

Applications of analyzing paragraphs of psychological and cognitive tests using the coefficients of difficulty and discrimination and their role in strengthening and stability

Rafahiya Bouchareb ¹, Kadri Abdelhafid ², Trifa Mohamed ³

¹ Universiy Batna2, Institute of Physical and Sport Activities, Algeria, <https://orcid.org/0000-0001-9491-3350>

² Universiy Batna2, Institute of Physical and Sport Activities, Algeria, <https://orcid.org/0000-0002-4318-3260>

³ University of Mohammed Seddik Benyahia Jijel (Algeria), mohamed.trifa@univ-jijel.dz

Received: 11/2023, Published: 12/2023

Abstract

The research paper aims to identify the methods of analyzing paragraphs of psychological and cognitive tests, where we focused on clarifying the methods of calculating the difficulty level and its degree of discrimination, and their significance between the higher and lower levels in the trait or ability measured by the test. In light of a specific internal or external test, the purpose of paragraph analysis is to improve the degree of sincerity and reliability of the total test, in addition to knowing the good sequence of paragraphs.

Keywords: Paragraphs of psychological and cognitive tests, Difficulty and discrimination coefficients, Honesty and stability.

Introduction and research problem:

Since the test or scale includes a set of vocabulary, paragraphs or questions, the characteristics of the test It also depends on the characteristics of its vocabulary, and in order to build an effective test, the vocabulary it contains should be carefully examined using qualitative descriptive methods and quantitative statistical methods, in order to analyze the test items (Item Analysis), especially the statistical characteristics of each of its vocabulary.

These characteristics are related to the level of difficulty , and the degree of its discrimination between the higher and lower levels in the trait or ability measured by the test in the light of the test. As an internal or external criterion, the internal criterion may relate to performance in a group of vocabulary in general or the internal consistency of vocabulary, and the external criterion relates to performance outside the scope of the test itself, such as success in a particular study, or in performing a certain work, and the attention in both cases is focused on assessing the validity of each vocabulary. (Salah al-Din Mahmoud Allam, 2000, p. 266.)

The purpose of paragraph analysis is to improve the degree of sincerity and stability of the total test, in addition to knowing the good sequence of paragraphs, by examining the evaluation of each paragraph of the test to determine its validity and strength to

distinguish between Those who have a strong ability, and those who have a weak ability in a certain field, and the analysis of paragraphs goes through two important steps, the first is the study of the difficulty of the paragraph, and the second is the study of the significance of distinguishing the paragraph. (Moqaddam Abdel Hafeez, 2003, p. 136)

The tests have been formulated in an accurate, sound and balanced manner so as to lead to the achievement of the objectives for which they were designed through the availability of the required specifications for a successful test that contains the elements of objectivity, honesty, stability or reliability, to be more effective When applied, he must have the ability to reveal the capabilities of the examinee with that process, i.e. the process of examining the examinees' responses to each paragraph to evaluate the question or judge it, and this comes through the difficulty coefficient and the discrimination. (Jaafar Abdel Kazem Al-Mayahi, 2011, p. 171) .

Methods and analysis

The idea is to present the most important characteristics of paragraph analysis, and how to calculate the coefficient of difficulty and distinction and their significance, and its relationship to strengthening both the stability and sincerity of the tool, through the following detailed theoretical presentation:

1-Paragraph Analysis:

Paragraph analysis is the study that depends on the logical, statistical and experiential analysis of the test units to know their characteristics, and to delete, modify, replace, add, or rearrange these paragraphs in order to reach a consistent, honest and appropriate test in terms of length and level of difficulty. (Mohammed Nasr El-Din Radwan, 2006, p. 315).

1.1. The main objectives of paragraph analysis:

- Detecting whether the paragraph is very difficult, very easy, or moderately difficult, as the statistical analysis reveals the difficulty level of the paragraph, or what is known as the index of difficulty.

