Ed Process Int J

www.edupij.com

ISSN 2147-0901 (Print), 2564-8020 (Online)

EDUCATIONAL PROCESS
INTERNATIONAL JOURNAL

Research Article

Cite this article: Yassin, O.S.B., Freihat, A.M., & Al-Tarawneh, S.H. (2025). Differences of Half-Split Equations on Estimating Test-Reliability Coefficient. *Educational Process: International Journal, 14*, e2025015. https://doi.org/10.22521/edupij.2025.14.15

Received December 01, 2024
Accepted December 27, 2024
Published Online January 02, 2025

Keywords:
half-split, reliability, tests

Author(s) for correspondence:
Omar Saleh Bani Yassin[1]
Aiman Mohammad Freihat[2]
omarsa@bau.edu.jo[1]
aiman.freihat@bau.edu.jo[2]
Educational Sciences Department, Irbid University College, Al-Balqa Applied University

Differences of Half-Split Equations on Estimating Test-Reliability Coefficient

Omar Saleh Bani Yassin, Aiman Mohammad Freihat, Sabri Hassan Al-Tarawneh

Abstract

Background/purpose. This study aimed to investigate the differences among the equations used in estimating the reliability coefficient using the half-split method. These equations demonstrate Spearman-Brown's, Rulon's, Guttman's, Mosier's, Flanagan's, and Horst's.

Materials/methods. The study instrument was a 43-item scale for evaluating the computerized mathematics curriculum for the tenth grade in southern Jordan. It was applied to a sample of 303 male and female teachers and educational supervisors.

Results. The results showed that all values of the reliability coefficients estimated in the six equations were acceptable. In addition, the best equation for estimating the half-split reliability coefficient was the Spearman-Brown equation, followed by two equations by Flanagan and Rulon.

Conclusion. Considering the results of the current study, the researchers do not recommend using Mosier's equation because it gave the lowest reliability-coefficient value.

1. Introduction

Tests are one of the most common measurement and evaluation instruments used to measure examinees' achievement, and they must demonstrate significant validity, reliability, and objectivity. Each researcher must verify its general psychometric properties and reliability when adapting, developing, or constructing a scale or test. It becomes vital if the results of these measures and tests are used to make educational decisions related to the student's future, classification, and guidance. Therefore, measurement instruments have received the attention of researchers who have conducted many studies in this field (Saeed, 2015, 2019, 2023).

Reliability is an important test characteristic. The test is reliable when it gives similar scores or results to the same respondent if reapplied more than once. The test is highly reliable when it gives an actual description of the measured trait. Reliability (test and retest) is assessed using several methods, such as equivalence reliability and equivalence-stability reliability. Besides, some methods require the application of the test once to assess internal reliability. They include Cronbach's Alpha, Kuder-Richardson equations (20 and 21), and the half-split method (Aiken, 2003; Kim & Feldt, 2008).

In the Half-Split method, the test is divided into two parts, and each respondent has scores for each test section. The Pearson correlation coefficient is calculated between the scores of the two sections. The value of the correlation coefficient is equal to the reliability coefficient for half of the test. Equations were applied to calculate the reliability coefficient for the test as a whole (Al-Ghareeb, 1998; Al-Majeed, 2010; Melhem, 2002).

The test is divided into two halves in several ways: odd items from the first part and even items from the second part. The upper half of the items are from the first part, and the lower half of the items are from the second part. The division is made according to the values of the difficulty coefficients of the items, or the items are placed randomly in the two-section test (Al-Nabhan, 2004). Some disadvantages arise from dividing the test into two parts in an upper and lower manner. Among these disadvantages, students in the upper part may be excited, active, and motivated to answer the upper items (the beginning of the test). However, their enthusiasm and motivation may decrease. They may feel tired and bored when answering the lower items (the end of the test), which affects their performance in both halves (Allam, 2010). Many researchers divide the test into two parts according to item numbers, placing odd and even forms in the first and second parts of the test, respectively. This is because the division problems are eliminated using the lower and upper parts. Thus, the examinees' activity, fatigue, and boredom spread over both halves of the test. There will be odd and even items answered by the examinees while they are equally active and highly motivated. In addition, there will be odd and even items answered when the examinees are tired and bored (Al-Turairi, 1997; B. Ismail, 2004; H. Ismail, 2014).

