Contemporary Studies Journal in Education and Psychology
P-ISSN: 3079-773X

Vol 1, Issue 1, March 2025

النظرية التقليدية الكلاسيكية في القياس قراءة تحليلية

The Traditional (Classical) Theory of Measurement: An Analytical Reading*

أ.د. ضرغام جاسم محمد النعيمي & علي محمد أحمد معلا

Prof. Dr. Dhurgham Jasim Mohammed Al-Nuaimi[1] & Ali Mohammed Ahmed Malla Al-Zakholi[2]

[1] College of Physical Education and Sports Sciences, Mosul University, Iraq, **Email:** Dhurghamjasim@uomosul.edu.iq

[2] Department of Physical Education and Sports Sciences, Faculty of Education, Zakho University, Iraq, Email: Ali.al-zakholi@uoz.edu.krd

* receiving date: 03-01-2025

* Accepting date: 15-02-2025

* Publishing date: 15-03-2025

الملخص:

هدفت هذه الورقة إلى تسليط الضوء على نظرية الاختبار الكلاسيكية التي تُعد واحدة من أبسط النظريات وأقدمها في القياس النفسي والتربوي وتتتبع الورقة تطور نظرية الاختبار الكلاسيكية منذ بدايات القرن العشرين على يد رواد مثل تشارلز سبيرمان وغالتون ولورد وتوفلر كما هدفت إلى توضيح المفاهيم الأساسية للنظرية مثل الدرجة الحقيقية ودرجة الخطأ والثبات والصدق مع إبراز دورها في تطوير الاختبارات وركزت الورقة أيضًا على التطبيقات العملية لنظرية الاختبار الكلاسيكية في قياس السمات والقدرات البشرية مع التأكيد على تأثيرها في تفسير درجات الاختبارات. ومن النتائج التي توصلت إليها الورقة تميز نظرية الاختبار الكلاسيكية بين الدرجة الملاحظة والدرجة الحقيقية ودرجة الخطأ تعكس الدرجة الحقيقية القدرة الفعلية بينما تمثل درجة الخطأ المؤثرات العشوائية ويُحسب الثبات كنسبة تباين الدرجة الحقيقية إلى التباين الكلي وعلى الرغم من ظهور نظريات جديدة مثل نظرية الاستجابة للمفردة إلا أن نظرية الاختبار الكلاسيكية لا تزال مستخدمة على نطاق واسع نظرًا لبساطتها وسهولة تطبيقها.

الكلمات المفتاحية: نظرية القياس التقليدية الكلاسيكية، نظريات القياس السيكومتري والتربوي، تشارلز سبيرمان، غالتون توفيك، الدرجة الحقيقية، درجة الخطأ، الثبات، الصدق، الاختبارات النفسية والتربوية، قياس السمات البشرية، قياس القدرات العقلية.

Abstract:

This paper aimed to shed light on the Classical Test Theory (CTT), which is one of the simplest and oldest theories in psychological and educational measurement. It traces CTT's development from early 20th-century by pioneers like Charles Spearman, Galton, Lord, and Toffler. Also, it aimed to clarify the basic concepts of the theory like true score, error score, reliability, and validity, highlighting their role in test development. The paper also focused on practical applications of CTT in measuring human traits and abilities, emphasizing its influence on test score interpretation. The results that the article reached to are: CTT distinguishes between observed score (X), true score (T), and error score (E). True score reflects actual ability, while error score represents random influences. Reliability is calculated as the ratio of true score variance to total variance. Despite appearing new theories like Item Response Theory (IRT), CTT remains widely used due to its simplicity and ease of application.

Keywords: Traditional (Classical) Measurement Theory, Theories of Psychometric and Educational Measurement, Charles Spearman, Galton, Tuffik, True Score, Error Score, Reliability, Validity, Psychological and Educational Tests, Measurement of Human Traits, Measurement of Mental Abilities.

The Classical Measurement Theory:

The Classical Measurement Theory clearly emerged through the efforts of Spearman in 1932, where he presented the mathematical foundation for his intelligence theory (the two-factor theory), including terms such as "true score," "reliability," and "validity." While the dominant statistical theory relied on indicators from Karl Pearson since 1900, the classical theory was largely built on the concepts of correlation.

The interaction between measurement methods and statistical mathematical theories on one hand, and psychological theories on the other, led to the emergence of measurement theories, including the classical theory. In the 1930s, several scattered statistical methods emerged, with the goal of uniting them within a comprehensive scientific framework, resulting in the establishment of the Journal of Psychometrika in 1935, which focused on statistical equations in the field of measurement. In Britain, other journals such as Education and Psychology emerged, which focused on psychological and educational measurement (Allam, 2007: 96).