Detecting whether the paragraph has the ability to distinguish between distinguished and non-distinguished individuals, as the discriminative power of the paragraph is statistically revealed, or what is known as the index discrimination, which is an indicator of the validity of the paragraph.

Show why a particular paragraph is not performing its function adequately, and how this paragraph can be modified to make it more effective.

The statistical procedures for analyzing the paragraph depend on the patterns of the paragraphs used in the test, the number of examinees, and the purpose of the analysis. Confidence increases in the paragraph when the statistical analysis is done on one hundred examinees or more, and decreases when the number is less than one hundred. (Muhammad Nasr al-Din Radwan, 2006, pp. 315-316)

2.1. Paragraph Analysis Steps: Paragraph analysis is done according to the following steps:

First Step: Arrange the answers in the group score card from highest to lowest.

Second Step: The answers in the group scorecard are categorized into three subgroups as follows:

*The highest group: represents about 27% of the number of answers.

* The lowest group: represents about 27% of the number of answers.

The middle group: represents about 27% of the number of answers.

The highest and lowest groups are used as antagonistic groups that reflect the maximum differences in performance on the test.

The size of each of the two groups (the highest and the lowest) depends on the appropriate number for the statistical analysis process, as well as on the total number of answers. Measurement specialists have agreed that this percentage ranges from 25% to 33% for each group.

The third step: Count the number of those who answered correctly on each item of the test for the upper group.

Fourth step: Calculating the number of those who answered correctly on each of the test items for the lower group

1. 3. Study the difficulty of the paragraph: (Item Difficulty)

The difficulty of the paragraph means how difficult or easy it is for the respondents, and we say the difficulty of the paragraph if it measures a certain ability, but if the paragraph is a characteristic of personality, values or trends, we study the extent of its popularity.

The study of the difficulty level of the paragraph achieves three basic objectives: the first is the appropriateness of the paragraph, the second is to place the paragraph in the appropriate place for it among the test units, and the third is to ensure that the difference in the level of difficulty between the unit before it and, the one after it is equal, so as to ensure that the test is graded in degrees of approximately equal distances. Somehow, it happens on a ruler (like a distance scale).

The level of difficulty is studied by applying the paragraphs that were initially developed for the test on the rationing sample, then we calculate the success rate in this sample for each paragraph of the test, which is equal to the number of successful people dividing the number of sample members x 100, and the higher the success rate, the more difficult the paragraph or It is easy, and the lower the percentage indicates its difficulty, and vice versa, the paragraph in which 30% passes is more difficult than the one in which 50% passes, and a good test is the one that allows the average difficulty of its paragraphs with the success of 50% of the sample members tested. (Moqaddam Abdel Hafeez, 2003, p. 137).

Easy questions do not mean questions in which we do not expect any student to fail, and difficult questions do not mean impossible that one of the students cannot solve, but rather they are questions that require an estimated mental effort, and the test should begin with easy questions up to the difficult ones in order to raise the students motivation and appetite for the test. (Ahmed Yaqoub Al-Nur, 2007, p. 209).

The method of ratios is often used to prepare standardized tests, but it is affected by the guesswork factor, especially if the paragraphs are of the type called multiple choice, as the individual's answer to this type of test may be affected by chance or guesswork, or in a way that does not indicate an increase in his information or thinking.

In the given problem, and the seriousness of this defect, increases as the number of tests or alternatives decreases, the test that includes choosing one answer from two answers, the probability of success in it by guessing is 50%, which includes choosing one answer from three answers or alternatives, in which the probability of success is 30% In order to avoid this defect, Guilford (1973).

proposed a law to correct the percentage of correct answers to the paragraph from the effect of the guesswork .