The correlation coefficient results from the half-split method are the correlation coefficients between the two halves of the test but not for the entire test. Therefore, correcting the correlation coefficient between the two halves is necessary until the reliability coefficient is obtained for the entire test, correcting for attenuation. Metrologists have developed several equations to calculate the reliability coefficient:

1) Spearman-Brown Equation: This equation assumes that increasing the number of test items increases the reliability coefficient. One advantage of this equation is that it can be used in any half-split. Some critics have pointed out that it does not work with timed tests (speed tests). The reliability coefficient is calculated using Equation 1.

$r_{xx} = \frac{2r_{12}}{1+r_{12}}$ Where:

r_{xx}: coefficient of reliability of the total test

r_{12}: correlation coefficient between the two halves of the test (Abu-Saree', 2004; Crocker & Algina, 1986).

2) Rulon Equation: This equation is considered a shortened method and does not require calculating the correlation coefficient between the two parts of the test. It estimates the reliability based on the ratio of the real variance in performance. In addition, this equation was calculated without correcting the length. Thus, the reliability coefficient is calculated using Equation 2.

$r_{xx} = 1 - \frac{\delta^2 d}{\delta^2 x}$ Where:

δ^2 d: Variance in the differences between the scores of the two halves of the test.

$\delta^2{}_x$: Total variance of test (Stanley & Hopkins, 1998).

3) Guttman Equation: This equation is based on the same logic as the Rulon Equation. The difference between them is only in the more intuitive calculations, and they do not require an equation to correct the length. Thus, the reliability coefficient is calculated using Equation 3.

$r_{xx} = 2\left(1 - \frac{\delta 1^2 + \delta 2^2}{\delta^2 x}\right)$. Where:

$\delta 1^2$: Variance in scores of the first half of the test.

$\delta 2^2$: Variance in the scores in the second half of the test

$\delta 2_x$: Variance in the entire test (Al-Nabhan, 2004).

4) Horst Equation: This equation is considered a correction equation similar to the Spearman-Brown equation. It is used when dividing the test into two sections with an unequal number of items (Al-Tarawneh, 2022). The reliability coefficient is calculated using Equation 4.

$r_{xx} = \frac{r_{12}\sqrt{\ r_{12}{}^2 + 4AB(1 - r_{12}{}^2) - r_{12}}}{2AB(1 - r_{12}{}^{2)}}$ Where:

r_{12}: Correlation coefficient between the two parts of the test.

A: Ratio of the items number in the largest part of the test.

B: Ratio of the items number in the lowest part of the test (Al-Turairi, 1997).

5) The Glackson Equation: Proposed by Glackson to calculate the reliability coefficient for tests that were affected by time and speed. The reliability coefficient is calculated using Equation 5.

$r_{xx} = r_{11} - \frac{\mu}{\delta^2 e}$. Where:

r_{11}: Reliability coefficient calculated using the Spearman-Brown method.

μ: Average of items left over at the end of the test.

$\delta^2{}_e$: The least variance in the number of items in the test part (Al-Majeed, 2013).

6) Mozier's Equation: The researchers used the standard deviation of the examinees' grades, the grades' variance, and the correlation coefficient between one of the test halves and the total test score. The reliability coefficient is calculated using Equation 6.

$r_{xx} = \frac{r_{ox} \times \delta x - \delta o}{\delta^2 x + \delta^2 o - 2 r_{ot} t \delta o \delta x}$. Where:

r_{ox}: Correlation coefficient between respondent item scores and total test scores.

δx: Standard deviation of the total test item scores.

δo: Standard deviation of the degrees of odd items.

$δ^2$ x: Variance for total test item scores.

$δ^2$ o: Variance in the degrees of odd items (Al-Turairi, 1997).

7) Flanagan Equation: This equation is based on both the standard deviation and the variance of the two test parts and the correlation coefficient between the two test parts. The reliability coefficient is calculated by Equation 7:

$r_{xx} = \frac{4\delta 1 \times \delta 2 \times r_{12}}{\delta^2 1 + \delta^2 2 + 2\delta 1 \delta 2 \times r_{12}}$. Where:

δ1: The standard deviation of the first part of the test

δ2: standard deviation of the second part of the test

r_{12}: correlation coefficient between the two parts of the test

$δ1^2$: Variance for the first part of the test

$δ2^2$: Variance for the second part of the test (Al-Majeed, 2013).

