Verifying the Criterion reference-based hierarchical Assumptions for high-order thinking skills Using Item Response Theory

Asst.Prof. Balqees Hammood Kadhim
University of Baghdad/Collage of Education Ibn Rushd
Department of Education and psychological science
Balqeeshmood24@gmail.com

Received:1/7/2019
Accepted:14/1/2020

Abstract

The present study aimed to investigate the assumptions of the hierarchical structure of skills (thinking of high rank). To achieve this, a reference-based test was built based on the description of the behavioral field of an educational institution according to the North Carolina Classification through the following steps to determine the purpose of the test, the content components and the group to which they will be applied and analyze the topic in detail, verify the validity of the analysis of the test items, extract the standard properties, and then verify the assumptions of the validity of the hierarchical structure of these components.

Keywords: Verification, Hierarchical assumptions, Criterion- Referenced, high-order Thinking skills, Item response theory.

التحقق من افتراضيات البناء الهرمي محكي المرجع لمهارات تفكير عالي الرتبة باستخدام نظرية استجابة الفقرة

أ.م.د بلقيس حمود كاظم
جامعة بغداد/كلية التربية أبن رشد للعلوم الإنسانية
قسم العلوم التربوية والنفسية

المستخلص:

هدفت الدراسة الحالية الى التحقق من افتراضات البناء الهرمي محكي المرجع لمهارات (تفكير عالي الرتبة) وتم الاختبار المحكي بناءا على وصف المجال السلوكي للمؤسسة التعليمية وفقا لتصنيف ولاية كارولينا الشمالية من خلال الخطوات التالية : تحديد الغرض من الاختبار ومكونات المحتوى وتحليل الموضوع بالتفصيل ، والمجموعة التي سيتم تطبيق الاختبار عليها والتحقق من صحة تحليل فقرات الاختبار واستخراج الخصائص القياسية وفق نظرية استجابة الفقرة ، ثم التحقق من صحة الافتراضات الخاصة بصحة البناء الهرمي لهذه المكونات.

الكلمات المفتاحية: التحقق، الافتراضات الهرمية، محكي المرجع، مهارات التفكير عالي الرتبة، نظرية استجابة الفقرة.

Research Introduction

The field of psychological and educational measurement in recent decades has witnessed a remarkable interest by measurement experts in many countries of the world in order to build a test items that helps to accurately measure learning outcomes and come up with objective quantitative estimates of the features that emerge from the responses of students or individuals to those test items.

In the literature of educational and psychological measurement there are different approaches to measurement, namely the standard reference in which the student is classified in the light of the grades of the group to which he belongs. Through real-time statistics, measures of central tendency and dispersion, the distribution should be moderate. The other approach is the criterion-reference in which students are categorized by other statistics derived from the test content indirectly through the cutting score.

Therefore, it appeals to education officials and educators of the importance of caring for the quality of building tests.

Research problem

The weakness of the educational system may be due to a weakness in the tools and means of measurement and evaluation used, so the reform of the process depends on the identification of problems associated with its tools as it is a good approach to reform education and improve the quality of outputs.

The assignment of the tasks of psychological and educational measurement to non-specialists and the expertise of psychological specialists with limited measurement theories. This problem is related to the previous problem in assigning the tasks of building tests to non-specialists. They have received some training and experience in this field, enabling them to properly prepare the tools and interpret their results. The UNESCO Regional Seminar recommended the need for development to take into account modern methods and techniques in this field. (Allam, 2011: 53-54)

Those interested in building tests previously, especially in the educational field, were limited to standardized reference tests that measure a sample of educational tasks in some areas of content, where they include skills and goals that do not represent the content in general, so it is not possible to reach accurate measurements when performing. This reduces the accurate diagnosis of the results and mental skills acquired. Hence the interest in constructing spoken tests that define areas accurately, clearly and sequentially has appeared . (Walsh & Betz, 1995: 17)

In tests, the degree varies depending on the nature of the material, as the degree is more important for the hierarchical content such as engineering and science than for materials of a different nature. The improper transfer of the student to the next unit improperly (error a) hinders the learning of this unit. Accurate classification gives another opportunity to learn the units and skills, while the student does not move mastered to the next unit and classification (error b). It increases the burdens of the educational system on the one hand and the student loses direction towards the goal on the other. (Mahmoud, 2: 2006)

Therefore, there was a need to build a clear and precise test. This is a necessary and important step to move from mastering from one skill to another. This reliability feature in the hierarchical structure helps to exclude any unnecessary sub-skill that does not contribute to the acquisition of the final skill. (Allam, 2001: 59). Thus, the existence of the skill is not urgent in the test as it did not achieve the sincerity of the hierarchy.

Research Importance

officials of education and educational institutions demand the importance of taking care of the quality of measurement and evaluation tools that contribute to making objective decisions on scientific grounds, and that the evaluation is an integral part of the educational system ,where it is to determine the extent to of the objectives of the educational system.

One of the most important types of evaluation that contributes to planning and building processes for educational institutions is the criterion reference test, because it is concerned with measuring and diagnosing the knowledge and skills achieved by the student or the learner and the behaviors acquired in relation to educational content. The grades in this type of tests are interpreted based on the performance of the measured task, i.e. , the location of the individual depends on a number of important rules that contribute and facilitate the process of interpreting the results on these tests, that is, the construction of the tests must be in accordance with sound scientific rules (Crocker & Alging, 1986)

Since the test is a criterion reference, it is closely related to the learning process, the validity of the hierarchical structure of the analytical products on which the test building is based is urgent (Allam, 2001: 86).

Measurement and evaluation methods play an important role in educational decision-making .These methods identify the effectiveness of educational programs and the study writes its importance as it aims to verify assumptions hierarchical construction of thinking skills (high rank) using the theory of response to be objective and accurate measurement. Omnia Kadhim (1984) considers the importance of benefiting from individual response models to solve objective measurement problems reflected in both the field of university education and the field of general education, and in the field of mental measurement and utilization. (Kadhim, 1996: 398)

Therefore, the use of the response theory of the item in light of the advantages enjoyed by this theory in the construction of metrics or the development of these scales to conform to their models, in order to provide objective conditions for the attribute or ability of independent assessment of the characteristics of the vocabulary characteristics of .individuals and independent estimates of individuals from the component vocabulary sample Scale (Anwar AL-Sharkawi et al., 1996-302)

Therefore, the building of any developed society requires attention to the knowledge building, and thinking of the most important pillars of development, especially university education and upgrading requires taking parallel steps towards scientific progress (Hassan, 2004: 1) , where higher education occupies a prominent place in the process of society with all its contributions to meet the needs of human development in various sectors of life.

Theoretical significance.

1. The present research contributes to present a test of thinking skills with a hierarchical structure that helps to be used by the institutions concerned.
2. The use of the theory of item response achieves objective characteristics in measurement.

Research Aim.

The aim of the present research is to present investigate the assumptions of a criterion reference-hierarchical structure of high-order thinking skills using the item response theory.

Research limits.

The current research is limited to university students in Baghdad province.

Definition of Basic Terms

- Assumption in education and psychology for the Pioneering Glossary (Considering the case to be correct based on the available evidence, or taking the contrary to what is accepted and invoked to prove the case , or the assumption of the researcher assumption to resolve a problem (Al-Raed, 1965).

https/www.almougem.com

- Hierarchical construction Defined by :

*** Allam (2001)**

(A structure that is concerned with identifying knowledge and auxiliary skills to be learned in order and sequence of hierarchical construction according to the priorities of their contributions to the formation of the main skill and start with the behavior of the input base of the pyramid and then the levels of behavior range from the simplest to the most complex until it reaches the top of the pyramid, which is the sufficiency or the main skill.) (Allam, 2001: 51)

***Uprichard & phillips 1997**

A method used to analyze a particular skill or behavior in a hierarchical way, forming a series and steps of tasks, each representing a functional level of the given concept.
(uprichard & phillips .1997.7)

- Validation of Hypotheses for Hierarchical Construction.

Walbesser finds that (It is through the dependence of the hierarchical analysis of outputs, i.e. the dependence of each level of outputs on the other. You must achieve the auxiliary goals).