The Classical Measurement Theory emerged in its current form through Galton in 1950 with his book Theory of Mental Tests, which is considered one of the pioneering works in this field. Galton provided a comprehensive and detailed presentation of the classical measurement theory in a precise mathematical style, which helped to establish fundamental concepts such as "true score," "reliability," and "validity."

This theory was founded on the efforts of Spearman in the 1930s and had a significant impact on the development of psychological and statistical sciences. It integrated the statistical concepts of correlation theory with psychometric measurement methods, leading to the development of a wide range of tools and tests used in research and practical applications.

Among the important contributions in this field, specialized scientific journals emerged, such as Psychometrika in 1935, which addressed

statistical equations in measurement, as well as British journals like Education and Psychology, which focused on psychological and educational measurement. These journals played a significant role in spreading and developing these theories on a wide scale (Al-Zubaidi, 2018: 42).

Education Measurement: Theory and Practice

In 1968, Lord and Tovick developed and applied the Classical Measurement Theory within using it widely in the fields of education and psychology. Their work was based on modern statistical and mathematical methods, which led to a reformation of the theory's fundamental attributes. Their book Statistical Theories of Mental Test Scores is considered one of the most important references in this field, as it presented the latest innovations in the theory. Then, in 1982, Thorndike continued to develop this theory through his works, most notably his book Educational Measurement: Theory and Practice.

The Classical Measurement Theory is characterized by its simplicity and ease of understanding, which helped in its widespread adoption among teachers and professionals in the field, despite the emergence of newer theories. The theory is based on fundamental principles such as reliability, validity, and true score, which were central to improving measurement tools and psychological and educational tests (De Batisti, 2004: 14).

The roots of the Classical Measurement Theory go back to the English psychologist Charles Spearman, who between 1904 and 1913 presented a well-known set of logical and mathematical evidence regarding the nature of scale data (test scores) and the measurement errors they are subject to. His research showed that the correlation between observed scores is lower than the correlation between true scores due to the influence of these errors. Spearman laid the foundations for a model known as the Classical True-Score Model, which is also the term used for this theory. This model contributed to providing an accurate understanding of the characteristics of

psychometric measurement, aiming to distinguish between the true score, which represents the ability (or trait/characteristic/behavior) measured by the scale (test), and the observed score affected by random errors (Ebel, 1972: 10).

The Classical Measurement Theory is based on a key fundamental assumption in the design of tests and measures related to psychological and educational fields, which states that individuals' scores (research sample) on the trait, behavior, or characteristic measured by the test or scale follow a normal distribution. This distribution reflects the hypothetical distribution of true scores and is considered a common model for interpreting human traits. This distribution is influenced by several factors, most notably the nature and characteristics of the sample of individuals being tested, as well as the statistical features of the test items, such as difficulty and discrimination. Based on this, this assumption is used in result analysis to estimate individuals' true scores and determine the appropriateness of the test for its intended purpose.

This theory is based on the concept of the principle of individual differences, which assumes the existence and availability of variations between individuals in behavior, traits, characteristics, or attributes being measured. This principle is considered the foundation of psychological measurement (Kazem, 1990: 36).

According to this theory, the score an individual obtains on a test is an imperfect function of the amount of the trait or attribute being measured, due to irregular factors or chance variables. Therefore, the score obtained by the individual is called the observed score or the error-prone score. The actual amount of the trait possessed by the individual is called the true score.

The difference between the observed score and the true score is referred to as the error score. The error score may cause the observed

score to be very high at times, while at other times it may cause it to be very low. Therefore, it is essential to analyze the nature of both true variance and error variance, as well as the relationship between them. To achieve this, reliability must be calculated, which refers to the ratio of the true variance of the score derived from the test (Mahasneh, 2013: 138).

The Concept of Classical Measurement Theory

Classical Measurement Theory, also known as True Score Theory, is one of the oldest and simplest theories in the field of psychometrics and educational measurement. Despite its simplicity, this theory formed the foundation for the development of tests and measures in the early 20th century. Although significant advancements have occurred in the field of measurement in recent decades, such as Item Response Theory (IRT) and Generalizability Theory, Classical Measurement Theory continues to be widely used in many areas. Modern theories are considered more complex and precise in handling measurement data, offering analytical tools that address the weaknesses associated with True Score Theory, such as its sensitivity to sample nature and item characteristics. However, the classical theory retains its importance as a conceptual and practical foundation in the field of measurement.

Traditional measurement methods still have a strong impact until today. Many current tests are based on the traditional foundations of measurement, demonstrating the effectiveness and continuity of these methods in providing accurate data characteristics (Mahasneh, 2013: 95).

The primary aim of Classical Measurement Theory is to estimate the reliability of measurement by analyzing the relationship between observed scores and true scores. This theory is sometimes referred to as True Score Theory, due to its reliance on the well-known mathematical model called the True Score Model.

This theory focuses on distinguishing between the true part of the score, which reflects the measured trait or characteristic, and the part resulting from random errors. Through this, it provides a fundamental tool for estimating the reliability of tests and measures related to both the psychological and educational fields, making it a cornerstone in the analysis of tests since the early development of measurement tools (Mikhail, 2015: 99).

The basic idea of Classical Measurement Theory

Classical Measurement Theory, also known as True Score Theory, is one of the oldest known theories in the field of psychometrics and educational measurement. This theory is based on the principle and foundation of two main concepts: the true score and the error score.

The theory assumes that if a test is administered to an individual multiple times using a variety of different items and under varying conditions, the observed scores will fluctuate due to the influence of random errors. However, the average of these scores is considered an unbiased estimate of the individual's true ability (behavior) or true score.

Thus, the theory serves as a foundation for understanding how random factors intersect with observed measurements, allowing researchers to estimate the reliability and accuracy of measurement tools and develop more reliable tests.

Due to the presence of errors in any type of measurement, the observed score also reflects the impact of several factors such as:

1. Time.
2. Type of questions.
3. Psychological conditions, and others.

This makes it different from the true score, and additionally, we cannot apply the test multiple times to the individual. In other words, there is a difference between the observed score and the true score due to measurement errors. This can be mathematically expressed as follows:

$$X = T + E$$

Where:

X = Observed score

T = True score

E = Error score

This means that the observed score (X) for each student is the algebraic sum of the true score (T), which reflects the student's actual level, and the error score (E), which arises from what is called random error. Therefore, the test must be designed in a way that allows us to measure each student's true score. In other words, all precautions must be taken to minimize the impact of random error, so that its proportion in the observed score is as small as possible. (Al-Dosari, 2000: 96)

Basic Assumptions of Classical Measurement Theory:

The basic assumptions of this theory that explain the relationship between the true score and the error score are

1. The observed score for the scale or test is the sum of the true score and the error score.
2. The average of the errors for all individuals' scores is equal to zero.
3. The correlation between the true score and the error score is equal to zero

4. The correlation between the error score in identical tests and the error score in any other version of the same test or a different test is equal to zero (Oliver, 2024: 30).

Although this theory is characterized by the simplicity of its assumptions and ease of application, it primarily depends on the sample used, which limits the possibility of applying its tests to other populations (Hernandez, 2009: 40).

The normal distribution of scores refers to the distribution of scores on the ability continuum measured by the test, where the majority of scores are centered around the mean, with a gradual decrease in frequency as the scores move away from this center toward the extremes. This distribution reflects the fundamental assumption in many psychological and educational tests, which is that human traits or abilities tend to be distributed normally, with most individuals at the average level, while the extremes represent those with very high or very low abilities. The number of individuals at a particular ability level corresponds to the expected number according to the probability density function.

$$(E\pm) + T = X$$

The observed score (X) is the sum of the true score (T) plus the measurement error (E), where (E) can take negative or positive values. Thus, it is expressed by the equation:

- In Classical Measurement Theory, the variance of raw scores is defined as the sum of the variance of true scores and the variance of error scores. This means that the total variance of the observed scores in the test reflects two essential components: the first represents the true differences in the traits or abilities measured by the test, and the second represents the variance caused by random errors or imperfect measurement conditions. This mathematical relationship is fundamental to understanding how random errors affect observed scores and how to

minimize them to improve the accuracy of the test. Furthermore, the correlation between the two sets of raw scores derived from parallel tests, or the reliability of the tests in a given population, is expressed by the ratio of the variance of true scores to the variance of raw scores. Finally, increasing or decreasing the length of the test affects the reliability of its scores. It is important to note that the true score cannot be directly known or measured, but can be inferred or estimated from the observed score. This is done by calculating the average of the independent observed scores resulting from applying the test many times, with the accuracy of estimating the true score increasing as the number of administrations grows.