Several studies have been conducted to verify the reliability of these tests. Zimmerman et al. (1984) conducted a study to compare the reliability of matching type, multiple-choice, and complement tests with the participation of 73 male and female students from Carleton University. The results indicated that the reliability coefficient of the matching test type was greater than that of the multiple-choice and complement tests. Many studies have been conducted on the relationship between the number of alternatives for multiple-choice tests and their reliability. Trevisan et al. (1991) conducted a study to show the impact of the number of alternatives in multiple-choice tests on the reliability and actuality of the test. They used a test with three images that differed in the number of alternatives. The study sample comprised 435 male and female American high school students. The results indicated statistically significant differences among the reliability coefficients attributed to the number of alternatives for the test items. The differences favored the test, which included three multiple-choice options.

Al-Zahrani (2000) conducted a study to compare eight methods for estimating the reliability of criterion-referenced tests. A test was developed for mathematical competencies in mathematical calculations. The study sample comprised 659 male and female basic sixth-grade students in the Jeddah Governorate. The results indicated that the number of test items should not be less than 20 to obtain acceptable reliability coefficient values. The results also indicated that the beta coefficient differed from the rest of the coefficients in its impact on the change in the variance value and number of test items.

Walker (2006) conducted a study to compare the Spearman-Brown, Flanagan, and Rulon equations in the half-split to estimate reliability under different conditions of the coefficient of variance. The differences among Spearman-Brown's, Flanagan's, and Rulon's equations were examined. The variance coefficients of the two halves of the test were of different ratios, and there were different correlation coefficients between the two halves of the test. The results indicated using the Spearman-Brown equation to estimate reliability is unguaranteed when the variance ratios of the two halves of the test are between 0.9 and 1.1.

Thompson et al. (2010) conducted a study to evaluate the largest reliability value estimated by the half-split method. They calculated the reliability coefficient value using all half-split methods. The maximum value was then chosen to represent the reliability coefficient value. The maximum reliability value was compared to 10 values to estimate reliability using the internal reliability method. The results indicated that the maximum reliability values calculated were the most accurate reliability coefficient values.

Onn (2013) conducted a study to compare classical theory and the item-response theory in terms of the number of items and reliability using a physics test. The study sample comprised 69 Nigerian male and female students. The results indicated that the reliability-coefficient values for both theories were low. The reliability coefficient value of the item-response theory was higher than that of the classical theory.

Al-Qatawna (2015) conducted a study to compare the reliability results by criterion-referenced tests according to classical and item-response theories. Al-Qatawna used a criterion-referenced mathematics test. The study sample comprised 531 male and female students in the tenth grade of Kerak Governorate. The results indicated no statistically significant differences among the reliability coefficients estimated according to classical theory. However, statistically significant differences among the reliability coefficients were estimated according to the item-response theory. The differences were in favor of the reliability coefficient for respondents.

Zare' (2021) conducted a study to compare the reliability values of thirteen reliability coefficients. They exemplified the Guttman minimum reliability coefficients (λ1, λ2, λ3, λ4, λ5, λ6); Cronbach's coefficient (α); the total omega coefficient (ωt); the hierarchical convergent omega coefficient (ωt); the largest minimum coefficient (glb); the rank alpha coefficient for multiple partitions (αpoly); the worst half-partition coefficient beta (β); the stratified Alpha coefficient (αstrata); and the maximum reliability coefficient (Maximal Reliability). Zare' used data generated by the Monte Carlo method. The results showed four reliability coefficients, which gave the highest reliability estimate coefficient. These reliability coefficients were considered unbiased and outperformed the traditional Cronbach's Alpha coefficient. These reliability coefficients exemplified the total omega coefficient (ωt), the maximum limit coefficient (glb), the minimum Guttman coefficient (λ4), and the alpha coefficient for multiple sections (αpoly). The total omega coefficient (ωt) was the best, which gave the highest reliability value.

Viewing the previous studies, it is clear that some of them investigated the impact of the type of test items on reliability (Zimmerman et al., 1984). Some studies investigated the impact of the number of item alternatives on the reliability of the test (Trevisan et al., 1991). Other studies have investigated the comparison between classical theory and item-response theory to estimate the psychometric characteristics of the test (Onn, 2013). A few studies have compared methods for estimating the reliability of the criterion-referenced test (Al-Zahrani, 2000). Several studies have compared reliability values using the Spearman-Brown, Flanagan, and Rulon equations (Walker, 2006). Some studies have evaluated the greatest value for estimating reliability using the half-split method (Thompson et al., 2010). However, no study compared six equations of half-split to estimate reliability. Thus, this study aimed to complement previous studies that revealed the impact of the equation used in the half-split method in estimating the reliability coefficient test.