The hypotheses are achieved by five percentages (consistency ratio, ranking adequacy ratio, inverse consistency ratio, inverse efficiency ratio, completeness ratio). (Allam, 2001: 69-74)

Criterion Reference Defined by :

***Kawafihuh (2005).**

(It means the degree sufficiency in limited skills, i.e. its focus is on the extent to which the individual reaches the level of performance in a skill at the level set by the examiner in advance). (Kawafihuh, 2005: 60)

*** Popham (2014)**

(The individual's performance in these reference tests is attributed to the content of the test itself) (Popham;2014;62).

*** Definition of Glaser**

(It is a test in which the details of the content are clearly defined. Through it, you can determine what an individual can and cannot do). (Glaser, 1994:27)

The skill is known by:

*** Fatlawi (2006)**

(A series of procedures or steps that can be observed, whether direct or indirect), measurement, repetition and repeat when needed (Fatlawi, 2006: 349)

*** Shehata and Naggar (2003)**

(A particular operation is carried out with the speed and proficiency with the shortcut in the effort to reach the end). (Shehata and El-Naggar, 2003: 302)

High Order Thinking Defined by :

*** Alatoom (2012)**

(It is a mixture of several forms of advanced thinking and appears in the form of a set of separate mental activities that require making mental decisions, and analysis of complex

situations according to different criteria involving multiple solutions. It avoids simple solutions.) (Alatoom, 2012: 231)

***(Limpman 1991)**

(It is the individual 's preoccupation in formulating predictions and analyzing data through a number of equations and formulas (Limpman. 1991: 98).

*** (Lawrence 2000)**

(It is a thinking that requires the acquisition of the following skills including application, analysis, installation and evaluation through Organization of knowledge and mental openness). (Lawrene, 2000:2)

Item Response Theory defined by :

Allam. (2001):

(That the probability value of an individual's response to a test item is a function of both the attribute or the particular test that the individual is supposed to measure and the characteristics of the item or individual that he or she tries to answer) (Allam, 205: 2001).

Chapter II
Theoretical framework

Criterion – based Measurement

Measurement sciences represents an important input in the study and appreciation of various phenomena, which led to increased accuracy in measurement in the educational and psychological aspects, which helped to understand the phenomenon and to the accuracy of prediction and its control. This made, those made who are interested in measurement to urge to use accurate measurement methods where the so-called criterion – based measurement has appeared to be the, as the true beginning of these tests and methods was in the sixties where credit for the emergence of this term is due to Glasser (1963), who distinguished in one of his studies between the criterion – based test and the norm – based test, which refers to a standard, which was defined by both Glazer and Netco in the educational measurement book published by Thorndike (1971) as tests that include diagnosing and measuring the knowledge, skills, and behaviors of the individual educated person (Glasser & Nitko, 1971: 130).

But it seems that some who are interested in this type of measurement did not use the word criterion in the sense that Glasser meant, so they used the word standard (standard) and did not differentiate what is meant by the word criterion in the standard measurement, since the interpretation of the word criterion led to the diversity of definitions in this type of testing (Allam,1986: 17). This is also has been clarified by Husk and Babam by defining the contents of the criterion – based measurement 1969. These data obtained led to the development of the criterion – based measurement to collect data that the interpretation of the student's degree is absolute and not approximate (Crocker and El-Gina, 2009: 53) (Abu Nahia, 1994: 52 -53).

The explanatory framework for the tests referred to is the criterion – based tests to what the individual can do by specifying criteria or levels of performance or the degrees of cutting to verify the adequacy of the subject and thus there is an analogy between the criterion of achievement and the performance of the individual and thus can translate the degree of the individual in the test directly into behavior that can be due to His observation and measurement, that is, the degree that the individual gets, which is the raw degree, has a meaning in itself (Allam, 2000: 262). In this way, we can explain the results resulting from the criterion – based test as focusing on the tasks themselves and based on a standard that represents the specified acceptable level of proficiency. Which is Similar to the frame of criterion that explains the test results on the performance model for

individual elements. Thus, the performance represents the model for a group of individuals. (Thorndike & Hegen, 1986: 76). Most of those who are interested in educational and psychological measurement emphasize the best use of it, and tests have indicated the criterion of reference in measuring educational outcomes directly (Al-Anzi, 2004: 43-44).

Characteristics of criterion – based tests :

The tests are based on a set of characteristics, as seen by Ababneh (2009). It enjoys an accurate description of the individual's performance in terms of the criterion, as it provides real information in evaluating the effectiveness of the teaching process, and contributes to the evaluation of mastery learning and the final evaluation, and provides an appropriate method for diagnosing learning difficulties (Al-Zayla'i, 2014:18). Also, the term level or criterion does not necessarily imply the full mastery of the connected attribute. Rather, the spoken levels of the examiners can be determined at any point of the caller, and this point is a specific behavior that describes certain tasks that the examiner is able to perform and the comparison is made in light of these levels. These levels do not remain constant for the individual examined, but rather change according to his developmental change (Abdul Salam,1996: 195).

It also depends on a set of measurable behavioral or procedural goals that are more detailed and comprehensive to measure an individual's ability than standard testing requires (Michael and Buffalo, 2008: 23). Whereas, the distribution of grades in it takes the form of a twisted twist and is negative based on elaborate learning. It also employs teaching strategies that rely on self-learning and learning out of mastery (Tawq and Adas, 2001: 522). The ranks of individuals at stake are not affected by the person giving the grades, as they are related to the job that is measured by the test (Al-Shawwara 2015: 14).

Types of criterion – based tests :

There is a diversity around the definitions of the criterion – based tests, which led to a difference in the behavioral range to which the individual's degree is attributed, in terms of an explanation within it, its content, and the degree of complexity. These tests are distributed as follows:

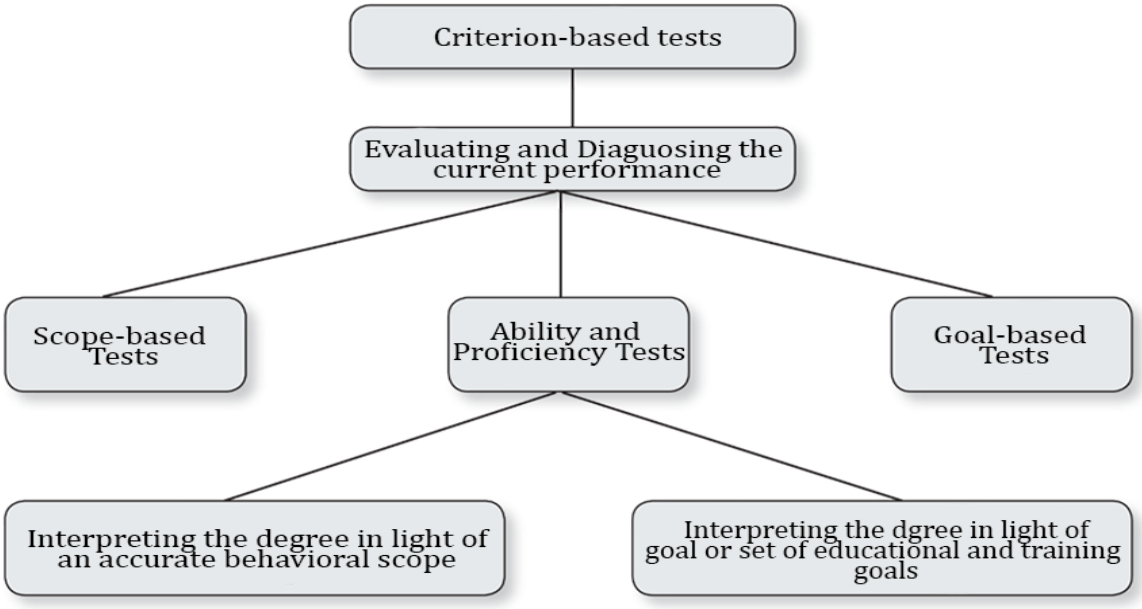


Figure (1) shows the reference of the tests, the purpose of the test and the interpretation of the scores

1- Reference-Scope tests :

It is the one that depends on its construction on a broad and comprehensive identification of tasks or skills, a clear and accurate identification, from which an item sample from this range is randomly chosen or random previews, and the scores of this test are used to obtain statistical estimates of the probability of answering the examined individual or a group of examined individuals for items of The overall scope represented by the items of the test is a true, correct answer at a specific time, which helped to generalize on a comprehensive scale (Allam, 25: 2001) (Abdul Salam, 94: 1996).

2- Proficiency or mastery tests.