- There is no direct correlation between the true scores and the measurement errors of the examinees in the test. This means that it is not necessarily the case that the measurement errors for examinees with higher true scores are smaller than those for examinees with lower true scores.
- There is no correlation between the measurement errors of two different tests on the same examinee.
- Measurement error is considered a random error that affects the accuracy of measurement, and this concept is also referred to as test reliability, which reflects the extent to which measurement errors are minimized. On the other hand, systematic error is associated with the concept of measurement validity or test accuracy, which refers to the extent to which the test accurately and consistently measures the intended trait or characteristic without bias.
- This theory assumes that the variance of measurement errors is equal for all examinees who are administered the test or measurement tool.
- The assumptions of Classical Measurement Theory may be violated in some cases that affect test results. Since we are typically unable to

determine the true score (T) or the measurement error (E), or to validate the suitability of these assumptions, what we can do is to estimate when these assumptions will be appropriate or applicable (Al-Saykhan & Al-Momani, 2021: 140)

Criticisms directed at Classical Measurement Theory:

1. The unequal measurement units and its linearity.
2. Its limitations in addressing some modern psychometric challenges, such as creating and managing question banks, which require more advanced measurement models to ensure test fairness and adaptability to different individual abilities. Additionally, the theory struggles with the construction of criterion-referenced tests and the use of computers in constructing and interpreting certain types of tests.
3. The scores of individuals on any test depend on the nature of the item sample included in the test, as this sample directly affects the accuracy and validity of the measurement. Additionally, Classical Measurement Theory assumes that the variance of measurement errors is equal for all individuals on whom the test is administered, which may not always be accurate in real-world environments (Wright, 1997: 35).

Norm-referenced measurement:

This system emerged under the educational philosophy that prevailed in the early 20th century, which focused on classifying individuals based on their relative positions among a known group of peers in various abilities and skills. Psychologists played a major and more influential role than educators in the development of this type of measurement. For example, Alfred Binet, the first person (scientist) to design a test aimed at distinguishing mentally disabled children from normal children, created a foundational tool for measuring mental abilities. He also contributed to the emergence of the first collective test used for classifying officers and

soldiers before the period of World War I, which aimed to enhance the specific role of norm-referenced measurement. The success of this test greatly encouraged the adoption of quantitative measurement methods on a broader scale, leading to rapid development of measurement tools used in both educational and military fields. (Al-Jubouri, 2012: 95)

Norm-referenced measures, which are the most common and widely used, rely on measuring the individual differences between people in order to distinguish them. In these tests or measures, an individual's score can only be understood when compared to a standard based on the performance level of a group or peer group to which they belong.

This standard is represented by the average scores of this group or peer group, while the comparison is determined based on how much an individual's score deviates from the value of this average. These deviations are calculated using statistical measures such as the standard deviation, allowing for an estimate of the individual's relative position within the group and determining how they compare in terms of excellence or lag when compared to their peers (Al-Jubouri, 2012: 95)

With the spread of the concept of Mastery Learning, the focus in the educational field shifted from measuring individual differences among students to measuring the extent to which they achieve the defined educational objectives. This shift led to moving away from relying on the traditional normal distribution used in norm-referenced measures, as educational activities aimed to ensure that all students master what they learn.

Glaser is considered the first scholar to draw attention to this new approach in the measurement process, and he was the first to coin the term "Criterion-Referenced Tests." Later, Carver further reinforced this direction by being the first to use the term "Criterion-Referenced Edumetry" as an alternative to "Psychometrics" (norm-referenced measurement).

This development highlighted the fundamental difference between the two approaches; criterion-referenced measurement focuses on evaluating the extent to which individuals achieve pre-established standards or criteria, while norm-referenced measurement aims to compare an individual's performance to that of their peers.

The goal of criterion-referenced measurement is to determine the level of an individual's performance and assess their progress toward the established measurement objectives, regardless of their standing among their peers (Abd Al-Majid, 2010: 116).