2. Study Problem

Tests must be reliable because important decisions are based on their scores. Metrologists and evaluators have adopted many methods and equations to calculate the reliability-coefficient values of tests. One of these methods is the half-split method. Many equations are used in the half-split method through which the reliability–coefficient value of the test is estimated. Therefore, it is significant to determine the best equation that gives the best reliability-coefficient value. This study particularly addressed the following research questions:

RQ 1: What are the reliability-coefficient values of the tests using the half-split method expressed by Spearman-Brown's, Rulon's, Guttmann's, Mosier's, Flanagan's, and Horst's equations?

RQ2: Are there any statistically significant differences in the values of the reliability coefficients of the tests using the half-split method depending on the equation calculated?

3. Study Significance

3. 1. Theoretical Significance

This significance stems from highlighting the half-split method for estimating the reliability coefficient using gender equations and determining the best and simplest equation for this purpose. It also highlights the consequences of their results for educational decisions.

3. 2. Practical Importance

This significance stems from developing reliable tests to measure the true level of respondents. It helps test preparers improve these equations to verify the reliability of the tests when building them. Besides, it helps teachers estimate the reliability coefficients of their tests by using the best and simplest of these equations, as it depends on applying tests once.

4. Limitations of the Study

The results of the current study must be interpreted regarding the following limitations, which might have influenced these results:

1. Using the half-split method to estimate the reliability
2. Using the Spearman-Brown equation, the Rulon equation, the Guttmann equation, the Mosier equation, the Flanagan equation, and the Horst equation.
3. Gathering real data from a previous study by researcher Sabry Tarawneh.

6. Method

6. 1. Methodology and Procedures

The researchers used an analytical methodology to address the research questions stated earlier.

6. 1. 1. Study Instrument

The study instrument is a scale for evaluating the computerized mathematics curriculum for the basic tenth grade. It was developed by Al-Tarawneh and Al-Qadi (2016). It comprised 43 items and was verified by presenting it to 10 competent referees selected from the teaching staff at Mutah University. They are specialists in mathematics and educational technology. They are supervisors and mathematics teachers concerned with teaching mathematics to the tenth graders. The correlation coefficient was also calculated using the experimental sample's responses on each item and the responses on the scale as a whole. All correlation coefficients were significant at the level of (α=0.01).

The scale reliability was verified by applying it to an experimental sample comprising 25 teachers and educational supervisors from the study community. The reliability coefficient was calculated using the Cronbach Alpha, which was 0.90, and it was equal to 0.80 when the repetition method was applied.

6. 1. 2. Study Sample

The study sample was from a previous study by the researchers Sabri Al-Tarawneh and Al-Tarawneh, which examined the evaluation of the computerized mathematics curriculum for the basic tenth grade from the viewpoints of mathematics teachers and supervisors in the southern region. The sample comprised 303 male and female teachers and educational supervisors (Al-Tarawneh & Al-Qadi, 2016).

6. 2. Study Procedures

The researchers adopted the following procedures:

- Determining the data for the scale comprising 43 items of the five-point Likert scale type and the number of respondents (303).
- Selecting six equations for estimating the reliability of the half-split, namely: Spearman-Brown equation, Rulon equation, Guttmann equation, Mosier equation, Flanagan equation, and Horst equation. The researchers did not use the Glexon equation because its tests depend on time, and some items were left unanswered (Al-Majeed, 2013).
- Calculating unknown values (elements of each equation) in the six equations listed in Table 1 using the statistical program (SPSS).
- Calculating the reliability coefficient of the scale as a whole by using the six equations, pen, paper, and calculator.
- Finding the critical value of the test (M), which follows the chi-square distribution at degrees of freedom equal to (the number of reliability coefficients -1). It was at the significance level (α =0.05), equal to (11.07).
- Finding the critical value of the test (W), which follows the distribution of F with degrees of freedom (n1-1, n2 – 1), using the Arabic processor program in statistics (APSS), which is equal to (1.227).
- Coming up with results and writing recommendations.