The design of this type of test is to determine whether the examined individual has acquired the behavior that the educational program aims to develop. These tests help and contribute to making decisions regarding an individual's mastery of a goal, skill, or a range of specific skills. These may depend on their constructing a range reference or a target reference. Perfection may be complete or relative, with a fluctuation between mastery and lack of mastery. The location of the individual is determined by this continuum (Allam, 25: 2001). These tests, which are based on mastery of learning, performance contracts, and evaluation of educational programs, are among the most important applications of criterion-based tests (Allam, 1986: 92-95).

3- Goal-based Test :

These tests are applied upon completion of a specific program or educational unit with the aim of classifying the individuals tested into two groups, one of which is perfect for the goals and the other is not perfect for the goals in light of a specific percentage of the items that must be answered correctly and identify the goals that no one was able to

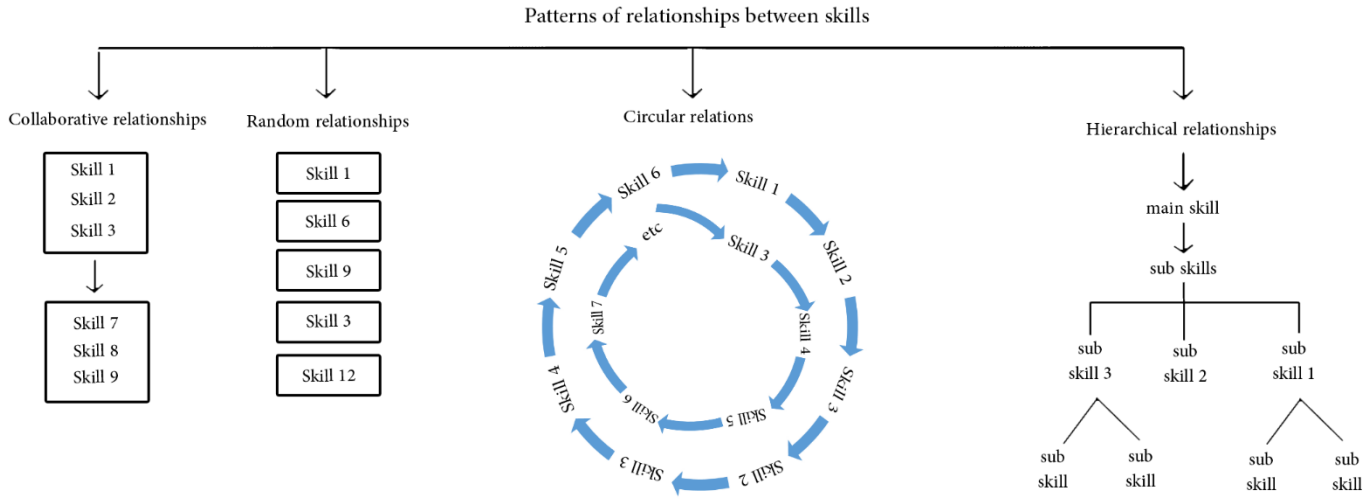
achieve, and they can also be classified in several groups. The construction of reference tests to the target differs from constructing group reference tests that are used to estimate the overall degree of the educated individual about his overall performance, while the goal-reference tests aim to verify the learner's acquisition of basic competencies and skills and reflect specific educational outcomes and to identify the sources of errors (Allam, 2001: 25-41). These tests are used to measure a number or a set of educational, behavioral, and procedural goals, due to the lack of vocabulary, which represents the behavioral range. There are several methods for analyzing competencies and skills at the test site in objective reference tests (Popham, 1973: 93).

Methods for analyzing competencies or skills.

Since analyzing the main competencies or skills of the educational or training program and the processes and knowledge it contains requires knowledge of patterns of relationships between skills and how these skills are arranged relative to each other, there are several patterns that explain these relationships between skills that contribute to knowing the skills that have been achieved and the skills that have not been achieved (Allam : 2001: 45 - 51). And Figure No. (2) illustrates this

Figure No. (2) shows the types of patterns of relationships between skills from the researcher's work

What is important for the subject of the current study is the hierarchical construction



method, we will provide it with some brevity.

Hierarchical construction and analysis method:

To conduct a hierarchical constructive analysis of a major skill or proficiency, we begin with the following question for ourselves. What are the requirements or behavioral components that must be met by the learner in order to achieve proficiency or main skill after defining these requirements or components? We will repeat the same question for each of them, and thus determine the requirements for each of these behavioral components until it reaches the introductory behavior of the learner who was previously learned. Thus, these requirements, i.e. knowledge and supporting skills, are arranged in a hierarchical and constructive way, according to their contribution to the formation of the

main skill, as this behavioral entrance represents the base of the pyramid. Thus the levels of behavior fall from the simplest to the most difficult until we reach the top of the pyramid, which represents the achievement of the main skill (Majeed, 2013:233) (Eisenberg & Walbesser, 1971: 244).

This must distinguish between competencies or cognitive skills that require or include a linear successive series of several tasks, such as solving a complex mathematical problem. This analysis, which requires a written move from left to right in sequential steps, is called an analysis of procedures and as shown in the following figure (2) :

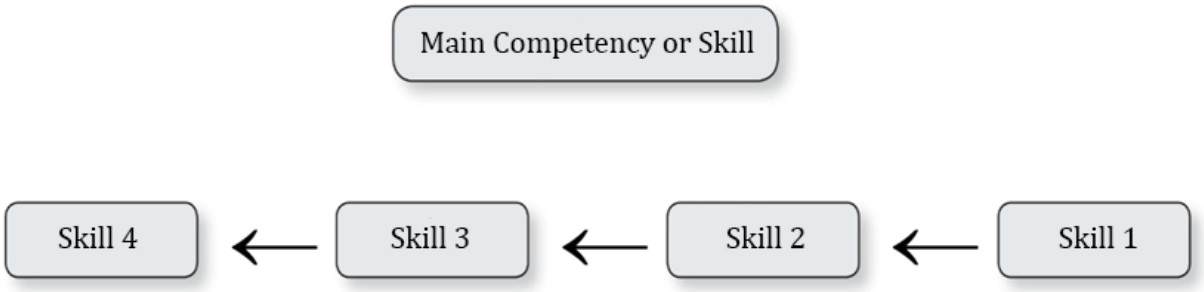


Figure No. (2) illustrates the types of patterns of relationships between skills from the researcher's work, and what matters most to the subject of the current study is the hierarchical method of construction, and we will explain it with some brevity.

Where the sufficiency or the main skill in this form is considered as an umbrella for the four written procedures and thus the main skill or the final behavior is not performed by the learner, but it is only a general description of the set of written procedures where the learner does not move from the performance of the fourth procedure or skill to the performance of the final behavior that Represented by the main skill. Thus, we can through the successive procedures that the main skill has been achieved and between the competencies or skills that require a hierarchical sequence where the main or final skill is formed from sub-sub-skills that help contribute to achieving the sufficiency or the main skill as shown in Figure (2) of the hierarchical relationships, where It is not only an umbrella as in the series of written procedures. Rather, the learner should perform this main skill (concluding), which represents the general subject, because it includes all of these skills, and the sub-skills depend on each other (Allam, 2000: 75-58). There are several hierarchical models that appeared within the new theory, which made a remarkable contribution in the second half of the twentieth century, namely:

Firstly. Hierarchical model . (Rank order)

This hierarchical model depends on the Aristotelian style of classification in relying on the categories and on the categories within the categories, and then the classification method, such as the inverted tree, has its roots higher and its branches down. And it assumes that there are many levels of factors, the higher the level in which the factor is higher, the more extensive and comprehensive nature. Among the most common hierarchical models are the Vernon model and the Fouad al-Bahi al-Sayyid model.

A- The hierarchical model of Fouad El-Bahi El-Sayed.