Corrected percentage of correct answers=

$$\frac{\text{Number of alternatives or optional answers for each paragraph} \times (\text{The combined ratio}) - 1}{\text{Number of alternatives or optional answers for each paragraph} - 1}$$

$$\text{Number of alternatives or optional answers for each paragraph} - 1$$

1. 1.3. Difficulty index calculation: (index of difficulty)

The difficulty index is calculated for each of the test items using the results of steps (3,4), and applying the following equation:

There is another way to calculate the paragraph difficulty index without dividing the subjects into two groups (highest and lowest), where:

The relationship between the coefficient of ease and difficulty is a direct inverse relationship, as:

$$\text{Ease coefficient} + \text{Difficulty coefficient} = 1$$

(Number of alternatives or optional answers for each paragraph x The number of correct answers in the lowest group

The number of subjects in the higher group + the number of subjects in the lower group

×100

The number of people who answered correctly in the paragraph

Difficulty index = $\frac{\text{The number of people who answered correctly in the paragraph}}{\text{The total number of subjects to whom a test was applied}} \times 100$

The total number of subjects to whom a test was applied

The difficulty of the test vocabulary is one of the characteristics that play an important role in the norm-referenced tests, and affect individuals' answers about their vocabulary. It is not useful in revealing the differences between them in what the test measures, and psychometric and empirical studies have shown that the test can distinguish to the maximum extent possible between the tested individuals.

if the average level of difficulty of the vocabulary it includes is 0.50 Approximately, that is, he can answer 50% of them for each of his vocabulary, but the optimal distribution of vocabulary difficulty levels in this type of tests also depends on the relationships that exist between the test vocabulary. The average difficulty is 0.50, but if the correlation between each item and the other decreases, we expect that the difficulty of the vocabulary will be less extensive, so it is preferable to select the test items so that the extent of their difficulty Moderate, but the average difficulty of all vocabulary is approximately 0.50.

It should be emphasized here that the characteristic of the "individual difficulty" in the framework of the classical theory of tests is not limited to the item, because it also reflects the ability of the members of the reference group who answered that item. (Salah al-Din Mahmoud Allam, 2000, p. 268, 269)

1. 2.3. Correction of the value of the difficulty factor from the effect of guessing (chance):

When using the law of finding the difficulty coefficient of paragraphs (vocabularies) from multiple, or right or wrong whose answer is either right or wrong, the effect of geussing should be taken into account.

It is known that some individuals arrive at the correct answer by guessing alone, i.e. by pure chance. The probability of correct guessing for the answer to the vocabulary of multiple choice that includes five alternatives is 0.02, and four alternatives are 0.25, while the probability of correct guessing if the item is of the type of right or wrong.

0.50, since it contains only two alternatives which affects the value of the difficulty factor of the individual.

Therefore, the following question arises:

Can the total score obtained by an individual in the test be considered equal to the number of his correct answers on the test vocabulary? Or should it be corrected from the effect of random (chance) guessing to exclude that part of the total score that the individual is supposed to have obtained by chance, that is, without actually knowing the correct answer for some vocabulary?

The truth is that there is no agreement among specialists in measurement about the answer to this question, as some believe that it is not necessary to make this correction, and their argument in that is the following:

- 1- The ranking of individuals among them does not change if we correct their total scores in the test from the effect of guesswork.
- 2- Correcting the scores of individuals who do not resort to guessing leads to a lowering of their overall scores in the test, while this does not happen to those who are accustomed to guessing in their answers.
- 3- The correction of scores from the effect of guesswork depends on the assumption that the correct answer is either known or unknown, while some individuals resort to guessing based on partial information they have.
- 4- Correcting the scores from the effect of guessing, which may lead to the wrong assumption that the resulting scores are the real scores of the individuals in the test.

Therefore, the owners of this approach suggest some guidelines that can contribute to reducing the impact of estimation significantly, as follows:

- 1- Warn the testers not to guess randomly in their answers about the test items.
- 2- Allow enough time for each individual to be able to attempt to answer each of the test items.
- 3- Inclusion of the alternatives included in the vocabulary of the multiple test with partially correct information or common errors related to the domain measured by the vocabulary.

As for others, it is necessary to correct the value of the test vocabulary difficulty from the random effect of chance, in order to estimate the percentage of the number of individuals who answer correctly for each item based on their actual information and to exclude those who answer correctly by chance.