Table 1. Statistics Calculated Using the SPSS Program

Number	Statistical	Symbol	The Reliability Coefficient
1	The correlation coefficient between the two parts of the test	r_{12}	Spearman-Brown, Hurst and Flanagan
2	Variance of the difference between the scores of the two halves of the test	$\delta^2 d$	Rulon
3	The overall variability of the test	$\delta^2 x$	Rulon, Guttman, Mosier
4	The Variance in the scores of the first half of the test	$\delta 1^2$	Guttman and Flanagan

Number	Statistical	Symbol	The Reliability Coefficient
5	Second-half test score variance	$\delta 2^2$	Guttman and Flanagan
6	The ratio of the number of items of the bulk of the test	A	Horst
7	The ratio of the number of items in the lower part of the test	B	Horst
8	Coefficient of correlation between respondent and total item scores	r_{ot}	Mosier
9	The standard deviation of total test item scores	δx	Mosier
10	The standard deviation of the degrees of odd items	δo	Mosier
11	Variance of the degrees of odd items	$\delta^2 o$	Mosier
12	The standard deviation of the first part of the test	$\delta 1$	Flanagan
13	The standard deviation of the second part of the test	$\delta 2$	Flanagan
14	The standard deviation of total test item scores	$\delta 2^2$	Guttman and Flanagan

7. Results and Discussion

7. 1. Results for the First Research Question

To answer RQ1, asking 'what are the values of the reliability coefficients of the test using the half-split method expressed by the following equations: Spearman-Brown, Rulon, Guttman, Mosier, Flanagan, and Horst?" the reliability coefficient of the test was calculated using the equations (1, 2, 3, 4, 5, 6, 7) as shown in table 2.

Table 2. The Reliability Coefficient by the Equation

The Equation	Reliability Coefficient
Spearman-Brown	0.849
Rulon	0.818
Guttman	0.812
Mosier	0.773
Flanagan	0.824
Horst	0.796

Table 2 shows that all the reliability coefficients were greater than 0.65, and they were acceptable. The table also shows apparent differences between the values of the reliability coefficients calculated by the six half-split equations. The lowest value of the reliability coefficients was calculated by the Mosier equation, which was equal to 0,773. The largest value was calculated by the Spearman-Brown equation, which was equal to 0,849. This may be due to the difference in the statistics used to calculate each equation to estimate the reliability coefficient.

7. 2. Results for the Second Research Question:

To answer RQ2, asking 'are there any statistically significant differences in the values of the reliability coefficients of the test using the half-split method due to the equation it calculated?' the test (M) was used as suggested by Hakstain and Whalen (1976). It follows the distribution of the chi-square test, with degrees of freedom equal to (the number of reliability coefficients-1). It is calculated using the following equation:

$$M = \frac{(j-1)(9n-11)^2}{18\,J(n-1)}\left(k - \frac{\left[\sum_{k=1}^{k}(1-r_{ak})^{\frac{-1}{3}}\right]^2}{\sum_{k=1}^{k}(1-ra_k)^{\frac{-2}{3}}}\right)$$ Where :

J: number of test items

ra_k: method reliability coefficient

n: the number of samplers used in the method.

K: the number of reliability coefficients.

The calculated value of the statistic (M) was equal to 63.35. It is greater than the critical value of the chi-squared test by 4 degrees of freedom at the significance α = 0.05 level, equal to 11.07. This indicates there are significant differences among the reliability-coefficient values. To find out in whose favor the differences were, the Feldt et al. (1987) equation was used, being calculated using the following equation:

$$W = \frac{1-\alpha 2}{1-\alpha 1}$$. Where:

$\alpha 1$: The coefficient of greatest reliability

$\alpha 2$: Smaller coefficient of reliability

The value calculated by Equation Number 9 is compared with a tabulated value of the F test with degrees of freedom (n1-1, n2-1).

(n1): the number of samplers for the first reliability coefficient.

(n2): the number of samplers for the second reliability coefficient.

Note: (n1) = (n2), where the reliability coefficients were calculated on the same sample members, and their number is (303) respondents.

Equation 9 was applied to check in favor of whom the differences would be. The critical test value (F) was found with degrees of freedom (302, 302) at the level of significance (α = 0.05) and equals (1.227), as shown in Table 3.

