Comparison of the Reliability Test of Semester Final Exam Scores for Graphic Media Courses Using Various Reliability Test Methods

Hari Sugiharto Setyaedhi[1], A. Ardianik[1], M. Hanif[1]

[1]Universitas Negeri Surabaya, Surabaya, Indonesia
harisugihartosetyaedhi@unesa.ac.id

Abstract. The purpose of this study was to determine the accuracy of the estimated reliability coefficients of several reliability coefficient methods. Data were analyzed using the Spearman-Brown formula, the Flanagan formula, the Rulon formula, the Hoyt formula, the Kuder-Richardson 20 and 21 formulas, and the Alpha-Cronbach formula. This research was descriptive quantitative. The data were obtained from the final semester examination answer sheets for the Graphic Media Course from 90 students. Based on the calculation results, it was known that: a) the reliability coefficient of all formulas has almost the same estimation, which is around 0.8. b) The average reliability coefficient is estimated to be around 0.8; c) Spearman Brown odd-even is 0.837; Spearman-Brown first-last is 0.820; Flanagan odd-even is 0.85; Flanagan first-last is 0.8; Rulon odd-even is 0.865; first-last Rulon is 0.805; Kuder Richardson 20 is 0.820; Kuder Richardson 21 is 0.8; d) All reliability coefficients have reliability above 0.70, which means they are high. The results showed that the odd-even Rulon coefficient was the highest, so it was appropriate for estimating the final semester examination

Keywords: Graphic media, Exam Scores, Reliability Test

1 Introduction

Higher education aims to develop the potential of students to become human beings who have faith and piety in God Almighty, have good morals, and are healthy, capable, creative, skilled, competent, and cultured (Huda et al., 2018; Pangalila, 2017). Graphic media development courses are designed to communicate facts, ideas, and messages clearly and powerfully (Manshur & Rodhi, 2020a). Development of graphic media, including visual media, that is conveyed by relying on the visual senses and the message conveyed is poured into the form of visual symbols to achieve learning objectives (Akbar et al., 2021). Graphic media has the function of channeling messages from sources to message recipients in the form of visual communication symbols (Manshur & Rodhi, 2020b). Graphic media, in the form of pictures and writing, are the elements contained in graphic media (Purwani et al., 2019). The main function of graphic media

A. Mustofa et al. (eds.), *Proceedings of the International Joint Conference on Arts and Humanities 2023 (IJCAH 2023)*, Advances in Social Science, Education and Humanities Research 785,
https://doi.org/10.2991/978-2-38476-152-4_123

is to attract attention, clarify the presentation of ideas, and illustrate what might be forgotten more quickly if not graphically (Yulia Pramusinta, 2017). Lecturers as educators and creators must have high levels of creativity, namely by designing graphic media lectures as well as possible and delivering them in a fun way. This is done so that graphic media can be properly utilized by students (Halimatus et al., 2019). The development of graphic media on campus can improve the quality of learning (Subhi et al., 2022). Lecturers measure student learning outcomes by using test instruments in the form of end-of-semester exams (UAS).

Lecturers who measure learning outcomes aim to determine student understanding after studying for one semester (Supriyadi, 2017). The test instrument used to measure learning outcomes is called “a learning achievement test." (Magdalena, Syariah, et al., 2021). The learning result test is used as the basis for decision-making. The learning achievement test is used as an assessment tool to determine whether the student has achieved the learning objectives (Tri Jampi Setiyorini, Zyah Rochmad Jaelani, 2022). This is understandable because the test is a tool to measure whether educational goals are achieved (Hikmah & Muslimah, 2021). Tests can also determine the success of a learning program (Apsari & Acep Haryudin, 2017). The test also functions as a tool to inform students about their mastery of the learning material (Wenno et al., 2021). Tests can determine student achievement so that they can make the right decisions (Ulfah et al., 2020). The test is a planned effort carried out by lecturers to show learning outcomes to students (Kurniawati, 2019). so that students and lecturers can make the right decisions about learning. A quality test instrument produces reliable information.

The most popular test instrument used in colleges and schools is multiple choice or objective testing (multiple choice) (Kurniawan & Andriyani, 2018). Multiple-choice objective tests with answer choices (options) are available. Multiple-choice results can be scored quickly and objectively by the examiner (Magdalena et al., 2021; Zainal, 2020). Objective tests have the advantage of being time-efficient, being able to measure large numbers of test takers, and being easy to score (Murti et al., 2018). Multiple choice tests are qualified to measure reliability, and others (Nusantari, 2016). Multiple choices can be used to measure validity and reliability (Neti, 2020).