El-sayed was interested in building the hierarchical model of mental abilities in the book Intelligence (1959) until the year (1976), where he classified the factors into two basic types. They are mutual and individual , so the mutual factors are divided into three other factors, or types are secondary, sectarian, and general. As for the individual factors, they are divided into two types which are private and expatriate. As for the general factors or capabilities, they are divided into two types, which are the general mental

ability or sectarian abilities are divided into four types which are major, and complex. , Primacy and simplicity (Mr. DT: 34-35). As in figure No. (3)

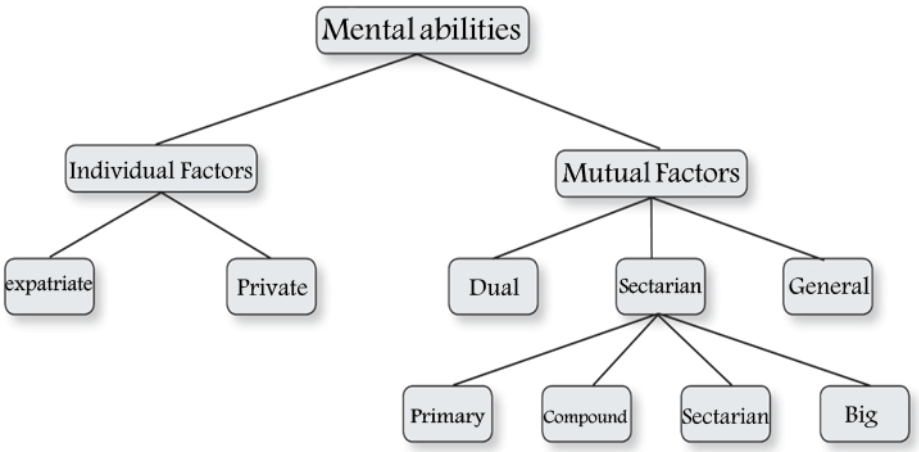


Figure No. (3) shows the hierarchical model of Fouad El-Bahi El-Sayed for mental capabilities

B- Vernon Hierarchical Model.

Vernon organized the mental abilities and classified them as a hierarchical classification organized in four levels and it begins with the general factor and falls into sectarian factors, which in turn are divided into two works, one of which is linguistic - educational, and the other practical - mechanical, which in turn are divided into qualitative factors, each of them related to specific tests (Allam, 2011: 366). As stated in the study of Jassem and Ali (2015), a mental abilities test was built according to the strategy of analyzing the hierarchical structure according to this theory.

Second: The Matrix Model

The matrix has two dimensions: the first with lines and the other with columns. It is possible to increase the dimensions of the matrix model to three dimensions or more, such as the Guilford model and the Qusi model, which includes immediate dimensions (after the content, after the operations, after the outputs). It included the content (content of shapes, symbols, symantec (meanings) and behavioral content (El-sayed, dt, 38-39)). After the operations, which are (factors of cognition, remembering, convergent and divergent thinking, and evaluative factors), and after the results, i.e. what results from performing operations on the content and the resulting thought formula (Allam, 2011: 360) it is, as stated in Al-Bahi, represented (unity, class, relationship, system, transformation, inclusion). This is called the three-dimensional model. The study of both Al-Kahlout and Hamouri (2009) came in building Standard diagnostic tests based on hierarchical structure.

Thinking and Thinking Skills.

There are a number of criteria that are used to judge the quality of thinking of the examined individual in his treatment of an issue, where (clarity, correctness, accuracy, relevance, breadth, logic, importance and consistency) are the most prominent of these criteria to judge the quality of thinking (Elder & Paul, 2001: 42). The nature of the thinking activities practiced varies in terms of their quality, as they may be simple and direct, such as their association with familiar things, including the vehicle as a solution to a complex mathematical problem, or thinking of new solutions. (Zaaloul, 2009: 287)

There are many types of thinking, as there are no two persons who are similar in nature or in their ability to work and think until identical twinning. They differ in the nature of thinking, such as the Almighty saying ((each works according to its similar)). (Al-Isra - verse 84)

This highlights the importance of learning thinking skills and related processes that remain always renewed in terms of their usefulness and uses in processing information whatever the type of this information, where Compay and Sternberg point out, "Knowledge is important, of course, but it may become old. As for thinking skills, they are always renewed. It enables us to gain knowledge and infer it regardless of the place, time, or types of knowledge that are used in thinking skills. "(Quimby & Sternbery).Perkins, (1985) believes that thinking is an intellectual activity that has multiple characteristics and is the ability to perceive the basic relationships in the problematic situation, and destiny it is based on the alternative choice from a large number of available alternatives and the ability to gain insight and organize ideas in order to reach new ideas, and thinking does not grow automatically because it is a process that is not acquired randomly or as an accidental result through attempts to accomplish for other purposes, but rather a process that requires continuous technical and educational performance from A lifelong development for the learner (Razuki and Mohamed, 2019: 13).

This is inferred from thinking through three basic aspects. The first is a cognitive process that occurs within the human mind and is inferred indirectly, and the second is a set of processes that occur within the mind, and the third and last aspect is the behavior directed at every problem (Al-Ghurery: 2007: 14).

Higher-order Thinking

Lipman (1998) argues that high-order thinking is thoughtful and requires self-regulation of the thinking process and always seeks exploration, but when the existing state of thinking is lacking and lacking (concepts, self-organization and exploration) we cannot be seen as high-order thinking (Lippman, 1998: 34) Also, there are a number of characteristics, as Resnick (1987) put it:

1- It includes self-organization, that is, thinking in thinking requires a kind of autonomy through evaluation and self-monitoring.

2- It gives multiple solutions instead of giving a unique solution.

3- The way it works is not predetermined and determined by mathematical relationships and functions.

4- Discover new meanings and concepts

5- Interested in knowing and acknowledging causal or logical relationships. (Al-Atoum et al., 2006: 201-203)

Assumptions Underlying High-order Thinking:

1- These skills are amenable to learning because they prepare for the individual and the learner to adapt and engage in life in order to serve the community. It also serves to define alternatives and priorities.

2- The previous ideas of the learner interact in order to achieve mental goals that can be added and modified. Where learning occurs through re-organization and prior information with new information to be a new mental activity. Resulting in personal and social interaction (Razzuki, 2019: 60).

Riznak believes that high-level thinking is a set of cognitive mental activities that require a mental trial of different situations according to specific criteria. Riznak (1985) classified high-ranking thinking skills in five skills (problem-solving skills, decision-union skill, critical thinking skills. , Metacognitive Skills, Creative Thinking Skills) (Retnonaite & et.al, 2017: 416).

As for the North Carolina classification of high-ranking thinking skills prepared by the Department of Education for the State of North Carolina in 1994, it is a modified

version of the classification of the American Association for Curriculum Development and Education prepared by Marzona and his colleagues in 1988. All of the skills (focus, information gathering, remembering) were combined and included in one field which is knowledge, and another area of Bloom's taxonomy is added. Bloom's taxonomy is similar to Piaget and Vygotsky's assumption that complex forms of thinking cannot be achieved until the simpler forms are managed, which was neglected by Marzona and his colleagues in the classification of the American Association for Curriculum Development (Razuki and Muhammad). , 2019: 69).

But Al-Rimawi and others (2008) see that there are so-called basic thinking skills that include (observation, recall, coding and setting goals, questioning, classification, comparison and application) and there are high-thinking skills that require from an individual an advanced level of mental and mental treatments to be practiced effectively and includes (organized thinking) , High-ranking thinking, creative thinking and metacognitive critic) (Rimawi et al., 2008: 320). This is, consistent with what was mentioned in the 1988 Marzona and colleagues classification. This classification includes four main areas and their sub-skills, which are as follows:

Whereas, this hierarchy of high-order thinking skills and their meanings of skills as stated in Al-Otoum, et al. 2006, Zuqi and Muhammad 2019 are:

First: Analyzing skill:

It includes identifying and examining relationships by distinguishing components and attributes. The following sub-skills are included :

1- Determination of properties and components: It is the identification of properties and parts of a universe through its stored knowledge and clarification of these parts that make up the whole.

2- Defining relationships: It is the one that enables the learner to clarify the type of these relationships between the components, whether they are temporal or partial relationships, part-to-whole relationship, reason, outcome, or others.

Second: Generating Skill:

It is a skill that involves generating and producing new meanings and ideas through previous knowledge and includes the following sub-skills:

1- Conclusion: It is the logical identification of the available information and coming up with something that is correct and sound.

2- Prediction: The educated individual anticipates the upcoming events as a result of the information obtained.

3- Expansion: It is the ability of an individual to add many details, clarification and information related to previous knowledge to improve the understanding process.

Third: Integrating Skill:

It is the ability to integrate previous knowledge and new information to build a new understanding. The following sub-skills are included:

1- Summarization: It is the exclusion of the unimportant and the choice of the important and the plural of the true dimensions of the subject.