This estimate is based on the assumption that the individual's choice of an incorrect answer for an optional item is due to his lack of information required by the correct

answer, and not to the error of his information or misunderstanding, as well as the assumption that the individual who does not know the correct answer sees that all the alternatives included in the item can be represents the correct answer.

If we accept these two assumptions, we expect a certain percentage of individuals who do not know the correct answer to reach this answer by pure chance, and these percentages are the average percentages of the number of individuals who choose one of the other alternatives, i.e. the wrong answers.

It is possible to obtain an estimate of the percentage of the number of people who answer correctly for the item based on their actual information by subtracting this percentage from the total percentage of the number of individuals who answer the item correctly.

To correct an individual's degree of chance effect, we use the following formula:

Correction of the individual degree of the effect of chance = $\frac{\text{Number of correct answers} - \text{The number of wrong answers}}{\text{The number of alternatives to the multiplex test, or the true or false} - 1}$

The number of alternatives to the multiplex test, or the true or false - 1

However, it should be taken into account that in timed tests, that is, that depends on the speed in answering, some individuals may leave some vocabulary answers due to lack of time. , in order to distinguish between them, it can be assumed that all the vocabulary, including the last one that the individual answered, has tried to answer it based on his information, and as for the rest of the vocabulary that he left unanswered, we assume that he did not try to answer them. (Salah al-Din Mahmoud Allam, 2000, p. 273).

1. 3.3. Difficulty Index Meanings:

From the above it is clear that the paragraph difficulty index refers to the percentage of individuals or examinees who answered correct answers to the paragraph. This percentage is known as the "difficulty index", and this percentage can be evaluated as follows:

- 1- If this percentage (difficulty index) is 90%, this indicates that the paragraph is very easy, and then this paragraph becomes incapable of distinguishing.
- 2- If this percentage is 10%, this indicates that the paragraph is very difficult.
- 3- If this percentage is 100% or 0%, then this means that the paragraph is not valid because it cannot in any way distinguish between the two examinees.

4- The paragraph difficulty index is evaluated according to the levels defined by Ebel (1979) as follows:

Table No. 01: Evaluation of the paragraph difficulty index according to the levels determined by Ebel (1979).

| evaluation | difficulty indicator |
|--|----------------------|
| exclude the paragraph | % 80 |
| the paragraph can be accepted if it has an acceptable discrimination indicator | 71%-79 |
| the paragraph is good | 30%-70% |
| the paragraph can be accepted if it has an acceptable discrimination indicator | 20%-29% |
| exclude the paragraph | % 20 Under |

Paragraphs that have a difficulty index from 40% to 60% are classified as moderately difficult, and it is preferable to place items that have a difficulty index of more than 60% at the beginning of the test, and items that have a difficulty index of less than 40% at the end of the test, in order to ensure that the distribution of the items Characteristics of the normal distribution with respect to the difficulty level of the paragraphs. (Muhammad Nasr al-Din Radwan, 2006, pp. 320-321)

Through the following table, we judge the best coefficient of ease and difficulty for each type of question:

Table No. (02): The best coefficient for ease and difficulty for each type of

Through the following table, we judge the best coefficient of ease and difficulty for each type of question:

Table No. (02): The best coefficient of ease and difficulty for each

| Type of questions | Difficulty coefficient |
|---|------------------------|
| article | 0.5 |
| Correct/incorrect | 0.75 |
| Multiple choice with three alternatives | 0.67 |
| Multiple choice with four alternatives | 0.63 |
| Multiple choice with five alternatives | 0.6 |
| A general standard for all alternatives | 0.6-0.4 |

From the table it is clear that:

The paragraph should not be too easy so that all the examinees succeed in it, and it should not be so difficult that all the examinees fail in it, because the extreme paragraphs in their difficulty or ease levels, do not provide us with sufficient

information about the level of achievement of the examinees, and the measurement specialists have agreed that the difficulty level of the paragraph ranges from 30% to 70%.

The paragraph achieves the maximum ability to distinguish between the two subjects if its level of difficulty allows the success of 50% of the examinees who answered it.

To estimate the difficulty indicator of the paragraph, any of the two equations mentioned above can be used, as it is noted that the first equation is easier to apply because it deals with about 54% of the number of examinees, although it is faulted as it assumes that the middle group, which represents 46% of the number of examinees, is included in the distribution framework. This equation is much better in application in cases of large numbers, while the second equation is more appropriate in cases of small groups. (Muhammad Nasr al-Din Radwan, 2006, pp. 321-322)

It does not mean that the degree of difficulty applies to all paragraphs of the question (types) whether they are essay or objective questions. In essay tests, the level of difficulty is appropriate when it has the ability to obtain structured and broad information, while in objective test questions there is a possibility to specify The difficulty is more accurate compared to the paragraphs of the article question. (Al-Mayahi, 2011, p. 176)

4.1. Indication of the paragraph's ability to distinguish (paragraphs validity):

Another important characteristic that should be available in the vocabulary of the tests is the group's reference or the criterion, which means the extent to which individual differences can be measured using the vocabulary of these tests. Although there are multiple criteria that can be used in this regard, such as the opinions of teachers and the assessments of specialists in different academic fields, the total scores obtained by students in the same test are usually used as an appropriate test, whether in standardized achievement tests or in class tests written by the teacher himself. For his students, that is, the criterion used in these tests is an internal criterion. However, in constructing tests that measure inclinations, mental health, and some tests of mental preparation and skill aspects, it is preferable that the criterion be external. (Salah al-Din Mahmoud Allam, 2000, p. 277)

The discrimination indicator (evidence) refers to the degree to which the item (item) distinguishes between the levels of the examinees, meaning the ability of the item to distinguish between individuals who receive high scores, and individuals who receive low scores in the trait measured by all the items of the scale or test when A large number of examinees in the higher group get correct answers on the item, we say that the discriminatory ability of the item is positive, and if the test item has the ability to distinguish well, this means that the examinees who achieve high scores on the test will answer this item a correct answer, Contrasted with the examinees who have low scores

on the test, and the paragraph that is unable to distinguish is considered poorly constructed.

Negative discrimination indicates that most subjects in the lower group get correct answers to a paragraph.

The index (item validity) is known as the index (item validity) index, which is one of the important measures for item analysis, as it enables us to classify all test items into three types of items:

1- Paragraphs have a positive discrimination index.

2- Paragraphs have a negative discrimination indicator.

3- Paragraphs that have no discriminatory ability. (Muhammad Nasr al-Din Radwan, 2006, pp. 327-319)

1.4.1. Calculating the ability of a paragraph to discriminate: (ID): (Index of discrimination):

1- Indication of the difference between two percentages: The ability of the paragraph can be measured to find out the difference between the percentage of correct answers in the upper third of the upper group, and the lower third of it. An external criterion, such as work performance grades or academic grades, or the criterion may be the total score obtained by the examinees in the test. (Moqaddam Abdel Hafeez, 2003, p. 143)

Any of the following percentages can be used to determine the size of each group.

- 33% as terminal values for each group - Cureton 1957.
- 27% as terminal values for each group - Kelly 1939.
- 25% as terminal values for each group - Ibel 1979.
- 9% as terminal values for each group - Ibel 1979.

2- Calculating the number of examinees who answered correct answers to the target paragraph for the distinguished group (N1).

3- Likewise, the number of those who answered correctly on the same paragraph is calculated for the group that got the lowest score on the test (N2).

4- The number of correct responses for the weak group is subtracted from the number of correct responses for the distinguished group, then the result is divided by the total number of the distinguished group (N1) using the Marshall and Hallis equation

The number of correct answers for the privileged group - the number of correct answers for the weak group

| | | |
|----------------|-------|---|
| Discrimination | index | = |
|----------------|-------|---|

The number of subjects in one of the two groups

* Index of discrimination (ID):

1- The discrimination index ranges from (-1.00 to +1.00) but rarely reaches either of the two values.