Limitations of Measurement Classical Theory:

- Despite the widespread use of traditional test theory by practitioners and researchers in the field of measurement and testing, it has some limitations in analyzing test results. Some of the key shortcomings of this theory include:
- Reliance on sample characteristics: Psychometric properties derived from the traditional or classical approach, such as difficulty indices, discrimination indices, and reliability, are closely linked to the characteristics of the sample of individuals tested. These properties are also influenced by the difficulty level of the test items. This connection indicates that the accuracy and reliability of measurements in the traditional framework are not absolute but depend on the nature of the chosen sample and its diversity, as well as the test design itself.
- Limitations in generalizability: As a result, the classical approach is limited in generalizing its findings beyond the sample used, highlighting the need for more advanced measurement models (Hassan, 2006: 10)
- It is assumed that the scores obtained by individuals on the items of a test or scale should form a linear scale, meaning that the scores reflect a regular increase in performance or the level of the ability or trait being

measured. It is also assumed that all items of a given scale (test) are related to the variable or trait being measured and carry the same meaning or concept for all individuals (test-takers), regardless of their backgrounds or individual characteristics. This assumption is fundamental to traditional measurement theory, where the design of tests is based on the premise that all items contribute equally to the evaluation of the targeted trait.

- However, this assumption may face challenges in some contexts, especially when individuals interpret test items differently due to variations in cultural backgrounds or levels of understanding. Furthermore, this scale is typically represented as a curve. The fixed difference between two test scores may have different meanings depending on the location of the scores along the continuum of the trait or ability being measured (Hassan, 2006: 10).
- It is assumed that test scores reflecting the measured trait or ability show a steady, linear change, meaning that any increase in an individual's scores indicates a proportional rise in the amount of the trait or ability they possess. However, this assumption may not always hold true in reality. For instance, some individuals with high abilities may obtain low scores due to factors such as misinterpreting test items, stress during the test, or other conditions unrelated to their actual ability. Conversely, some individuals with lower abilities may achieve high scores due to luck or guessing correct answers. Therefore, interpreting the meaning of score changes requires considering additional factors to ensure fairness and accuracy in the results (Brahimi, 2017: 298).
- The composition and meaning of a test can change over time, meaning that the nature of the test items and their interpretation may evolve for the target groups the test is designed for. These changes can result from various factors, such as shifts in environmental conditions or

differences in testing circumstances that may not be consistently or adequately standardized. Additionally, any alterations made to the test items, whether through deletion, addition, or modification, may lead to unexpected changes in individuals' scores. These changes can significantly affect how well the items represent the behavioral domain they intend to measure, thus reducing the accuracy and validity of the measurement. Therefore, maintaining the stability of the test composition and considering environmental and temporal factors is critical to ensuring the accuracy and reliability of the results (Brahimi, 2017: 298).

- The results of individuals on a particular test are highly dependent on the characteristics of the item sample included in the test. This means that drawing different samples of items that vary in difficulty, even if all are from a large group designed to measure the same ability or trait, will lead to differences in the expected scores for individuals. This variation arises from the difficulty level of the items included in each sample. Whether the items are drawn from the same original group or from an equivalent group, the variation in the properties of these items directly affects the individuals' results. Therefore, controlling the properties of the items and ensuring the equivalence of the samples is crucial for ensuring fairness and accuracy in measurement (Al-Shamrani, 2014: 717).

- Focusing on only one measurement objective without considering other goals, such as estimating an individual's performance relative to other groups they do not belong to, can limit the scope of the assessment. For example, an individual's performance could be compared with the performance of individuals from different age groups (Al-Shamrani, 2014: 717).

- The estimation of an individual's score and the statistical processing of the results, including the calculations involved, are based on the assumption that these scores are equal units. However, these scores

estimate individuals' levels without adequately measuring the distances between these levels in a convincing way. The difference in the distance between consecutive scores in the test can lead to variability in the meaning and quantitative content of any specific score difference across the range of the test. In other words, the difference between two consecutive scores may not reflect the same amount or significance across the score scale, which affects the interpretation of differences between individuals. This discrepancy highlights a challenge in achieving quantitative consistency within the test, which may lead to variations in the accuracy and objectivity of the evaluation, especially when comparing individuals' performance based on these scores (Al-Shamrani, 2014: 717).

- There is no common scale for individuals and items. When observing the total scores for each item and individual, it is noticeable that the lowest observed scores correspond to the most difficult items in the test or measurement tool, which actually represent the higher levels of that variable. On the other hand, the highest scores correspond to the most capable individuals. Thus, individuals and items cannot be placed on the same continuous scale, making it difficult to interpret individuals' scores relative to the item scores (Al-Shamrani, 2014: 717).
- The traditional theory of measurement does not provide a clear explanation of the pattern of performance of the individuals being tested on the test items (study tool), despite the importance of this explanation in identifying the true causes behind some unusual responses, and thus making the appropriate decision regarding these responses (Basel, 2016: 189).
- The traditional (classical) theory of testing focuses more on the test as a whole than on each individual item within the test. For example, the True Score model does not provide considerations regarding how an

individual responds to a specific item, which is important in many test applications. When a test designer aims to construct a test for a specific group of test-takers, such as a test to differentiate between applicants for a particular certification, it becomes important to estimate the likelihood that a test-taker will provide an appropriate response to a particular question. This estimation is valuable when designing a test that matches the test-taker's level of the trait being measured (Basel, 2016: 189).