Table 3. Differences among The Equations Used in The Half-Split Method

The equation	The Spearman-Brown	Rulon	Guttman	Mosier	Flanagan	Horst
The Spearman-Brown	0.151	1.205	*1.245	*1.503	1.185	*1.351
Rulon	0.182	-	1.033	*1.247	1.034	1.121
Guttman	0.188		-	1.207	1.068	1.085
Mosier	0.227			-	*1.290	1.113
Flanagan	0.176				-	1.159
Horst	0.204					

**: A statistical significance at the level of significance (α = 0.05)

Table 3 shows the following results:

1. The differences in the reliability-coefficient values were between the values calculated by the Spearman-Brown equation and the values calculated by the Guttmann, Mosier, and Horst equations. The differences were in favor of the Spearman-Brown equation. This is because the calculated reliability-coefficient value, using this equation, depends on the correlation-coefficient value between the two-part test only. This equation did not include the standard deviation values of the test or any of its halves in its calculation, excluding the values of the total variance of the test or any of its halves. The rest of the equations entered the standard--deviation or total-variance values.
2. The differences in the reliability-coefficient values were between the values calculated by the Rulon equation and the values calculated by the Mosier equation. The differences were in favor of the Rulon equation. The Rulon equation depends on the variance of the difference between the scores of the two-half test and the total variance of the test. However, the Mosier equation depends on the correlation coefficient between the two-half test, the standard deviation, the variance of the odd items, and the total test's variance.
3. The differences in the reliability–coefficient values were between the values calculated by the Flanagan equation and the values calculated by the Mosier equation. The differences were in favor of the Mosier equation. The Mosier equation depends on the correlation coefficient between the two-half test, the standard deviation, the variance for the odd items' scores, and the overall variance of the test. However, the Flanagan equation depends on the correlation coefficient between the two-half test, the standard deviation, and the variance for the scores of the two parts of the test.

A comparison between the results of this study and those of the previously mentioned studies shows that the Spearman-Brown equation is the best equation for estimating the half-split reliability coefficient. However, Walker (2006) indicated that using the Spearman-Brown equation is not preferred at limited values of variance for the two halves or parts of the test. The results of the study by Thompson et al. (2010) differed from the current study's results. The former results indicated that the most accurate values of the reliability coefficient were at their maximum values, regardless of the equation used to estimate the half-split reliability coefficient. Nevertheless, the results of this study indicated that the best equations that give the highest value for the half-split reliability coefficient are, respectively, the Spearman-Brown, the Flanagan, and the Rulon. Therefore, according to the researcher's knowledge, this study confirms the scarcity of studies that have investigated the comparison among the six equations.

8. Conclusion

The researchers concluded that the reliability-coefficient values calculated by the six methods were greater than 0.65, which is acceptable. Besides, significant differences existed among the reliability-coefficient values calculated by the six equations. The researchers recommend using the Spearman-Brown equation because it gave the highest reliability-coefficient value. The Flanagan equation can be used in second place, and the Rulon equation in third place. The researchers do not recommend using the Mosier equation because it gave the lowest reliability-coefficient value.

9. Recommendations

The researchers make the following recommendations based on the current results:

- Using the Spearman-Brown equation to estimate the reliability of the half-split method.
- Conducting a study that compares the Spearman-Brown, Rulon, and Guttman equations.
- Conducting a study of the same variables to estimate the reliability-coefficient tests of different numbers of sample members and items.
- Conducting a study to estimate the reliability coefficient by the half-split method using other equations, like the Glackson equation.

Declarations

Author Contributions. All authors have read and approved the published version of the article.

Conflicts of Interest. The authors declare no conflicts of interest.

Funding. This research received no funding.

Data Availability Statement. Data can be provided by the corresponding author upon request.

References

Abu-Saree', R. A. (2004). *Data analysis using SPSS*. Amman, Dar Al Fikr.

Aiken, R. (2003). *Psychological testing and assessment*. Boston: Allyn and Bacon.

Al-Ghareeb, R. (1998). *Psychological and educational evaluation and measurement.* Anglo-Egyptian Library.

Allam, S. A. D. (2010). *Educational and psychological measurement and evaluation: Its basics, applications, and contemporary trends*. House of Arab Al-Fikr.

Al-Majeed, S. (2010). *Psychological tests (models)*. Safa's House for Publishing and Distribution.

Al-Majeed, S. (2013). *Foundations of the construction of psychological and educational tests and scales.* Debono Center for Teaching Thinking.

Al-Nabhan, M. (2004). *Fundamentals of measurement in behavioral sciences*. Dar Al Shorouk for Publishing and Distribution.

Al-Qatawna, A. (2015). *Reliability in the tests is a spoken reference in mathematics for the tenth grade according to the classical theory and the theory of item response according to the two-teacher model: A comparative study* (Unpublished Master Thesis). Mu'tah University, Karak, Jordan.