Test instruments that are of good quality and suitable for use are only possible if they are based on the applicable test development principles (Arifin, 2017). The test instrument can be said to be good if the item analysis has been carried out (Putri et al., 2020). In evaluating student learning outcomes, lecturers rarely analyze quantitatively (empirical) (Elviana, 2020). Ernawati's research stated that the test instruments used so far had been tested qualitatively, but quantitatively, the test instruments had never been carried out (Erawati, 2018). A good test instrument must fulfill the following requirements: 1) the item difficulty index is in the moderate category; 2) the item discrimination index can discriminate between clever and less intelligent test takers; and 3) distractors are selected at least 5% by all test takers. 4) good validity, and 5) high reliability (Arifin, 2017). The characteristics of a quality test instrument are valid, reliable, relevant, representative, practical, discriminatory, specific, and proportional (Arikunto, 2018; Zainal Arifin, 2019). Several stages must be carried out in developing test instruments, namely: designing tests, conducting trials, assessing validity and reliability, and interpreting test scores (Mardapi & Kertowagiran, 2011). The requirements for a good test are: a) reliable; b) valid; c) objective; d) discriminatory; e) comprehensive; f) easy to use (Purniasari et al., 2021). One of the requirements for developing a quality test instrument that is suitable for use as an evaluation tool for learning outcomes is having high reliability (Hanifah, 2014). This is following the thinking of measurement experts, who set the main criteria for measurement, one of which is to use the reliability formula (Azwar, 2016). Several aspects must be considered before the test instrument is used must have several conditions, such as 1) validity, the degree of accuracy of the measuring instrument for what should be measured (Sugiono et al., 2020), 2). reliable, meaning that if the test is tested on the same test takers up to several times, the test with different timeframes will give relatively the same results as long as the object being measured has not changed (Anastasi. Anne and Urbina, 2017; Farida & Musyarofah, 2021; Idrus, 2019; Sugiono et al., 2020). 3). objective, meaning that the test results are not affected by the subjective factors of the researcher, 4) balanced, meaning that the difficulty index of the items must be balanced with the objectives of the test, 5) discriminatory, meaning that the test must be able to distinguish between students who have high abilities and those who have low abilities; the test must be able to distinguish between high abilities and low abilities test-takers in the group; 6) norm, meaning that the test results must be easy according to certain standards or norms (Sulistianingsih, 2020). The thing that should be considered in a test instrument is whether it meets the demands of validity and reliability, namely the accuracy and consistency of the measurement results (Nengsi & Efrina, 2019). The measuring instrument is said to be qualified if it has been tested for validity and reliability (Dewi & Sudaryanto, 2020; Puspasari & Puspita, 2022). The test instrument is said to be good if it has evidence of validity and reliability so that it is suitable for use (Alfiatunnisa et al., 2022; Bashooir & Supahar, 2018; Budiantoro et al., 2021). This study only discusses reliability. Reliability is used to measure the reliability or constancy of a test instrument. Test instruments that have high reliability are expected to be a guide or reference for measuring student abilities (Sarwiningsih, 2017). The reliability of the test can be influenced by several factors, such as the characteristics of the test takers, test conditions, variations in test administration, errors and differences in scoring, length of the test, homogeneity of student abilities, and item difficulty level (D. Putri & Nahadi, 2019). The results of the study state that the reliability of the test is closely related to how the test is presented, the mood of the test takers, the attitude of the test takers when facing the test, motivation, the condition of the exam room, and so on (D. Putri & Nahadi, 2019). The procedures that must be carried out in developing a reliable test instrument are: 1) compiling test specifications; 2) writing test items; 3) analyzing the items qualitatively; 4) conducting test trials; 5) analyzing the items quantitatively; 6) revising the test; 7) composing the test; 8) carrying out the test; and 9) interpreting the test results (Ndiung & Jediut, 2020). Reliability is needed to determine whether the measuring instrument in the form of a test is reliable or consistent from time to time (Bahri, 2019; Erfan et al., 2020; Sanaky, 2021; Syahfitri et al., 2018) From the expert's opinion, it can be concluded that the concept of reliability refers to the consistency, stability, constancy, and reliability of measurement results (Arfah, 2021).

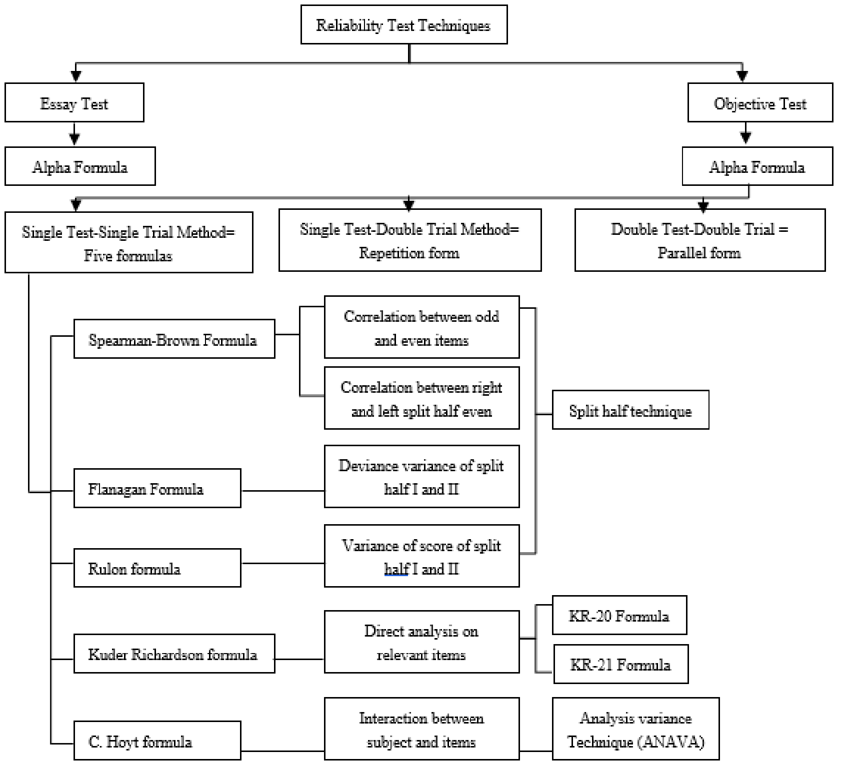
Lecturers are required to know how to determine the reliability coefficient of the test because the reliability of the test is related to the reliability of the test instrument, which is used to determine student learning outcomes. Facts in the field show that: 1) most of the lecturers do not understand how to determine the reliability of the test; 2) the lecturer has not made maximum use of the reliability coefficient of the test; 3) the lecturer has not fully utilized the reliability of the test; 4) the lecturer has not known the various reliability formulas that have been developed by the experts. Test instrument reliability can be tested with internal consistency (Suwartono et al., 2017). Gunartha stated that to estimate the reliability index of a measuring instrument, it can be tested through the following approaches: 1) internal consistency; 2) test-retest; and 3) parallel forms (Gunartha, 2022). Internal consistency is developed based on classical theory. The internal consistency approach aims to see consistency between items or sections in the test itself. For this reason, the test is divided into sections, and the number of items in each section is balanced. There are several formulas for estimating reliability with an internal consistency approach, such as 1) Spearman Brown's formula, 2) Flanagan's formula, 3) Rulon's formula, and 4) Hoyt's formula. 5) Kuder Richardson's formula (KR-20, KR-21), and 6) Alpha Cronbach formula (Gunartha, 2022; Puspasari & Puspita, 2022).

The urgency of this research is that educators or lecturers must understand how the reliability coefficient of a test is determined because accuracy in determining reliability is needed to determine the reliability of the test, which will later be used as a tool to determine student abilities. This is one of the expected improvements in the quality of education because high-reliability tests have a large influence on the analysis of student abilities as a reference for improving learning (Sarwiningsih, 2017). Testing the reliability of the device is considered very important because it is used as a measuring tool to obtain data and information related to the problem under study (Ayu & Rosli, 2020). Education and lecturers are important to know the reliability test because it determines how a measurement can be trusted because of its constancy (Yusup, 2018). Reliability is very important in measurement, especially in order to obtain consistent measurement results (Khumaedi, 2012).

Reliability refers to the stability of the measuring instruments used and their consistency from time to time. In other words. Reliability is the ability to measure instruments to provide similar results when applied at different times (Surucu & Maslakci, 2020). The study conducted by (Surucu & Maslakci, 2020) explains that in quantitative research, most of the predictor variables and outcomes are abstract concepts known as theoretical structures. Reliability is that the measuring instrument provides consistent results under the same circumstances. The validity of a measuring instrument to measure accurately without confusion with other features is defined as "validity". Validity is the level to serve the intended use of the scale. The use of valid and reliable measurement tools to measure these abstract concepts is an important factor in determining the quality of research.