2- Reconstruction: It is the ability to change the cognitive structure in order to enter and integrate new and previous information related and united to form a new cognitive constructive understanding.

Fourth: Evaluating Skill:

It is a structured process for gathering and analyzing information, i.e. estimating the reasonableness of ideas and results. It is a judgment on something according to a certain standard and includes two skills:

1- Standards Setting: The ability to develop to judge the quality of ideas. These standards are based on and derived from the experiences and expertise that the educated individual is going through.

2- Verification: It is the ability to confirm allegations made about a case. (Razzouki and Mohamed, 2019: 96-72) (Al-Otoum and others, 2006: 209-227).

And Figure No. (5) shows the hierarchical structure of high-order thinking skills according to the classification of Marzona and Rimawi, which is prepared by the researcher and will be followed by an explanation of these main skills, which includes (the skill of analysis, generation, integration, evaluation) with its sub-skills that constitute the main skill, which is high-ranking thinking.

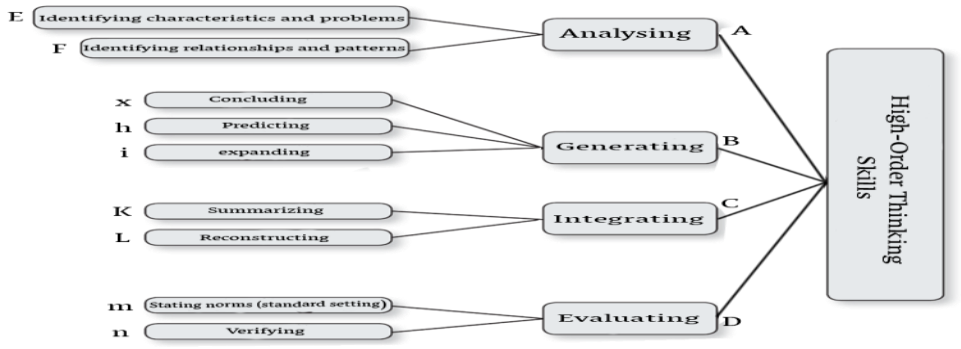


Figure No. (5) illustrates the hierarchical construction of high-order thinking skills (by the researcher)

Item Response Theory

This theory is considered one of the most suitable theories in analyzing and interpreting the items of psychological and educational tests. The response theory of the item (IRT) assumes that the performance of the examined individual can be predicted, or that their performance can be explained in a psychological or educational test, in light of a distinctive feature of this performance called the trait, and this feature is difficult to notice directly; therefore it must be estimated or inferred from the performance of the examined. It can be observed on a set of scale or test items (Hambleton, Swaminathan & Rogers, 1991:62).

Where several models have emerged for this theory, and from these models the Rush model is the simplest model of potential traits , and the main idea behind this model is that each item carries an emotional charge that contributes with other items in the scale in the formation of a total emotional charge that reflects the direction of the individual in accordance with his appreciation of those items according to the number of probable mathematical function used in the scale, and the model estimates this charge for each item according to the probabilistic mathematical function approved by the model, then verifies the conditions for the suitability of the items for the model (Ahmed Odeh, 1992: 155).

It is also characterized by the lack of the number of assumptions that are required from the data in order to give the model accurate estimates, and the ease of estimating the index of difficulty and statistical capacity. Fisher indicates that the raw score expresses the only information required in the data to estimate the ability, that is, all people who participate in a degree will be estimated for them the same ability (Fisher, 1973: 100)

The response theory of the item is based on several assumptions that must be verified in the data in order to perform its work objectively. These results can be trusted. Below is an explanation of the most important of these assumptions, as stated by Hambelton & Swaminathan (1985) and Swamnthian, which are as described below :

First: Unidimensionality: That is, the test items measure only one ability that explains the individual's performance on the item, meaning that all items measure one dimension.

Second: The assumption of local independence: Local Independence ,is the response of the subject to an item influences, positively or negatively, his response to any other paragraph, meaning that the subject's responses to the test items are statistically independent from the tested sample, i.e. free from the sample.

Third:ICC-Curve Characteristic Item : The item Properties curve is a mathematical function that links the probability of a correct answer to the item (P) and the ability of the examiner θ) which is measured by a set of items in the test that was built for that purpose.

Fourth: Assuming speed of performance: That is, the examiners who fail to answer the test items are due to their limited ability and not because the time is not sufficient to reach the answer to the item. (Hambelton & Swaminathan, 1985: 61)

Chapter III
Procedures

The Research Sample:

The aim of selecting the sample is to reach sound conclusions about the indigenous community from which the problem originated. This is done by choosing a category or sample that represents this community correctly.

In the current research, two samples were selected from the study population:

1. A sample of clarity of instructions and items, amounting to (50) male and female students.
2. Statistical analysis sample: The purpose of it is to analyze data derived from its responses to test items, and the sample size was (250) students who were randomly selected from the colleges of University of Baghdad (collage of Education Ibn Rushd's, Science, Law, Sharia'a, Languages, Engineering).

Tool of the study:

The researcher used the test of higher order thinking skills according to the classification of Carolina attributed to North Carolina (1994).

Test description:

The test consists of (63) items for various questions, which are pictorial and verbal items, distributed by seven items for each sub-skill of the nine skills, in addition to that the construction of its items was of a multiple choice type and are objective tests. The test consists of major and sub-skills. As explained in one of the figures of the second chapter. items are statistically analyzed according to the Rush model:

A- One-dimensional verification:

A one-dimensional assumption is a prerequisite for all response theory models, and this requirement is required in ability tests. Where Recasase Recas, (1979) proposes that one-dimensional verification can be achieved by examining the values of Eigen values. If global analysis gives one distinct factor, so that the potential root value of that factor forms a clear and high percentage of the total degree variance, then this means There is one dominant factor in the test, which indicates a one-dimensional assumption has been fulfilled (Wiberg, 2004: 5).

To verify this hypothesis, the researcher used the global analysis by means of the basic components (Principle Component) using the statistical program (SPSS) and according to

each of the (nine) skills of the test a total score of the statistical analysis sample of (250) examined. And the result was identical to the results of direct analysis (before rotation) depending on the lower limits (for two meetings) which are the potential root factor that can be explained equal to or more than one (Abdul Khaleq, 1983: 148).

Table (1)

The underlying root of each component of high-order thinking, the proportion of explained variance, and total variance

Component number	Skill	Sub skill name	Latent root	Explanation variance	Total variance
-1	analysis	Identifying characteristics and problems	2.244	33.112	7.331
		Identify relationships and patterns	6.131	51.138	11.294
-2	Generating	The conclusion	9.334	62.263	16.471
		Predicting	5.611	58.102	18.337
		Expansion	8.441	64.222	12.118
-3	integration	Summarizing	4.532	42.220	7.105
		Reconstructing	4.136	68.591	14.628
-4	Evaluating	Stating norms	9.111	25.413	8.332
		Verification	7.326	42.004	14.516

It is clear from the table that the components of the test are saturated with the general factor according to the lower limits of (GETMAN). The amount of saturation of the items of each component is by the general factor.

Table (3)

The Items are satiated with the first factor of each component

Sub-Skill No. / item number	First	Second	Third	Fourth	Fifth	Sixth	Seventh	eighth	ninth
1	0.632	0.371	0.641	0.659	0.476	0.461	0.431	0.418	0.511
2	0.522	0.392	0.750	0.418	0.872	0.639	0.629	0.375	0.673
3	0.396	0.816	0.667	0.843	0.379	0.388	0.611	0.591	0.447
4	0.481	0.381	0.812	0.491	0.322	0.510	0.512	0.692	0.546
5	0.694	0.539	0.612	0.335	0.558	0.418	0.414	0.450	0.710
6	0.331	0.562	0.411	0.522	0.413	0.764	0.620	0.602	0.366
7	0.442	0.382	0.409	0.496	0.582	0.622	0.601	0.511	0.340

From the above table, there were no items that were saturated less than (0.30).

B - Estimate the parameters of the test items:

Items were statistically analyzed according to each of the parameters of the difficulty of the paragraph and the ability of individuals as shown in tables No. (4, 5, 6, 7-8, 9, 10, 11, 12).

Table (4)

Values of the parameters of the difficulty of item and the capabilities of individuals for the skill of Identifying characteristics and problems

Item No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The first skill	**Distinction index: 0.426 / freedom degree: 14 / tabular value at (0.05) (23.68) level is**					
1-	0.612	0,253	11.534	*****	*****	0
2-	0415	0,130	9.623	4.321	0,217	105
3-	0,721-	0,182	20.672	7.311	0,156	16
4-	0,681	0,302	8.428	4.691-	0,327	22
5-	1,232	0,213	17.423	6.205	0,211	31
6-	2,115	0,177	14.531	0.481	0,528	13
7-	0,695	0.421	6.621	0.839	0.645	63

Table (5)

Values of the parameters of the difficulty of items and the ability of individuals to determine the relationships and patterns

Item No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The second skill	**Distinction index: 0.522 / Freedom degree: 11 / Tabular value at (19.68) (0.05) level**					
1	0.584	0,639	11.673	*****	*****	0
2	1.116	0.279	8.561	0,734	0.511	20
3	0.734	0.451	6,114	1.643-	0.472	35
4	0,620	0.731	16.734	0.497	0832	78
5	0.729	0.816	7.881	0,417	0.860	18
6	2.734	0.395	19.002	0,936	0.421	47
7	0,771	0.184	13.830	0.361	0.381	52

Table (6)

Values of the parameters of the difficulty of items and the abilities of individuals to deduce skill

Paragraph No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The third skill	**Distinction index: 0.615 / Freedom degree: 22 / Tabular value at (0.05) (33.92) level is**					
1	0.845	0.548	22.853	***	****	0
2	0.295	0.429	18.842	6.720	0.221	65
3	0.719	0.129	8.812	0.730	0.629	28
4	0.318	0,510	20,820	4.712	0.410	39
5	1.632	0.328	9.618	1.840	0.395	44
6	0.421	0.621	17.634	0.623	0.171	50
7	0,336	0.655	31.814	1.846	0.715	24

Table (7)

Values of the parameters of difficulty of Items and the ability of individuals to predict skill

Item No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The fourth skill	**Distinction index: 0.503 / Freedom degree: 9 / Tabular value at (0.05) (16.92) level is**					
1	0.263	0.429	5.612	****	****	0
2	0.612	0.321	14.672	1,412	0.552	46
3	0,419	0.522	9.831	1.852	0.729	41
4	0.391	0.212	2.639	0.892	0.402	72
5	1,110	0,720	7.923	1.944	0,488	32
6	1.836-	0.329	2.003	3.845	0.520	9
7	0.703	0.184	5.437	1.099	0.284	122

Table (8)

Values of the parameters of the difficulty of items and the capabilities of individuals for expansion skill

Paragraph No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The fifth skill	**Distinction index: 0.489 / Freedom degree: 16 / Tabular value at (26.30) (0.05) level**					
1	1.745	0.111	20.882	****	****	0
2	1.722-	0.734	18.927	1.271	0,447	67

3	1.943	0.429	9.104	4.110	0,589	73
4	0,943	0,416	18.956	1.909	0,720	10
5	0,720	0,629	11.942	3.911	1,013	49
6	0,431	0,274	5.002	1,117	0,629	51
7	0,551	0,178	17.663	4.821	0,417	5

Table (9)

Values of the parameters of difficulty of items and the abilities of individuals to summarize skill

Paragraph No. / Skill	**Difficulty coefficient - b-**	**Standard error S.E**	**The value of the Chi square**	**Ability index O**	**Standard error S.E**	**Repetition Freq.**
The sixth skill	**Distinction index: 0.412 / Freedom degree: 41 / Tabular value at (55.76) (0.05) level**					
1	0,432	0,339	37.110	*****	****	0
2	0,322	0,537	19.004	4.723	0,462	21
3	0,619	0,221	51.561	1.947	0,720	84
4	0,390	0,119	11.894	8.341	0,616	15
5	0.965	0,658	24.046	5.204	1,428	33
6	0,734	0,418	7.004	1.671	0,529	52
7	1.836-	0,620	10.884	7.119	0,405	45

Table (10)

Values of the parameters of the difficulty of items and the abilities of individuals to reconstruct skill

Paragraph No. / Skill	**Difficulty coefficient - b-**	**Standard error S.E**	**The value of the Chi square**	**Ability index O**	**Standard error S.E**	**Repetition Freq.**
The seventh skill	**Distinction index: 0.631 / Freedom degree: 26 / Tabular value at (38.89) (0.05) level**					
1	0.723	0,557	35.839	***	****	0
2	0.384	0,212	7.640	3,845	0,313	29
3	0.612	0,769	19.945	2,111	0,482	17
4	0.261	0,528	17.107	3.512	0,312	41
5	0.629	0,428	7.934	1,412	0,184	47
6	0.412	0,115	4.645	6.645	0,130	34
7	0.143	0,178	33.241	8.223	0,412	82

Table (11)

Values of the parameters of difficulty of items and individuals' capabilities for standard setting skill

Paragraph No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The Eighth Skill	**Distinction index: 0.558 / Freedom degree: 11 / Tabular value at (0.05) (19.68) level**					
1	0.561	0,261	2.830	****	****	0
2	0.823	0,285	5.912	2.423	0,731	21
3	0.341	0,184	12.845	1.836	0,538	70
4	0.412	0,167	19.005	7.615	0,820	49
5	0.319	0,144	17.931	5.003	0,619	58
6	0.712	0,175	11.256	3.012	0,481	73
7	0.519	0.216	15.002	4.174	0.629	0

Table (12)

Values of the parameters of difficulty of items and the capabilities of individuals for verification skill (proof)

Paragraph No. / Skill	Difficulty coefficient - b-	Standard error S.E	The value of the Chi square	Ability index O	Standard error S.E	Repetition Freq.
The ninth skill	**Distinction index: 0.684 / Freedom degree: 43 / Tabular value at (55.76) (0.05) level**					
1	0.391	0.174	17.520	*****	*****	0
2	0.732	0.418	5.118	1.840	0.992	66
3	1.634-	0.152	33.915	7.392	0.382	25
4	0.629	0.293	54.956	5.830	0.151	35
5	0.311	0.155	12.851	3.122	0.129	74
6	0.451	0.204	9.441	2.583	0.205	22
7	1.224-	0.281	14.295	2.849	0.418	28

Through tables No. (4, 5, 6, 7, 8, 9, 10, 11, 12) and viewing degrees of freedom and Chi calculated values for all skills and comparing them with the tabular value at the level (0.05) and as recorded in the tables above, there did not appear that the value The calculated value is greater than the value of the Chi square so the items have good coefficients.

Fittness test:

Kadhim, (1996) indicated that there are three basic criteria on which to choose the items that correspond to the assumptions of the model and delete the inappropriate items, which are:

The first criterion is that the item in its definition of the variable agrees with the rest of the items in the test:

When the value of the value is a statistically significant, the item should be deleted because it does not reflect the attribute expressed by the rest of the items, and after the researcher conducted data analysis of all the skills, no item was deleted since the value of the Chi-square is statistically significant at the level of (0.05).

The second criterion: that the item is independent of the sample:

The statistical analysis of the nine skills that represented high-order thinking skills showed that there is one discriminatory power, as the discriminatory power of the nine skills (0.412, 0.426, 0.489, 0.503, 0.522, 0.588, 0.615, 0.631, 0.684). And this test has been verified.

The third touchstone. The items should have an appropriate discriminatory power:

The following criterion, which is level (0.01), was chosen for the current research, and it is noted that the discrimination factors fluctuated between (0.412) and (0.684), and this indicates that the discrimination factors for the components of each component are within the acceptable range.

Objective measurement assumptions:

This assumption was validated by two indicators:

1. Carrying out the factor analysis of the total score for each of the six components. The factor analysis showed the presence of a general factor.
2. Analysis of the items of each component according to the "Rush Model" and the extent to which the items conform to the model's assumptions.

Second. Independence of measurement:

1. The measurement independence from the sample capacity that performs the test:

The researcher divided the statistical analysis sample for the nine skills into two samples, one of which is "high level", and the other "low level" after arranging the members of the total sample analysis of (250) individuals for each component. The high level sample consists from the first to the individual No. (125) and the other low from individual no. (126) to the individual (250), then the results of the responses of each sample were analyzed separately, in order to calculate the difficulty of the items and their standard errors, and the estimates of the power and its standard errors, and then a comparison was made for the indexes of the model (the difficulty and ability) as it is derived from analyzing the performance of the total sample, and from The two samples (high level and low Level), and to verify the statistical equivalence of these estimates, and the estimates are statistically equal Asymmetric if any difference between the two estimates did not exceed the total standard error for them as stated in (Kazim 0.526: 1996).

Then, the ability estimates corresponding to each possible total score derived from the performance of the total, high-level and low-level samples, and their standard errors, were extracted for the nine sub-skills that constitute a high-order thinking skill, as all differences were less than the sum of the standard error of the two estimates, and this indicates that the measurement The power is released from the sample to which the test was applied.

2. Independence of measurement from the items that individuals answer:

To verify this assumption, the test was divided into two tests (easy and difficult), according to the difficulty factors as calculated by the Raskall program, then the data for each test was analyzed using the Rush model method, as the difficulty of the item and the ability of individuals to both tests were estimated, then the results of the analysis of the two tests were compared with the results of the overall test. Then, the vertical tie was made between the easy and difficult tests.

It turned out that all estimates of the difficulty of the items were statistically equivalent, and the difference between any two analogous estimates did not exceed the sum of the standard error for them, which indicates that the measurement was freed from the difficulties of the test items.

The researcher then compared the total scores of five individuals from each component, derived from the easy and difficult tests, the reference gradient, the mutual gradient, and its standard errors.

And it turned out that all the differences were less than the sum of the standard error of the ability of the individuals corresponding to each grade of raw for the two grades (difficult, easy and total), and this indicates the liberation of the measurement from the application sample.

Validity and Reliability of the Tests:

The use of the Rush model in developing a test means the availability of objective requirements in the measurement of the power variable subject to measurement, and this implies that the validity and reliability conditionalities of the estimates of both test items and the capabilities of individuals have been verified (Kadhim, 1988: 98).

While the stability of the measurement is fulfilled by achieving the remaining objective requirements, which is the independence of measurement, that is, the independence of the measurement from the test items applied to the sample, and its independence from the sample to which the test is applied (Kadhim, 1996 A: 367). This is confirmed by the current research.

Final Description of The Test Components:

After completing the procedures for analyzing the test components and scaling them according to the test of Rush model, (63) test items are divided into (9) sub-skills according to the values of the parameters of difficulty of the items for each skill and the ability of the examiners is estimated in the log unit as calculated by the program.

Converting the Log Unit Gradient to the Centigrade Unit (watt):

To get rid of negative signs and fractions, the researcher converted the estimates from the log unit to the centigrade watt unit.

Verifying the hierarchical construction assumptions for high-order thinking skills:

First: Verification of the hypotheses related to the levels of analysis:

After validating the assumptions of the item response theory for high order thinking skills and verifying its honesty and consistency, the skills test was applied to (100) male and female students through which the researcher resorted to checking the validity assumptions of the hierarchy of each sub-skill of high-order thinking skills, as follows:

The first hypothesis: related to the skill of analysis (A) which includes the sub-skills (Identifying characteristics and problems) (E) and the skill of identifying relationships (F) .

The second hypothesis: related to the skill of generation (b) which includes the sub-skills (deduction skill) (x) and skill (prediction) (h) and skill (expansion) (i).

The third hypothesis: related to the skill of (combination and integration) (c) which includes the following sub-skills (summarization) (k) and skill (reconstruction) (l).
The fourth hypothesis: related to the skill of evaluation (d), which includes the following sub-foals (setting standards) (m) and skill (verification or proof) (n).
The fifth hypothesis: which includes the previous four hypotheses combined, which represents the main skill which is the skill of high order thinking, which is realized according to (A, B, C, and D).
The chosen hypothesis is for the individual to achieve the fifth hypothesis which he has to achieve the four hypotheses related to it and when the individual passes the degree of cutting he will have achieved that goal or skill and gets a score (1) while he gets a zero degree if he does not reach the degree of cutting.
Where the data were blanked in the form of arranged pairs, as follows to calculate the five ratios, namely (consistency ratio, arrangement adequacy ratio, reverse consistency ratio, inverse efficiency ratio, and completion ratio), Some possible outcomes for arranged pairs are as shown in Table (13).

Table (13)
The arrangement of some possible outcomes (arranged pairs) related to skill

1 1-1	1 0-1	1 1-0	1 0-0	0 1-1	0 1-0	0 0-1	0 0-0	
(1-1)	(0-1)	(0-1)	(0-1)	(1-0)	(0-0)	(0-0)	(0- 0)	

Equations for extracting ratios for the five hypotheses:

$$\text{Consistency ratio} = \frac{N(1,1)}{N(1,1)+N(1,0)}$$

$$\text{Arrangement competence ratio} = \frac{N(1,1)}{N(1,1)+N(0,1)}$$

$$\text{Completion ratio} = \frac{N(1,1)}{N(1,1)+N(0,0)}$$

$$\text{Reverse consistency} = \frac{N(0,0)}{N(0,0)+N(1,0)}$$

$$\text{Reverse competence} = \frac{N(0,0)}{N(0,0)+N(0,1)}$$.(Allam, 2001)

Table No. (14) shows data of the four hypotheses of high-order thinking skill.

Table (14)

Data and products for the five hypotheses of high-order thinking

Hypothesis number	Arranged couples	Iterations	Hypothesis number	Arranged couples	Iterations	Hypothesis number	Arranged couples	Iterations
First	(0-0)	12	**Third**	(0-0)	11	**Fifth Concluding**	(0-0)	8
	(1-0)	4		(1-0)	5		(1-0)	10
	(0-1)	6		(0-1)	9		(0-1)	12
	(1-1)	78		(1-1)	75		(1-1)	70
Second	(0-0)	8	**Fourth**	(0-0)	8			
	(1-0)	16		(1-0)	14			
	(0-1)	7		(0-1)	20			
	(1-1)	69		(1-1)	58			

Table (15)

The five proportions of the five hypotheses of high-order thinking

Hypothesis number	Consistency	Arrangement competence	Completion	Reverse consistency	Reverse competence
The first	0.93	0.95	0.87	0.75	0.66
the second	0.91	0.81	0.90	0.33	0.53
The third	0.89	0.94	0.85	0.69	0.55
Fourth	0.74	0.81	0.88	0.36	0.26
Fifth	0.85	0.88	0.90	0.44	0.40

As it appears from the table above that the values of the consistency ratios and the adequacy of the arrangement and completeness of the five hypotheses were from 0.85 or higher so they are indicators of the validity of the hierarchical structure, which is an indication of the adequacy of the arrangement for the hierarchical construction (Walbesser, 1968)

Second: Verifying the Validity of the Hierarchical Stuctare:

The different hierarchical structure varies in the degree of its sincerity. Therefore, the researcher verified the validity of the hierarchical structure by finding the ratio of the number of individuals who achieved the final skill to the total number of sample individuals. As the value of the sincerity of the hierarchical structure reached (0.70), this value is a measure of the effectiveness of the hierarchical structure as it represents an integrated hierarchical structure.

Third: Factories of pyramid complexity:

Hierarchical structures vary in the degree of their complexity, according to the number of supporting goals (sub-skills) in each of their hypotheses. This coefficient indicates the average number of auxiliary targets included in a particular hypothesis included in the construction. The hypotheses of the current research are four hypotheses, while we find that the supporting goals or sub-skills are (2-3-2-2), respectively, and the length of the hierarchical structure is (4), so the coefficient of the complex of the hierarchical structure

is (2.25) and in this way it is a construction A hierarchy of four related skills is (A, B, C, D).

Conclusions:

The researcher reached a number of conclusions as follows:

1. The effectiveness of the Rush model in developing the test, the current research tool, by matching test data to model assumptions.
2. The availability of several assumptions related to the approved model or objectivity requirements in the measuring instrument is essentially a test of the instrument's validity and reliability.
4. An estimate of the difficulty of the item was released on the ability of the individuals who answered the test.
5. The independence of the measurement from the group of items to which the examined respond
6. Checking the affinity of the subject's ability with the difficulty of the item, and this indicates that the measurement is free from the difficulty of the item.
7. Hierarchical construction can be analyzed by keeping the sub-skills and checking the sincerity of the hierarchy for educational skill or sufficiency or canceling the sub-skills that increase the complexity of the hierarchy.

