology and procedures, and thus the results showed that the indicators of the test paragraphs had the appropriate specifications for the steps of building the test according to the classic measurement theory, as those procedures were mentioned in the third chapter within the axis of the research tool and its psychometric properties clearly. Thus, the first part of this goal was achieved.

As for the second part of the target, the researcher has followed the necessary steps to follow it to codify the test according to the Rush model of the single parameter model, and the researcher has verified all assumptions of the underlying features theory based on the results produced by the computerized Raskell program, and the researcher extracted the values of the psychometric indicators of the test items that Its results were acceptable, in addition to obtaining the values of the parameter of individual ability and the difficulty of the vertebrae, which is a general factor in the Rush model and the vertical parity of the test items, in addition to following the steps to obtain the results of the psychometric properties of the test according to the model in terms of honesty and consistency of the vertebrae, and thus has

been achieved This part of the goal, as the image of the test whose paragraphs were prepared using the theory of features inherent in accordance with the Rush model has good indicators, and that image was adopted in the results of the research and thus the researcher achieved the goal on both sides, i.e. according to the traditional measurement theory and Rush model.

Conclusions: From the Results of the Research, We Conclude

1- The test items are compatible with their vocabulary and phrases with the current research environment of the university students category.
2- It is possible to judge the level of ability of university students and their interpretation with respect to the peers at each of the test passages, by testing repeated measurements that the results of the Raskal Program came to carry against the difficulty of the paragraphs through the use of statistical indicators for the test according to the underlying features theory (Rush Model) .
3- The researchers' use of methods and models of response theory for the period reduces the criticisms that may exist by specialists in this field because of its great importance in reducing the degree of standard error of measurement.
4 - Most book exercises contain more than one behavioral goal, which confuses students' thinking
Recommendations: Based on the results of the current research, the following recommendations have been formulated: -
1- The research tool that was codified using the Rush model is distinguished from achieving objectivity in measurement, so the current research recommends building achievement tests that are commonly used according to models of underlying features theory, for its control of some of the criticisms directed to these tests.
2 - Attention to establish training courses for all workers in the field of evaluation and psychological measurement and specialists in building tests and metrics on how to fill data using developed computer programs through which to obtain statistical indicators for the appropriate paragraphs with test factors or not appropriate for them according to the Rush model.
3- The introduction of the ability test developed in the educational field is important for the evaluation and early diagnosis of students 'capabilities not only in the advanced stages of study but also in the primary, intermediate and intermediate stages of study because of its importance in developing students' abilities after a period of time on multiple thinking skills.
3- Each exercise includes one application, meaning that it has one behavioral goal.
Suggestions:
Based on the research recommendations, the researchers suggest the following: -
1- Conducting a similar study on the bi-teacher model
2- Conducting a similar study on the three-parameter model
3- Conducting a similar study on other subjects

ARAB SOURCES

Abdel-Messih, Imad. (1991). The use of the monogram parameter Rush model in the analysis of vocabulary tests reference reference bipolar criterion. Journal of Research in Education and Psychology, Minia University, p. 4, pp. 443-475

Abdel Mawla, Shehta. (1999). Evaluation of building tests referred to criterion / criterion in the light of individual response theory and traditional theory, unpublished doctoral thesis, Ain Shams University, Faculty of Education.

Allam, Salah El-Din, (1985): Analyzing the data of mental tests using the probability logarithmic model of Rush: An Empirical Study, The Arab Journal for the Humanities, Kuwait, No. 17, Volume V.

Allam, Salah El-Din, (1995): Future directions for evaluating student achievement in light of the requirements of the twenty-first century, Journal of Education, Al-Azhar University, No. (49): 14-54.

Al-Anazi, Hussein Bin Aziz Hussein, (2006): Building an achievement test to measure the degree of mastery of basic skills in mathematics among students in the upper grades in the elementary stage, Master Thesis, King Saud University, College of Education.

Odeh, Ahmad Suleiman, (1985): Measurement and evaluation in the teaching process, Jordan: The National Press.

Awad Allah, Muhammad Abdul Rahim, (2000): A comparison between the two methods of the Rush model and the traditional method in building an intelligence test using the test of predicting academic achievement, unpublished doctoral thesis, University of Baghdad College of Education / Ibn Rushd.

Awad Allah, Muhammad Abdul Rahim, (2000): A comparison between the two methods of the Rush model and the traditional method in building intelligence tests using the test of predicting academic achievement, unpublished doctoral thesis, University of Baghdad, College of Education - Ibn Rushd.

Esawy, Abdel-Rahman Mohamed, (1985): Measurement and experimentation in psychology and education, Alexandria, Dar Al-Marefa Al-Jami'ia.

Fatihi, Mohamed, (1995): Methods of Measurement and Methods of Evaluation: Building Tests, Examinations, and Handling Results, 1st edition, Casablanca: Dida Kotika Publications.

Al-Qati`, Abdullah Ali, (1993): The relationship between the validity of arbitrators and the statistical concept of the validity of items, Journal of Psychological Studies, No. 1, pp. 53-61.

Kazem, Amina Mohamed, (1988): Using the Rush model in building a test achievement in psychology and achieving an objective interpretation of the results, Tikrit University

Foreign Sources

Adkins, P.C., (1974): Test Construction, Ohio, Abell & Howell Company

Aiken, L.R., (1979): Psychological Testing and Assessment, 3rd ed., New York, Allyn & Bacon.

Allen, M.J.; & Yen, W.M., (1979): Introduction to Measurement Theory, California: Brooks/Cole Publishing Company.

Anastasi, A. & Urbina, S., (1997): Psychological testing, 7th ed., New York: Prentice Hall.

Anastasi, A., (1988): Psychological Testing, 6th ed., New York, Macmillan.

Bechtoldt, H.P., (1959): Construction Validity, A critique American Psychologist, Vol.14, No.5, 619-629.

Berk (1982): Criterion-Referenced Measurement: The State of the Art, 2nd ed., Baltimore, London: The Johns Hopkins University– Press85.

Crocker, L. & Algian, J., (1986): Introduction to classical and Modern Test Theory, New York, CBS College Publishing

Cronbach, L.J. (1960): Essentials of Psychological Testing, New York: Harper and Row.

Cronbach, L.J. and Gleser, G.C., (1965): Psychological Testing & Personal Decisions, Snd.Cd, Urbena III: University of III. Pr110

Curetion, E.E., (1966): Validity, In E.F. Lindquist, Educational Measurement, Sixth Printing, Washington, American Council on Education.

Davis, F.B., (1962): Item Analysis in Relation to Education and Psychological Testing Psychological Bulletin, No.49, 97-121.

Eble, R.L., (1972): Essentials for Educational Measurement, New Jersey, Prentice Hall, Inc.

Elloitl, C. (1983). British Ability Scales Manual a: Introductory Handbook. Windsor England: National Foundation for Educational Research.

Essentials of Psychological Testing, 3rd ed :(1970) ,.

Ghiselli, E.E.; Campbell, J.P.; & Zedeck, S., (1981): Measurement Theory for the Behavioral Science, San Francisco: Freeman and Company

Gray, C.L., (1997): Basic Concepts in Item and Test Analysis Paper Presented at the Annual Meeting of the South West Educational Research Association, Austin, January, P.9.

Gustafsson, J., (1980): Testing and Obtaining Fit of Data to the Rasch Model, British, Journal of Mathematical and Statistical Psychology ,

Hambelton, R.K. et al., (1978): Referenced Testing and Measurement, A Review of Technical Issue and Development Review of Educational Research, 48, 4.467-510.