One of the main criticisms of traditional measurement methods is the lack of measurement objectivity, which is reflected in the variation of results based on both the test used and the sample level in the case of norm-referenced measurement, or the benchmark level in the case of criterion-referenced measurement. Therefore, the characteristics of the test-taker and the test itself cannot be separated; rather, both should be interpreted in light of each other (Basel, 2016: 189).

References:

1. Abdel-Majid, Nabil Abdel-Ghafour (2010): Using the Rasch Model in Analyzing Items of the Raven's Standard Progressive Matrices Test, Dhakar Research Journal, Baghddad: Al-Mustansiriya University.
2. Abu Fuda, Basel (2016): The Compatibility between the Classical Measurement Theory and Item Response Theory in the Matching of Criterion-Referenced Items in Analytical Geometry Unit, Arab Studies in Education and Psychology Journal, Cairo: Issue 73.
3. Alam, Salah El-Din (2007): Criterion-Referenced Tests in Educational and Psychological Fields, Egypt: Arab Thought House.

4. Al-Dosari, Ibrahim Mubarak (2000): The Reference Framework for Educational Evaluation, 2nd Edition, Riyadh: Arab Bureau of Education for the Gulf States.
5. Al-Jubouri, Rashid Saleh Mahdi (2012): Building an Achievement Test for Educational Supervision and Administration Based on the Latent Trait Theory for Teacher Training Institute Students, Baghdad: Al-Ustath Journal, College of Education Ibn Rushd.
6. Al-Saykhan, Rehab bint Saleh & Al-Momani, Rana Thani Daman (2021): A Comparison between Classical Theory and the Three-Parameter Model in Selecting Items for an Achievement Test in English, International Journal of Educational and Psychological Studies, Cairo: Volume 10, Issue 1.
7. Al-Shamarani, Mohamed (2014): Using Item Response Theory and Classical Theory to Estimate Psychometric Statistical Characteristics of a Critical Thinking and Learning Skills Test for Foundation Year Students at University, College of Education Journal, Issue 157, Volume 2, Cairo: Al-Azhar University.
8. Al-Zubaydi, Mohamed bin Hassan Yahya (2018): Building a Criterion-Referenced Achievement Test in the Course of Tests and Measurements Using the Unidimensional Rasch Model for Diploma Students in Education at Taif University, International Journal of Specialized Educational Studies, Volume 7, Issue 2.
9. Brahimi, Samia (2017): Studies in Psychological and Educational Measurement between Classical Theory and Contemporary Theory, Algeria: Mohamed Boudiaf University.
10. De Batisti, F. Salini (2004): Statistical calibration of Psychometric tests.
11. Ebel, Robert L. (1972): Essentials of educational measurement, Englewood Cliffs prentice Hall, New Jersey, USA.

12. Hassan, Sayed Mohamed Abu Hashem (2006): A Comparative Study between Classical Theory and the Rasch Model in Selecting Items for the University Students' Study Entrance Measure, Zagazig: College of Education Journal, Vol. 52.

13. Hernandez, R. (2009): Comparison of the Item Discrimination and Item Difficulty of the Quick-Mental Aptitude Test using CTT and IRT Methods. The International Journal of Educational and Psychological Assessment, April 2009, Vol. 1, Issue 1.

14. Kazem, Ali Mahdi (1990): Building a Standardized Scale for the Concept of Self Among Secondary School Students, Iraq, First College of Physical Education, University of Baghdad: Unpublished Master's Thesis.

15. Mikhail, Emptanious Nayef (2015): Psychological and Educational Measurement and Evaluation for Normal and Special Needs Individuals, Amman, Jordan: Al-Aisar Scientific Publishing and Distribution.

16. Muhassna, Ibrahim Mohamed (2013): Psychological Measurement in the Light of Classical and Modern Theories, Riyadh: Dar Jareer for Publishing and Distribution.

17. Oliver, Dean (2024): Basketball beyond Paper: Insights into the Game's Analytics Revolution, USA: University of Nebraska Press.

18. Write, B. D. (1997): Measurement Essentials, Wide Range, Inc.