2- When the discrimination index is negative, this means that the number of examinees who answered correctly the paragraph in the lowest group (N2) is more than those who answered the same paragraph in the higher group (N1), and then deletes such kind of paragraphs completely.

3- For tests of the standard reference (NR), the paragraph discrimination index is accepted when it is equal to (0.4) or more.

Table No. (03): Evaluation of the Paragraph Discrimination Index according to the levels determined by Ebel (1979).

| evaluation | discrimination indicator |
|---|--------------------------|
| good discrimination ability | 0.40 |
| it has a suitable discrimination ability, and may need adjustment especially with regard to paragraph | 0.30-0.39 |
| need revision | 0.20-0.29 |
| excludes the paragraph | 0.20 |

Important note: The previous levels of discrimination indicators are not applied to the benchmark tests (CR. (Mohammed Nasr El-Din Radwan, 2006, pp. 329-330).

Therefore, the higher the value of the individual discrimination coefficient, the better, but this depends on the purpose of using the test as we mentioned. In the group reference tests (the criterion) that aim to detect individual differences, their vocabulary should distinguish a high distinction between the levels of the trait that the test measures, but in Diagnostic tests referential criterion We should not pay much attention to the vocabulary discrimination coefficient, as the attention here is focused on the quality of the vocabulary in terms of its formulation, and its measurement of the components of certain skills or tasks so that using them can accurately identify

weaknesses, and the values of the vocabulary discrimination coefficient depend on the level of difficulty of the item, the more The value of a specific individual difficulty coefficient moved away from 0.50 by increasing or decreasing the maximum value for the discrimination coefficient, so we should not expect high values for the discrimination coefficient if the test items are very easy or difficult. (Salah al-Din Mahmoud Allam, 2000, p. 289)

Ebel (Ebel.1965) and Kryspin and Feldhusen (1979) illustrate the relationship between the values of the difficulty coefficient and the discrimination coefficient of test vocabulary in the following schematic form:

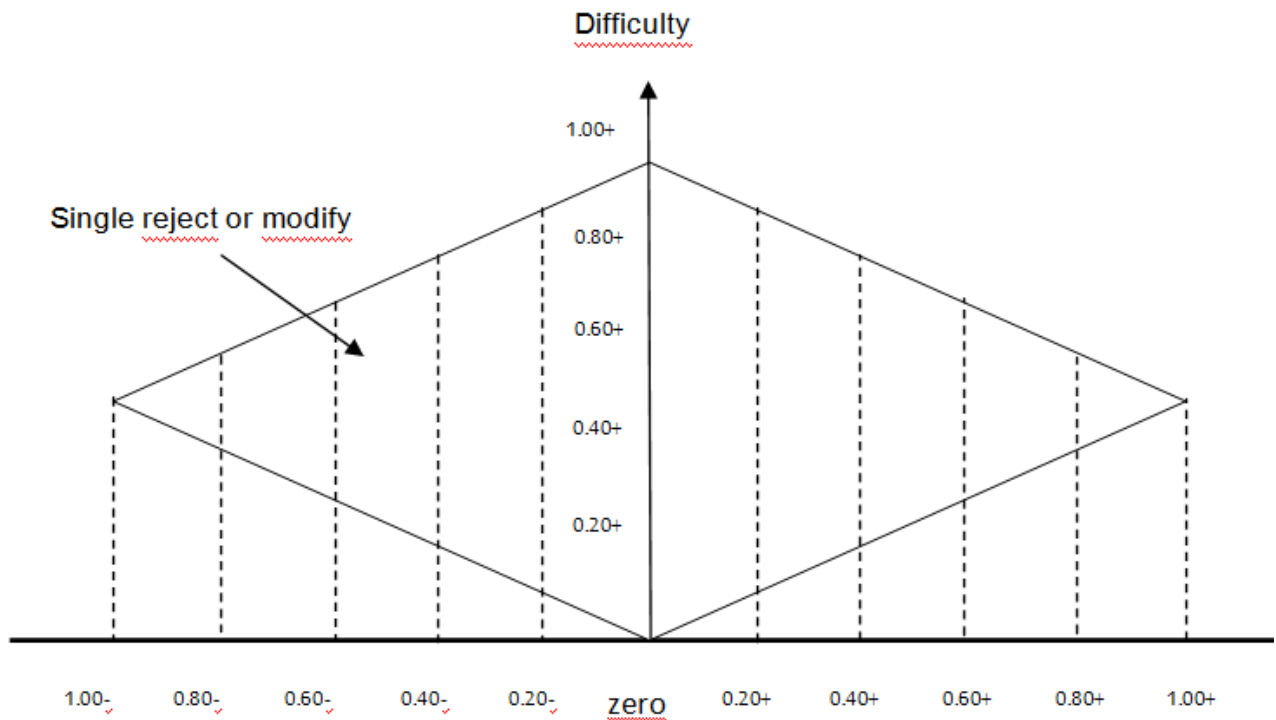


Figure No. (01): The relationship between the difficulty coefficient and the discrimination coefficient for test items.

It is clear from Figure (01) that there is a relationship between the values of the vocabulary difficulty coefficient and its discrimination coefficient. And the vocabulary whose difficulty coefficient ranges between 0.40 and 0.60 have high discrimination as shown by the two horizontal lines. The item discrimination coefficient is negative, that is, it fell in the rejection region to the left of the vertical axis, then this item can be excluded or reviewed and checked. (Salah al-Din Mahmoud Allam, 2000, p. 290)

Fakhri Khader (2003), citing Ahmed Al-Nour, indicates that distinguishing the test requires diversifying the levels of questions in terms of ease and difficulty, working to increase questions, and covering the content of the study material. They don't deserve it. (Ahmed Yaqoub Al-Nur, 2007, p. 207)

The degree of discrimination of the item is usually determined by comparing the performance of different groups on the item, with the performance of those groups on a specific test that is supposed to measure the trait under study to a large extent. College on the test and low correlation with other test items separately. (Zakaria Muhammad al-Zahir and others, 2002, p. 130)

5.1. Importance of test vocabulary analysis:

Those in charge of constructing tests and measurements are interested in writing or selecting high-quality vocabulary to accurately measure human traits, so they take into account many conditions in the formation and formulation of these vocabulary and verification by logical methods and expert judgments of the validity of the content of each individual item, as well as the validity of the content of the test as a whole. The accuracy of these methods does not dispense with the field experimentation of the test and the analysis of its vocabulary scores using statistical methods, and determining the relationship between what the vocabulary measures and the individuals' responses to it, in order to identify ambiguous or confusing vocabulary, or that encourages guessing, and very easy or difficult vocabulary. It does not reveal individual differences in the trait measured by the group reference test or standard, and thus it does not distinguish between the levels of this trait, nor does it contribute in any way to the validity of the test or the stability of its scores.

This is useful in the technical review and improvement of vocabulary so that each vocabulary contributes positively to what the test measures. Vocabulary analysis also helps in developing the skills of those who prepare tests and measures as a result of an informed examination of vocabulary, identifying weaknesses that may make some of these vocabulary invalid, and retaining vocabulary that meets the characteristics of The purpose of using the test or scale in the different areas of assessment, and reusing it if necessary. (Salah al-Din Mahmoud Allam, 2000, p. 267)

The selection of the test items in the referenced test depends mainly on the degree to which they measure the intended learning outcome. Items that have very high levels of difficulty or ease can be taken into consideration if they measure learning outcomes accurately, and this does not mean that the discriminatory ability is not important, but the measurement of the outcome Education is the most important.