Al-Tarawneh, S., & Al-Qadi, H. (2016). Evaluation of the 10th grade computerized mathematics curriculum from the perspective of the teachers and educational supervisors in the Southern Region in Jordan. *Journal of Education and Practice, 7* (2), 39–47.

Al-Tarawneh, S. (2022). Principles of measurement and evaluation.

Al-Turairi, A. (1997). *Psychological and educational measurement: Its theory, foundations, and applications*. Riyadh: Al-Rushed Library for Publishing and Distribution.

Al-Zahrani, S. (2000). *Comparison of methods for estimating reliability in fun-telling measurement*. Makah, Saudi Arabia: Umm Al-Qura University.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. *Holt Rinehart and Winston.*

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient Alpha. *Applied Psychological Measurement*, *11*(1), 93-103. https://doi.org/10.1177/014662168701100107

Hakstain, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent Alpha coefficients. *Pyschometrika, 41*(2)19-231. https://doi.org/10.1007/BF02291840

Ismail, B. (2004). *Reference in psychological measurement.* Anglo-Egyptian Library.

Ismail, H. (2014). *Extracting the psychometric properties of the teacher quality standards scale on a sample of teachers in the state of El Oued* (Unpublished master's thesis). University of Blida.

Kim, S., & Feldt, L. (2008). A comparison of tests for equality of two or more independent Alpha coefficients. *Journal of Educational Measurement*, *45*(2), 179-193. https://doi.org/10.1111/j.1745-3984.2008.00059.x

Melhem, S. M. (2002). *Measurement and evaluation in education and psychology* (2nd ed.). Dar Al-Masirah.

Onn, D. (2013). *Classical test theory versus item response theory: An evaluation of the comparability of item analysis result*, Retrieved from https://ui.edu.ng/sites/default/files.

Saeed, M. (2015). *Modern trends in educational measurement and evaluation: Achievement file*. Dar Al Nahda Al Arabiya, Cairo.

Saeed, M. (2019). Actuality of secondary school exam scores in predicting the achievement of first-year students at the Faculty of Education, Beni Suef University. *Arab Journal of Measurement and Evaluation,1*(2) -84. https://doi.org/10104 10.21608/AJME.2020.200201

Saeed, M. (2023). Shifting from learning assessment to assessment for learning. *Journal of the Faculty of Education, Beni Suef University, 20*(11), 1–11. https://doi.org/10 10.21608/JFE.2023.337355.

Stanley, J. C., & Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation*. Prentice-Hall.

Thompson, B., Green, S., & Yang, Y. (2010). Assessment of the maximal half-split coefficient to estimate reliability. *Educational and Psychological, 70*(2) 232–251. https://doi.org/10.1177/0013164409355688.

Trevisan, S., Sax, G., & Michael, W. (1991). The impact of student's ability on test actuality and reliability. *Educational and Psychological Measurement, 51,* 829- 837.

Walker, D. (2006). A comparison of the Spearman-Brown and Flanagan-Rulon formulas for half-split reliability under various variance parameter conditions. *Journal of Modern Applied Statistical Methods, 5*(2), 443–451. http://digitalcommons.wayne.edu/jmasm/vol5/iss2/18

Zare', N. (2021). Comparison of the coefficients of the reliability of the test scores under sets of conditions: Monte Carlo simulation study. *Educational journal, 2*(88), 1108- 1174.

Zimmerman, D. W., Williams, R. H., & Symons, D. L. (1984). Empirical estimates of the comparative reliability of matching tests and multiple-choice tests. *Journal of Experimental Education, 52*(3), 179–182. https://doi.org/10.1080/00220973.1984.1101189

About the Contributor(s)

Omar Saleh Bani Yassin, PhD, Educational Sciences Department, Irbid University College, Al-Balqa Applied University
Email: omarsa@bau.edu.jo
ORCID: https://orcid.org/0000-0002-4495-1030

Aiman Mohammad Freihat, PhD, Associate Professor of Measurement and Evaluation at Educational Sciences Department, Ajloun University College, Al- Balqa Applied University
Email: aiman.freihat@bau.edu.jo
ORCID: https://orcid.org/0000-0002-4161-4143

Sabri Hassan Al-Tarawneh, PhD, Department of Psychology, Faculty of Educational Sciences, Mu'tah University
Email: Sabri_razan@yahoo.com
ORCID: