Comparing Some Methods of Split-Halve Reliability

Ammar Yasser Mohammed Al-Jubouri

+9647807808121

ammar.yasir2202m@ircoedu.uobaghdad.edu.iq

University of Baghdad / College of Education Ibn Rushd for Human Sciences

Prof. Dr. Balqees Hamood Kazem

+9647901812673

balqees.hmood@ircoedu.uobaghdad.edu.iq

University of Baghdad / College of Education Ibn Rushd for Human Sciences

Abstract:

The current research aims to Comparing Some Methods of Split-half Reliability. To achieve the research objective, the researcher adopted the intelligence test (Alfred W. Munzert, 1994), which is published by the American institution (Prentice Hall) and consists of 60 items. The researcher then prepared the test according to the procedures followed in the development of psychological tests and presented it to 17 experts to verify the test's apparent validity. The experts provided their feedback on the test items, leading to all necessary modifications and the removal of 4 items. Subsequently, the researcher administered the test to a pilot sample of 60 male and female students in the intermediate stage, paving the way for its application to the statistical analysis sample of 400 students, selected through stratified random sampling from the fourth and fifth grades in intermediate schools in Kirkuk Governorate during the academic year (2023-2024). The students' responses were then corrected, and the results were organized into a data table using Excel, after that, I made a comparison between the classifications (iterative Split, random split, equal difficulty, equal discrimination, equal reliability). Then, we used methods for split-half reliability (Spearman-Brown, Horst, Guttman, Flanagan-Rulon, Angoff-Feldt) and created a matrix for each classification method and each reliability method. This way, each reliability method was calculated with each classification method, and the results indicated that there were no statistically significant differences between the methods.

Keywords: split-half, reliability, test, intelligence.

split-half: The internal consistency coefficient is obtained by splitting the test into two halves, using one half of the test items to obtain one score, and using the other half of the items to obtain a separate score. The correlation between these scores reflects the reliability of the half-test. We then calculate specific equations to obtain the overall reliability coefficient (AERA, APA, NCME, 2014,page 223).

reliability: It means that the test measures what it is intended to measure and does not measure something different (Hogan, 2015,page 150).

Test: It is a standardized procedure for taking a sample of behavior and describing it in categories or degrees (Gregory, 2015, page 23).

Intelligence: Munzert (1994): It is an ability that indicates the presence of a general factor that appears in all mental activities and a specific factor that only appears in certain activities (Munzert, 1994, page 35).

Introduction:-

The differences in methods for estimating reliability make it very difficult to obtain an accurate estimate of the results from one method when only the estimate obtained from another method is known. Researchers often assume that the split-half method yields excessively high estimates, which is an inaccurate observation unless it is based on a defensible standard. Dividing a test composed of several items into two halves provides each splitting method with its own estimate of the reliability coefficient, and it is not necessary for all splitting methods to be equally valid and effective from a priori perspectives. There are significant differences in the value of the reliability coefficient depending on how the halves are divided Kuder & Richardson (1937). in study (Brownell, 1933) pointed out of split-half methods that it is unreasonable for the same test given to students to have a reliability coefficient expressed by values ranging from 0.20 to 0.48, depending on how the test is divided. He advised researchers to track the relationships involved between these correlations, which in turn lead to variations in reliability coefficients. He also noted that most studies conducted on this method focused on correction formulas for test length without addressing the split methods .

it was found that as more splits were made, it became increasingly difficult to maintain similarity between the halves, and the splits became noticeably different from one another (Cronbach,1946). we observed that the majority of these studies use the (odd-even) method when calculating the reliability of tests and measures through split-half reliability, regardless of how suitable it is for the content of the test and how accurate it is in obtaining results, despite the existence of more than one method for calculating reliability through split-half. This prompted him to conduct this research in an attempt to answer the following question, page What is the impact of split-half methods on the reliability of intelligence tests?

Goals of research

- Calculate the methods of splits (iterative Split, random split, equal difficulty, equal discrimination, equal reliability, less more difficult ,odd -even).

- Calculate methods of reliability split-half (Spearman-Brown, Horst, Guttman, Flanagan-Rulon, Angoff-Feldt).

Literature review

Classical Test Theory :-

To clarify the relationship between the reliability coefficient and the true score, we will present the mathematical derivation that illustrates the connection between them. Since the current research topic is split-half reliability, we will assume that we have applied a single test format and divided it into two parts. In this test, we will have two observed scores, each consisting of a true score and an error score, as follows, page

X = T + E……………………1

X' = T' + E'…………………2

To find the correlation between the two halves of the test, considering that the concept of test reliability is based on obtaining the same scores across different applications using the same tool or equivalent tools, we must recall the five mathematical principles that we presented at the beginning of this chapter regarding the basic assumptions underlying classical measurement theory. By adhering to these three assumptions, we will rely on the scientific basis for reaching the true score after isolating random errors. This derivation is as follows, page

$$
\begin{aligned}
\rho_{XX'} &= \operatorname{corr}(X, X') \\
&= \frac{\operatorname{cov}(X, X')}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(X')}} \\
&= \frac{\operatorname{cov}(T+E, T'+E')}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(X')}} \\
&= \frac{\operatorname{cov}(T, T') + \operatorname{cov}(T, E') + \operatorname{cov}(E, T') + \operatorname{cov}(E, E')}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(X')}} \\
&= \frac{\operatorname{cov}(T, T')}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(X')}} \\
&= \frac{\operatorname{cov}(T, T)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(X')}} \\
&= \frac{\operatorname{Var}(T)}{\operatorname{Var}(X)}
\end{aligned}
$$

We observe that mathematical derivation began by using the correlation coefficient to verify theoretical assumptions. The theory assumes the existence of two tests (in the current research, they are half-tests since the topic is split-half reliability). It places the variance between the two halves in the numerator and the observed variance and true variance in the denominator. We then begin deriving the degrees (true and error) from each half-test.

Next, the variance is distributed across all terms of the equation, resulting in four terms as follows, page

•The variance between the true scores of the two halves.

•The variance between the true score of the first half and the error score of the second half.

•The variance between the error score of the first half and the true score of the second half.

•The variance between the error scores of the two halves.

Returning to the assumptions set by classical test theory, we find that the correlation between true scores is constant, the correlation between true scores and error scores equals zero, and the correlation between error scores also equals zero. Thus, we conclude that the remaining variances are true variances based on the variances of the observed scores in the two test halves, repeated over multiple applications. Therefore, we express reliability as the ratio of the deviation of the observed scores from the deviation of the true scores (Wu et al, 2016),

Reliability can be calculated in this way by administering the test once to a group of individuals, then dividing it into two equivalent halves. After that, we calculate the correlation coefficient between these two halves, and the correlation coefficient obtained from this method is considered the half-test reliability. To find the reliability of the entire test, the Spearman-Brown formula is applied, and this method is one of the ways to calculate the internal consistency of the test .Participants usually do not know that the test will be divided into two parts, nor do they know which items belong to which half, and often the test administrator does not decide how to divide the items until they see the test results. From (Gulliksen, 1950) perspective, effective split procedures are desirable, and it is preferable to determine the appropriate split into two parts beforehand, "before printing the test." It is clear that the half-test reliability provides a measure of consistency concerning the sample content. The reliability of results over time is not included in this type of reliability, as the test is administered in a single session (Anastasi, 1976).

The aim of creating two parallel test halves is to target or measure the same true scores with a high degree of accuracy. There is a way to verify whether the tests are parallel by ensuring that several statistical indicators are equal between the two halves of the test, which are (Equal mean, Equal standard deviations (Price, 2017)

Methods of split-half reliability: -

1. **Odd and even**:-

This is one of the most common methods of classification, as almost no measurement source is devoid of this approach. Some sources limit the method of half-splitting to this classification only. This method was mentioned by the scholar Spearman in a scientific paper, where he believed that this method is suitable for alternating the split of test items and considered that any difference between the two groups in the test items when using this method is due to chance, The idea behind this method is to divide the test items into two halves from beginning to end, where the first half consists of items with odd numbers and the second half consists of items with even numbers. The advantage of this method is that it controls for fatigue factors that affect performance, and these are accounted for in the calculation of error variance estimated by split-half reliability methods. (Spearman, 1910, page 275).

2. **Split-Half Reliability Based on Difficulty Index:-**

Using the split-half method to estimate reliability requires that the halves are parallel in content and statistical properties. Cronbach introduced a method for constructing parallel halves. Similar to the parallel-forms method, the split-half method requires the test developer to divide the test items to create two similar models in terms of difficulty level and logically consistent content. To do this, the test developer first conducts an item analysis on a set of papers. Information from the item analysis, specifically the difficulty index, is then used to select pairs of items with similar difficulty levels (Johnson & Penny, 2005,page 651). In this method, items are ordered according to their difficulty level, based on the difficulty index (p), with odd-numbered items forming the first subtest and even-numbered items forming the second subtest (Crocker & Algina, 2008, page 136). This produces two parallel subtests with the same difficulty level and homogeneous items. The parallelism of the tests can be verified by checking for equality of means and standard deviations. The purpose of parallel tests is to measure the same true scores with a high degree of accuracy. If the halves contain items with partially disparate content, they will measure different true scores; therefore, the halves must be matched based on the difficulty level and content they measure (Price, 2017, page 230). This method has been shown to produce halves with equal differences, with results similar to random splitting. This can lead to an increased proportion of comparable splits, saving effort, training, and practice (Cronbach, 1946, page 477, 480).

3. **Iterative-Split Method:**

This method is known as the Iterative-Split method and was developed by the Indian psychologist Satyendra Nath Chakrabartty. It is based on dividing a test into two halves, but unlike traditional methods, it does not rely on simple random splitting or dividing items into odd and even groups. Instead, this method splits the test based on the frequency of responses to the test items.The frequencies of responses for each item are calculated, and then the items are distributed between the two halves according to these frequencies so that the halves are as equivalent as possible. Chakrabartty demonstrated in his studies (Chakrabartty, 2013; Chakrabartty, 2015) that this method produces more balanced halves compared to traditional methods such as odd-even splitting. The study (Chakrabartty, 2013) showed that the iterative-split method yields superior parallel halves.

4. Suggested Methods:-

The researcher reviewed split-half methods found in measurement and assessment sources and references, as well as previous studies. Other methods were mentioned in Brownell's (1933) study, such as the "most-least difficult," "equal reliability in consistency," and "most consistent-least consistent" methods. These methods calculate the coefficient of internal consistency between test halves based on the general principle of split-half reliability, defined by the following steps, page

- Calculate the index value (difficulty, discrimination…..) for each item
- Distribute the items between the two halves based on the calculated index value (difficulty, consistency, etc.), according to the desired split.
- Calculate the correlation coefficient between the two halves.
- Correct the correlation coefficient using the Spearman-Brown prophecy formula. In addition to a set of arbitrary splits, as described by Cronbach (which did not rely on a specific principle) (Brownell, 1933; Cronbach, 1946).

  Devellis (2003) also mentioned the causal path method. He views this as a method that allows for obtaining equivalent subtests from the original test, considering each item as a subtest. The method relies on linking the two halves of the subtest using the causal paths of the latent variable to each half of the test. Therefore, the values of these paths equal the correlation between the tests. If the path values must be equal (which they are under the assumptions of this model), then the correlation between the tests equals the square of the path value from the latent variable to either test. Assuming this is a standardized path coefficient, it is also the proportion of variance in each test influenced by the latent variable. This, in turn, is the definition of reliability; therefore, the correlation between the two tests equals the reliability of each (Devellis, 2003, page 41). Based on this, the researcher proposes conducting half-splits relying on other statistical indicators that have been mentioned, which include split using indicators of (discrimination, .). The researcher also intends to compare methods (least-most difficult, least-most discriminating) and is based on the following fundamental principles, page

  •This method has not been mentioned in previous studies to the best of the researcher's knowledge, but the idea for this method has emerged.

  •Relying on the opinions of both Cronbach (1946) and DeVellis (2003), who indicated the possibility of conducting a large number of parallel splits based on certain attributes such as content and statistical characteristics, with discrimination and standard deviation being considered statistical properties.

  • Classical measurement theory relies on the principle of measuring individual differences, and these indicators are seen as markers of the existence of these differences, as they measure the deviation of data from the mean, meaning the deviation of some individuals' scores from the normal score that represents the average of the distribution. Accordingly, we will have the following methods,

  Split according to the item discrimination index,

  Through this method, items are divided into two halves based on the item discrimination index (Brownell, 1933).

Split-Half method of Reliability

Ebel and Frisbie (1991) state that test splitting means the scores used to determine reliability represent half the length of the test (half of the total performance sample). Therefore, it is necessary to correct the test correlation to estimate the reliability of the full-length test.In general, longer tests tend to be more reliable than shorter ones because measurement errors resulting from sampling the content decrease as the test length increases. Additionally, test length affects the variance of true and observed scores (Crocker & Algina, 2009).

Since the number of possible splits that can be created from a test depends on the total number of items in that test, obtaining a consistent estimate of reliability improves as the number of items increases, and conversely, it decreases as the number of test items decreases (Carmines & Zeller, 1979). On this basis, methods for calculating split-half reliability are classified into two types:

I. **Spearman-Brown Method**:
The Spearman-Brown formula was independently developed by the mentioned researchers to quantitatively show the relationship between a half-test and the full-length test. This method calculates the correlation coefficient between the two halves of the test (this correlation represents the reliability coefficient for each half, rather than the whole test). Then, the Spearman-Brown prophecy formula is applied to this correlation value to estimate the overall reliability (Geisinger, 2013). The formula appears in several versions,

$$\boldsymbol{r_{xy} = \frac{2r}{1+r}}$$

Where:
r_xy = reliability coefficient for the full-length test
r = correlation coefficient between the two halves of the test(Cohen & Swerdlik, 2010, page146)

Users of this formula should note that the method primarily relies on the assumption that the two halves of the test are parallel. The more this assumption is violated, the less accurate the results become (Crocker & Algina, 2009, page 187). The concept of equivalence assumes equality in the following statistical measures between the two halves:

- Means
- Standard deviations
- Intercorrelation coefficients
- Difficulty level of test items (Price, 2017)

II. **Horst's Method: -**

Horst (1951) proposed an alternative approach to the Spearman-Brown formula for calculating split-half reliability. This method estimates reliability in a way similar to Spearman-Brown by dividing the test and calculating the correlation between the two halves. However, it is distinguished by its ability to estimate reliability when the two halves are unequal in length. This equation is used when we have two halves that are comparable in all respects except length. Thus, reliability can be estimated using this method if we know the proportion of responses on each half of the test (Horst, 1951:368). The formula is defined as follows:

$$\boldsymbol{H} = \frac{\boldsymbol{r}\sqrt{\boldsymbol{r}^2 + 4\boldsymbol{p}_1\boldsymbol{p}_2(1 - \boldsymbol{r}^2)} - \boldsymbol{r}^2}{2\boldsymbol{p}_1\boldsymbol{p}_2(1 - \boldsymbol{r}^2)}.$$

H = Horst's reliability coefficient.
r = correlation between the two halves of the test.
r^2 = square of the correlation between the two halves of the test.
p_1 = the larger half.
p_2 = the smaller half.

III. Guttman Method:-

Guttman (1945) proposed a general equation to calculate reliability by computing the variances between the two halves of a test (Lambda). This equation represents the fourth limit (L4) among Guttman's six limits. Due to the strong influence of measurement independence, it is essential to focus on the information obtainable from a single measurement, which provides an estimate of the minimum reliability. Consequently, this calculation approach eliminates the assumptions regarding the need for measurement at two or more time points and the methodological difficulties arising from the assumption of measurement independence (Guttman, 1945: 255–260).

One of the features of Guttman’s calculation equation is that it does not require equal variances between the halves, nor does it require a correction formula as is the case with the Spearman–Brown equation (Geisinger, 2013:34).

In 1950, Gulliksen demonstrated that assuming equal variances ($\sigma 1^2 = \sigma 2^2$), the formula (L4) is equivalent to Spearman’s equation. Since this formula provides a lower bound for reliability, the test’s reliability coefficient may sometimes be lower than its true estimate. This fact may explain why corrected correlations between halves are sometimes higher than those from a single test. The general formula of Guttman’s (L4) equation is as follows:

$$r_w = 2\left(1 - \frac{\sigma_a^2 + \sigma_b^2}{\sigma_w^2}\right)$$

r_w= Test reliability coefficient

σ_a^2= Variance of the first half

σ_b^2=Variance of the second half

σ_w^2=Total variance (Johnson & Penny, 2005, page 652).

IV. Flanagan -Rulon Method:-

Flanagan (1937) and Rulon (1939) proposed a different formulation for split-half reliability, which takes into account the variation in results. Their interpretations, together with Spearman's concept, formed the foundation of the classical explanation of reliability. They introduced a formulation using correlation and standard deviations. Moreover, the Flanagan-Rulon equation is considered a special form of the more commonly used alpha coefficient in reliability calculations

This equation is used in cases where there are unequal standard deviations for the two halves of a test. Regardless of the correlation between the halves, this formula provides the best estimate of split-half reliability in such situations (Walker, 2005, page 443).

Guilford (1942) points out that if this formula is applied to calculate a test's split-half reliability with any division of the halves, the reliability coefficient is estimated fully as the reliability of half the test doubled. Therefore, this method does not require applying the Spearman-Brown correction formula , The equation aims to estimate reliability by considering only the true variance in performance, emphasizing that the difference between the logical variance assumed to be expressed by the test and the error variance, which ranges between 0 and 1, represents the actual reliability coefficient of the test. Hence, in the case of absolute reliability, it is expected that there will be no difference in an individual's performance across the two halves of the test. The equations appeared in the following forms:—

$$\rho_{\text{Flanagan}} = 2\left(1 - \frac{var_1 + var_2}{var_X}\right)$$

$$\rho_{\text{Rulon}} = 1 - \frac{var_e}{var_x}$$

$$\rho_{\text{Flanagan-Rulon}} = \frac{(4r_{1,2}SD_1SD_2)}{var_x}$$

var_1: Variance of the first half var.

var_2:Variance of the second half.

var_x: Total test variance.

SD_1: Standard deviation of the first half.

SD_2: Standard deviation of the second half .

r: Correlation coefficient between the two halves (Warren, 2015, page 4).

Angoff-Feldt Method:-

This coefficient is used to estimate the consistency of results derived from two different halves of a test and is considered an important alternative to Cronbach's alpha and Spearman-Brown coefficients in certain cases, such as when there are differences in the level of standard deviation between the two halves. The coefficient takes into account variations in the standard deviation between the two parts and is regarded as a good option for estimating test reliability when split into two halves, especially when achieving equivalence between the parts is difficult. Feldt and Charter (2003) demonstrated that the equation can be expressed in the following form:

$$r_{XX'} = \frac{4(r_{12})(S_{X_1})(S_{X_2})}{S_{X_{\text{tot}}}^2 - \left(\frac{S_{X_1}^2 - S_{X_2}^2}{S_{X_{\text{tot}}}}\right)^2}.$$

$r_{XX'}$= Anhof Feldt coefficient
r = Correlations between deviations of the two halves
S_{X_1} = Standard deviation of the first half
S_{X_2}= Standard deviation of the second half
$S_{X_{\text{tot}}}$= Overall standard deviation of the test (Warren, 2015, page 6).

Feldt and Charter (2003) identified the cases in which the Spearman-Brown, Flanagan-Rollon, and Angoff-Feldt formulas are used based on the differences in standard deviations between the two halves, as follows: • If the differences between the two halves do not exceed 5%, the Spearman-Brown formula is used. • If the differences between the two halves range from 15% to 30%, the Flanagan-Rollon formula is used. • If the differences between the two halves exceed 30%, the Angoff-Feldt formula is used .

Research Methodology and Procedures:-

To create two parallel halves of the test, the researcher followed the method outlined by Chakrabarty (2013) to achieve equivalence between the halves, which is based on two main principles, page

- Arranging the responses in ascending order based on their statistical indicators.
- Using the method of reciprocal distribution to allocate items between the two halves, followed by applying the split-half method.

We can confirm that the two halves are parallel, as noted in various literatures and previous studies, including Cronbach (1946), Devillis (2003), and Price (2017). This is evidenced by the closeness of the means and standard deviations between the two halves, as this high degree of proximity enables them to accurately measure the same true scores with a high level of precision. Items can be distributed manually or automatically via computer. To carry out this process more accurately and to prevent distribution errors, the researcher modeled these steps in the form of programming codes, which were then input into the R program.

- Step One, page After obtaining the statistical characteristics of the items, they are arranged in ascending order based on their statistical indicators, as shown in Table (1).

N	Number item	Item Difficulty	Number item	Standard divisions	Number item	Frequency
1	5	0.202	5	0.401	5	81
2	26	0.207	26	0.406	26	83
3	2	0.210	2	0.407	2	84
4	34	0.215	34	0.410	34	86
5	7	0.230	7	0.420	7	92
6	25	0.230	25	0.420	25	92
7	45	0.237	45	0.425	45	95
8	28	0.240	28	0.427	28	96
9	11	0.245	11	0.431	11	98
10	19	0.25	19	0.434	19	100
11	52	0.258	52	0.438	52	103
12	20	0.26	20	0.439	20	104
13	36	0.268	36	0.443	36	107
14	42	0.292	31	0.448	42	117
15	29	0.305	42	0.455	29	122
16	53	0.308	29	0.461	53	123
17	47	0.315	27	0.462	47	126
18	23	0.322	53	0.462	23	129
19	4	0.332	47	0.465	4	133
20	46	0.332	49	0.466	46	133
21	22	0.335	23	0.468	22	134
22	18	0.345	4	0.472	18	138
23	39	0.348	46	0.472	39	139
24	38	0.370	22	0.473	38	148
25	43	0.372	18	0.476	43	149
26	3	0.375	39	0.477	3	150
27	6	0.39	24	0.482	6	156
28	48	0.395	40	0.482	48	158
29	9	0.408	38	0.483	9	163
30	44	0.408	43	0.484	44	163
31	30	0.41	3	0.485	30	164
32	13	0.415	6	0.488	13	166
33	10	0.428	48	0.489	10	171
34	50	0.432	1	0.492	50	173

35	17	0.435	8	0.492	17	174
36	32	0.455	9	0.492	32	182
37	12	0.462	30	0.492	12	185
38	16	0.478	44	0.492	16	191
39	35	0.478	13	0.493	35	191
40	37	0.482	10	0.495	37	193
41	33	0.49	17	0.496	33	196
42	51	0.507	41	0.496	51	203
43	21	0.52	50	0.496	21	208
44	15	0.545	14	0.497	15	218
45	14	0.556	54	0.497	14	222
46	54	0.557	12	0.499	54	223
47	41	0.565	15	0.499	41	226
48	1	0.59	32	0.499	1	236
49	8	0.593	16	0.5	8	237
50	40	0.634	35	0.5	40	253
51	24	0.635	37	0.5	24	254
52	49	0.682	33	0.501	49	273
53	27	0.693	51	0.501	27	277
54	31	0.723	21	0.69	31	289

• Step Two, page - Apply the method of alternate distribution between the two halves in order to distribute the items between the two halves of the test concerning the index to be divided into halves. This method involves the following steps, page

 Divide the items in the test into groups of four, so that each group contains four items.

 Alternate the distribution of the groups between the two halves by placing the first item in the first half of the test, then the second item in the second half, the third item in the second half, and the fourth item in the first half. This is done to minimize the differences between the two halves as much as possible, ideally making them zero or close to zero, as illustrated in Tables (2) and (3).

Table (2)explain distribution of item between halves

Differences	Second half	First half
<= 0	Item two	Item one
> = 0	Item three	Item four

After that, the distribution of the paragraphs between the two halves was as follows, page

Table(3)

Table (3) applied distribution of item between halves

N	Difficulty-split				Standard divisions-split				Iterative-split			
	n	F	n	S	n	F	n	S	n	F	n	S
1	5	0.202	26	0.207	5	0.401	26	0.406	5	81	26	83
2	34	0.215	2	0.210	34	0.410	2	0.407	34	86	2	84
3	7	0.230	25	0.230	7	0.420	25	0.420	7	92	25	92
4	28	0.240	45	0.237	28	0.427	45	0.425	28	96	45	95
5	11	0.245	19	0.25	11	0.431	19	0.434	11	98	19	100
6	20	0.26	52	0.258	20	0.439	52	0.438	20	104	52	103
7	36	0.268	42	0.292	36	0.443	31	0.448	36	107	42	117
8	53	0.308	29	0.305	29	0.461	42	0.455	53	123	29	122
9	47	0.315	23	0.322	27	0.462	53	0.462	47	126	23	129
10	46	0.332	4	0.332	49	0.466	47	0.465	46	133	4	133
11	22	0.335	18	0.345	23	0.468	4	0.472	22	134	18	138
12	38	0.37	39	0.348	22	0.473	46	0.472	38	148	39	139
13	43	0.372	3	0.375	18	0.476	39	0.477	43	149	3	150
14	48	0.395	6	0.39	40	0.482	24	0.482	48	158	6	156
15	9	0.408	44	0.408	38	0.483	43	0.484	9	163	44	163
16	13	0.415	30	0.410	6	0.488	3	0.485	13	166	30	164
17	10	0.428	50	0.432	48	0.489	1	0.492	10	171	50	173
18	32	0.455	17	0.435	9	0.492	8	0.492	32	182	17	174
19	12	0.462	16	0.478	30	0.492	44	0.492	12	185	16	191
20	37	0.482	35	0.478	10	0.495	13	0.493	37	193	35	191
21	33	0.49	51	0.507	17	0.496	41	0.496	33	196	51	203
22	15	0.545	21	0.520	14	0.497	50	0.496	15	218	21	208
23	14	0.556	54	0.557	54	0.497	12	0.499	14	222	54	223
24	1	0.59	41	0.565	32	0.499	15	0.499	1	236	41	226
25	8	0.593	40	0.634	16	0.5	35	0.500	8	237	40	253
26	49	0.682	24	0.635	33	0.501	37	0.501	49	273	24	254
27	27	0.693	31	0.723	51	0.501	21	0.69	27	277	31	289
Sum		10.886		10.883		12.592		12.795		4354		4353
Average		0.403		0.403		0.466		0.473		161.259		161.222
Standard Division		0.141		0.137		0.039		0.056		56.640		56.262

After that, we calculated the reliability using each method of the split, and the results are as shown in the following table(4)

Table (4) result split-half reliability

N	Method Of split	correlation	Method reliability				
			Spearman -Brown	Horst	Guttman	Flanagan -Rulon	Angoff -Feldt
1	**Odd-Even**	0.724	0.840	0.840	0.839	0.840	0.840
2	**Random**	0.739	0.847	0.847	0.846	0.846	0.846
3	**Standard division**	0.735	0.847	0.847	0.847	0.847	0.847
4	**Iterative**	0.735	0.847	0.847	0.847	0.847	0.847
5	**Item difficulty**	0.735	0.847	0.847	0.847	0.847	0.847
6	**Item discrimination**	0.748	0.851	0.851	0.851	0.851	0.851
7	**First-second**	0.733	0.846	0.846	0.845	0.846	0.846

Through comparing methods of split-half reliability, it has been found that there are no statistically significant differences between these methods in terms of their impact on reliability coefficients. Thus, the two null hypotheses, which indicate no differences between reliability and validity coefficients across different split-half methods, are accepted. This means that the split-half method has approximately the same reliability value regardless of how the division is made. This aligns with Cronbach's study, which indicated that results are similar in random splits and parallel splits, especially in tests measuring a single general factor (Cronbach, 1946, page480).