The relationship of the difficulty index to the stability of the test:

Measurement scientists agree that the difficulty levels of the paragraphs affect the stability of the test, and that the optimal levels of difficulty that give the highest stability of the test, which depend on a number of factors, the most important of which are the number of test items and the type of items. Specialists recommend the following:

The stability of the test increases when the number of paragraphs is 30 or less, and tends to decrease whenever the number of paragraphs exceeds 30.

- In true-false or (agree-disagree) paragraphs, the paragraph difficulty index must be above 70%.

- And that this indicator for the test items between (2 of the alternatives) should be 85%, and between (3 of the alternatives) 77%, and between (4 of the alternatives) 74%, and in between (5 of the alternatives) 69%.

The stability of the paragraph (item) degree can be calculated by applying the following equation:

$$\text{Paragraph (item) stability} = \frac{\text{Paragraph frequency} \times \text{Number of choices "answer possibilities" coefficient}}{\text{frequency} - (1 \div \text{number of choices}) \times \text{highest relative}} \times \text{Number of choices} - 1$$

On the other hand, the difficulty index can be used to calculate the standard deviation (p-SD) for the paragraph using the following equation:

$$\text{standard deviation} = \sqrt{\text{Paragraph Difficulty Index} \times (100 - \text{Paragraph Difficulty Index})}$$

It is also possible to calculate the variance of the paragraph where (the variance - p²), and it is noted that the highest value of the variance of the paragraph is 0.25, which is the value at which the stability index is equal to about 50%. (Mohammed Nasr El-Din Radwan, 2006, pp. 323-326)

2. Conclusion:

Through this study, the most important methods of analyzing paragraphs were identified that depend on the logical, statistical and empirical analysis of the test units to know their characteristics.

The methods of calculating the difficulty and discrimination coefficients and their most important roles were identified, especially in the operations of deleting or modifying the items of paragraphs, or rearranging these paragraphs so that Access to a consistent, honest and appropriate test in terms of length and level of difficulty, so that we can build a psychological or cognitive test that truly measures the various aspects that make up the trait to be measured, especially that the processes of knowing the ease, difficulty, and discrimination of the paragraph have an effective and direct impact on

obtaining honest and then consistent content Because the paragraphs have been modified in order to avoid ambiguity and interpretation, and thus obtain an objective test.

***References**

1-Ahmed Nasreddin Radwan. (2006). Introduction to measurement in physical education and sports. Cairo: Al-Kitab Center for Publishing.

2-Ahmed Yacoub Al-Nour (2007). Measurement and evaluation in education and psychology. Jordan: Dar Al-Janadriyah for publication and distribution.

3- Jaafar Abd Kazem Al-Mayahi (2011). Psychological measurement and educational evaluation. Jordan: Treasures International Knowledge House for publication and distribution.

4- Zakaria Muhammad Al-Zaher, Jacqueline Tamerjian, Jawdat Ezzat Abdel-Hadi. (2002).

Principles of Measurement and Evaluation in Education, Jordan: Dar Al-Ilmiya and Dar Al-Thaqafa for Publishing and Distribution.

5- Salma Zaki Al-Nashif (2011). Your guide to test design. Amman: Dar Al-Bashir. Salah El-Din Mahmoud Allam (2000).

Educational and psychological measurement and evaluation (its basics, applications and contemporary trends). Cairo: Dar Al-Fikr Al-Arabi.

6 - Mustafa Hussein Bahi, Ahmed Kamal Nasari, Amin Abdul Ghani. (2013). Introduction to Tests and Measurements in the Sports Field: Cairo. Anglo-Egyptian Library.

7 - Mustafa Hussein Bahi, Ahmed Kamal Nasari, Mukhtar Amin Abdul Ghani. (2013). Introduction to Tests and Measurements in the Mathematical Field. Cairo: The Anglo-Egyptian Library.

9- Moghaddam Abdel Hafeez (2003). Statistics, psychological and educational measurement, with models of scales and tests. Algeria: University Press Office.