Recommendations:

In light of the researcher's findings, the researcher recommends the following:

1. The current study recommends the use of item response theory theory models in developing and building skills for major usage ideas to overcome the criticisms leveled at these metrics.
2. Providing computerized statistical programs to facilitate the process of analyzing test data according to the theory of underlying features.
3. The Ministry of Education and other relevant ministries are adviced to take advantage of this test of high-order thinking skills after verifying the sincerity of its hierarchical structure, in selecting students and guiding them to different disciplines.

References :

1- A,Edward Uprichard , Evan R.Phillips .1997 An Intra concept Analysis of Rational Number , Ahierarchy for.R.B.Kana.Florida dournal of Educational Research

2- Al-Atoum, Adnan Yousef, 2012. Development of thinking skills, theoretical models and practical applications. Amman, Dar Al-Masirah.

3- Al-Fatlawi, Suhaila Mohsen Kazem, 2006. Effective educational and teaching curriculum, Dar Al-Shorouk Publishing, Amman, Jordan.

4-Abdul Salam, Nadia Mohammed. 1996. Constructing the objective testing of the reference and group reference (comparison study) in Anwar Al Sharkawi and his colleagues, contemporary trends in educational and psychological measurement and evaluation. Cairo, the Anglo-Egyptian Library.

5- Al-Anazi, Hussain Bin Aziz. 2004. Building an achievement test to measure the degree of mastery of basic skills in mathematics for students in the upper grades in the elementary stage, unpublished Master Thesis, King Saud University. Faculty of Education .

6- Al-Atoum, Adnan Yousef and Al-Jarrah, Dhiab Abdel Nasser and Bechara, Mowaffaq. 2006. Development of thinking skills, theoretical models and practical applications. Jordan Amman. Al Masirah House for Publishing and Distribution.
7-Al-Ghurery, Saadi Jasim, (2007). Teaching thinking (its concept and contemporary trends) Iraq. Baghdad . Al-Mustafa Press.
8- Al-Hammouri, Abdel-Majid, and Al-Kahlout, Ahmed Ismail, 2009. Suggested strategy for building reference test diagnostic tests. Volume X, Issue 2, Journal of Educational and Psychological Sciences. Kuwait
9- Allam, Salahuddin Mahmoud. 1986. Contemporary developments in psychological and educational measurement. Al-Azhar University, College of Education, Kuwait University.
10-_____2000. Educational and psychological measurement and evaluation, its fundamentals, applications, and contemporary trends. Cairo . Arab Thought House.
11- _____ . 2001. Diagnostic tests benchmark reference in the educational and psychological areas, Arab House of Thought Cairo.
12- _____2011. Educational and psychological measurement and evaluation, its fundamentals, applications, and contemporary trends. Arab Thought House. Cairo.
13 - Al-Rimawi, Muhammad Awda, and others. 2008. General Psychology, 1st floor, Dar Al-Masirah. Ammaan Jordan
14 - Al-Zali, Muhammad Bin Ali. 2014. Building a reference-oriented diagnostic test to measure scientific research skills for graduate students. A magister message that is not published . College of Education, Umm Al-Qura University, Makkah Al-Mukarramah.
15 - Anwar, Muhammad al-Sharqawi, Suleiman al-Hadari, Umniah Kazem, and Nadia Muhammad Abd al-Salam 1996. Contemporary trends in measurement and psychological and educational evaluation. The Anglo-Egyptian Library. Cairo .
16 - Crocker, Gina, C, 2009. An Introduction to Traditional and Contemporary Theory of Measurement (Zeinat Dana Translated). Amman Dar Al-Fikr.
17 - Eisenberg&H.Walbesser .1971 . Learning Hieraarchies : Numerical Considerations PP.244-256 . htts/www.jstor.org.
18 - Elder, Linda&Paul , R.2001. Gritical Thinking to some purpose . Journal of Derelopmental Education , 25 (1)
19 - Fischer ,G. H. (1973) : The Linear Logistics model as an instrument In educational Research and psychological , voI .37
20 - Glaser , R.1994 B:Griteerion-Referenced Tests part 2, Unfinished Business . Educational Measurement Issues and practice, Washington,Win,Vol . 13, No.4.PP.27.30
21- Grocker, & Algina .1686. Introduction to classical and modern test theory , New York : Holt . Inc
22- Glasser R . Nitko .1972. The standard problem meaning and valuesin measurement and evaluation, New Orleans, August. 30,1974. New jersey
23− Hambleton, R.K. Swaminthan, H. (1985). Item Response Theory: rinciples and Application, Kluwer. Nijhoff Publishing, Boston.
24- Hassan, Wafaa, 2004. Teaching Teaching Strategies, George Washington University, Center for Applied Languages.
25- Hambleton, R., Swaminathan, H. and Rogers, J., Fundamental of Item Response Theory, Second Edition, Newbury Perk California, 1991.

26- Kadhim, Amina Muhammad. 1996 b. Using the Rush model in building an achievement test in psychology and achieving an objective interpretation of the results, in Anwar Al Sharkawi and his colleagues. Contemporary Trends in Psychological and Educational Assessment and Evaluation, The Anglo-Egyptian Library. Cairo .
27- kawafihah, Tayseer Mefleh, 2005. Measurement, evaluation, measurement and diagnosis methods. 2nd floor. Jordan. Amman, Al Masarra Publishing House.
28- Lawrence Erlbaum .N.J.2000.Perspective on thinking Learning and Cognitive Styles. Associates. The educational psychology series English , Online.
29- Lipman, 1991. Learning to think .Thinking to learn ; strengthening Reasoning and Judgment Through Philosophy , (Eds) UK.Oxford
30- Mahmoud, Sumaya Shukri. 2006. The effectiveness of some degree estimation methods in predicting subsequent achievement in engineering for middle school. Unpublished search. Master Thesis submitted to the Faculty of Education, Mina University. Egypt.
31- Majeed, Sawsan Shaker. 2013. The foundations of building psychological and educational tests and standards, Jordan. Amman . National Books House.
32- Michael, Matmanius, Buffalo, Yasser. 2008. Educational evaluation. Appointment training. Syria . Ministry of Education, Directorate of Preparation and Training.
33- Molenaar,I.&Hoijtink, H.1996. person-fit and the rasch Model with an Nakamura.Y.2001. Rasch Measurement and Item Banking Theory . education Resources Information.
34- Mr. Fouad Al-Bahi. Dt. Theories of mental capabilities, development of innovative capabilities, and special talents.
35- Odeh, Ahmed, Salman (1992) 0 The extent of compatibility between the Rush model and the traditional indicators in selecting the paragraphs of the seven-stage trend scale, Journal of the College of Education, UAE University, Issue (8) June, 153-179
Pdf.www.pdffactory.com
36 - Pioneer . Gibran Masoud. Arabic lexicon
37- popham , W, James .2014 . Griterion Referenced Measurement ; Half a century Wasted Educational Leadership .
38 - Razuki, Raad Mahdi, Muhammad, Nabil, Rafeeq. 2019. The Thinking Series and its Patterns Part 3. Objective thinking, high-level thinking, totalitarian, and Al-Sayer, rational, irrational. Lebanon . Beirut, Scientific Books House.
39- Retnowait,E,Ghuforn,A&atl.2017. Character Education for 21st century . Global Citizens .Universities Negeri Yogyakarta Indonesia (INCOTEPP 2017) 21-22 October. International conference of teaching education.
40 - Shehata, Hassan, Al-Najjar and Zainab (2003). Encapsulated educational and psychological terms, Cairo, Egyptian Lebanese House
41- Shawawreh, Shadi Yusuf. 2015. The tests are based on the reference construct and analysis (a comparative study), Jordan, the Islamic Scientific College.
42-Thorndike, Elizabeth Hagen, 1986. Measurement and evaluation in education and psychology. Translated by Abdullah Zaid Al-Kilani, Abdul Rahman Adas. Amman, Jordanian Book Center.
43- Touq, Mohy El-Din, Adas, Abdel-Rahman. 2001. The foundations of educational psychology. I 1. Jordan Amman . House of thought for publication.
44-Walbesser, H. (1968). An evaluation model and its application. Washington, D. C.: American Association for the Advancement of Scienc.
45- Zaghloul, Emad Abdel Rahim. 2009. Principles of educational psychology. Al-Maysara Publishing House. Jordan. Amman . I 1