After conducting the half-splits (odd-even, least-most difficult, equal difficulty fairness, least-most discrimination, equal discrimination fairness, equal fairness), the reliability was then calculated using each of the methods (Spearman-Brown, Horst, Guttman, Flanagan-Rulon, and Angoff-Feldt) for each split. Table (2) shows the effect of these different splits on the reliability and validity coefficients.

References:-

1. Anastasi, A. (1976). Psychological testing (4th ed.). MacMillan Publisher Co.

2. Brownell, W. A. (1933). On the accuracy with which reliability may be measured by correlating test halves. The Journal of Experimental Education, 1 (3), 204–215. https://doi.org/10.1080/00220973.1933.11009901
3. Carmines, E. G., & Zeller, R. (1979). Reliability and validity assessment (Vol. 17). SAGE Publications.
4. Chakrabartty, S. (2013). Best split-half and maximum reliability. IOSR Journal of Research & Method in Education (IOSR-JRME), 3 (1), 1–8. http://www.iosrjournals.org
5. Chakrabartty, S. N., Wang, K., & Chakrabartty, D. (2015). Uncertainty in test score data and classically defined reliability of tests & test batteries, using a new method for test dichotomisation. India.
6. Cohen, R. J., & Swerdlik, M. E. (2010). Psychological testing and assessment: An introduction to tests and measurement (7th ed.). McGraw-Hill.
7. Crocker, L., & Algina, J. (2008). Introduction to classical and modern test theory . Wadsworth Publishing Company.
8. Cronbach, L. J. (1946). A case study of the split-half reliability coefficient. Journal of Educational Psychology, 37 (8), 473–480. https://doi.org/10.1037/h0054328.
9. DeVellis, R. F. (2003). Scale development: Theory and applications (2nd ed.). SAGE Publications.
10. Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Prentice Hall.
11. Feldt, L. S., & Charter, R. A. (2003). Estimating the reliability of a test split into two parts of equal or unequal length. Journal of Applied Psychology, 8 (1), 102–109. https://doi.org/10.1037/1082-989X.8.1.102
12. Flanagan, J. C. (1937). A proposed procedure for increasing the efficiency of objective tests. Journal of Educational Psychology, 27 , 17–21.
13. Geisinger, K. F. (Ed.). (2013). APA handbook of testing and assessment in psychology . American Psychological Association.
14. Guilford, J. P. (1942). Fundamental statistics in psychology and education . McGraw-Hill.
15. Gulliksen, H. (1950). Theory of mental tests . Wiley.
16. Guttman, L. (1945). A basis for analyzing test-retest reliability. Psychometrika, 10 (4), 255–282. https://doi.org/10.1007/BF02288892
17. Horst, P. (1951). Estimating total test reliability from parts of unequal length. Educational and Psychological Measurement, 11 , 368–371. https://doi.org/10.1177/001316445101100306
18. Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2 (3), 151–160. https://doi.org/10.1007/BF02288391
19. Munzert, W. A. (1994). Test your IQ (3rd ed.). Prentice Hall.
20. Price, L. R. (2016). Psychometric methods: Theory into practice . Guilford Publications.
21. Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. Harvard Educational Review, 9 , 99–103.

22. Spearman, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3 (3), 271–295. https://doi.org/10.1111/j.2044-8295.1910.tb00206.x
23. Walker, D. A. (2006). A comparison of the Spearman-Brown and Flanagan-Rulon formulas for split-half reliability under various variance parameter conditions. Journal of Modern Applied Statistical Methods, 5 (2), 443–451. https://doi.org/10.22237/jmasm/1167642160
24. Warren, J., & Matthijs, M. (2015). A comparison of reliability coefficients for psychometric tests that consist of two parts. Methods, Data, Analyses . https://doi.org/10.1007/s11634-015-0198-6
25. Wu, M., Tam, H. P., & Jen, T.-H. (2016). Educational measurement for applied researchers . Springer Nature Singapore. https://doi.org/10.1007/978-981-10-3302-5