Some relevant previous research examined the comparison of reliability. The research entitled “Quality Reliability Service Toward Student Satisfaction" stated that the reliability variable is a variable that has a dominant and significant influence on student satisfaction (Isnaeni et al., 2019). The study entitled "The Comparison Accuracy Estimation of Test Reliability Coefficients for the National Chemistry Examination in Jambi Province on Academic Year 2014/2015" concluded that various types of reliability coefficients have almost the same accuracy (Sarwiningsih, 2017). Research by (Widhiarso & Mardapi, 2010) on the "Comparison of Reliability Coefficient Estimation Among Classical Test Theory" concluded that each reliability coefficient has almost the same value. Test the validity and reliability of quantitative research instruments, stating that the split half test, KR-20, KR-21, and Cronbach's alpha obtained relatively similar results (Yusup, 2018). Those three pieces of research conclude the same result. Other research also revealed the relevant results. The study entitled "An Analysis of English National Final Examination for Junior High School in Terms of Validity and Reliability" calculated the results of the Middle School English National Final Examination reliability using the Kuder-Richardson Formula (KR-20). The results show that the National Middle School English Final Examination has fulfilled the characteristics of a good test in terms of reliability, with a reliability coefficient value of 0.89 (Sugianto, 2016). The research entitled "Validity and Reliability of Measurement Instruments of Educative Family Life" used Cronbach's alpha to test the reliability. The Cronbach Alpha reliability coefficient of this scale is 0.950 which is included in the "very good reliability" category. The results show that the structural dimensions of the child measurement instrument (CMI) and the structural dimensions of the mother's measurement instrument (MMI) were reliable (Sudiapermana & Setiawan, 2022).

Figure 1. Determination of the Reliability of Learning Outcomes Tests



The novelty of this study is that there have been enough previous studies that have used various reliability formulas of the split-half method and single test (single trial) to measure the estimation of test scores. The update in the context of this study is that the reliability coefficient is used to measure the extent to which Semester Final Examination (UAS) scores in graphic media courses are reliable. Cronbach's Alpha test is the most commonly used method for measuring the reliability of measurement instruments that have several items, but each reliability test has its strengths and weaknesses, therefore the novelty of this study is that the use of the five reliability tests mentioned above, the strengths and weaknesses can complement one another. By comparing the magnitude of the reliability coefficient of UAS scores for Graphic Media Courses using various technical tests, one can determine which reliability tests are reliable and can make a new contribution to understanding the reliability of measurement instruments in the context of graphic media courses.

This study aims to compare the reliability coefficient estimation accuracy of several reliability coefficient test methods for the final semester exams for graphic design courses. The hope is that after knowing the results of the comparison of these formulas, the researcher can choose the highest reliability coefficient to estimate the results of the graphic design test. The higher the coefficient, the smaller the measurement error rate (Khumaedi, 2012). The formulation of the problems in this study is as follows: 1) determine the reliability coefficient of the Spearman-Brown formula Odd-Even, 2) Determine the reliability coefficient of the Spearman-Brown First-Last formula, 3) Determine the reliability coefficient of the Flanagan Odd-Even formula. 4) Determine the reliability coefficient of the Flanagan first-last formula, 5) determine the reliability coefficient of the Rulon Odd-Even formula; 6) determine the reliability coefficient of the Rulon first-last formula; 7) determine the reliability coefficient of the Kuder Richardson formula (KR) 20. 8) Determine the reliability coefficient of the Kuder-Richardson formula (KR) 21. 9) determine the reliability coefficient of the Hoyt formula, and 10) determine the reliability coefficient of the Alpha-Cronbach formula. 11) Compare all the reliability coefficients (the reliability coefficient of the Spearman-Brown formula). Odd-Even, the reliability coefficient of the Spearman-Brown First-Last formula, the reliability coefficient of the Flanagan Odd-Even formula, a reliability coefficient of the First-Last Flanagan formula, the reliability coefficient of the Odd-Even Rulon formula, reliability coefficient of the First-Last Rulon formula, reliability coefficient of the Kuder Richardson formula (KR), a reliability coefficient of the Kuder Richardson formula (KR), a reliability coefficient of the Hoyt formula, reliability coefficient Cronbach Alpha formula) with the UAS in the graphic media development course, and 12) determine the correct reliability coefficient value for the UAS in the graphic media course.

2 Methods

This study examines comparisons by comparing reliability calculations using the following formulas: 1) Spearman-Brown Odd-Even; 2) First-Last Spearman-Brown formula; 3) Odd-Even Flanagan formula; 4) First-Last Flanagan formula; 5) Rulon Odd-Even formula; 6) First-Last Rulon formula; 7) Kuder Richardson (KR) formula 20; 8) Kuder Richardson (KR) formula 21; 9) Hoyt's formula; and 10) Cronbach's Alpha formula. Data analysis using Microsoft Excel The data were obtained from the Semester Final Examination (UAS) test instrument for graphic media development courses in the form of responses from students of the Curriculum and Educational Technology Department. The subjects of this study were 90 students from the 2021–2022 Education Technology and Curriculum Department. The following will explain the calculation of the reliability coefficient using the split-half method or the halved method, KR-20, KR-21, Hoyt, and Alpha Cronbach.

To estimate reliability with the split-half method or the halved method, namely Spearman-Brown, Flanagan, and Rulon (Retnawati, 2017). Determination of the reliability of learning outcomes tests in the form of objective tests is carried out by "halving," or the Split-Half Technique, or the Single Test, Single Trial item, namely the item number odd-even and the first-last (Jago, 2019). Calculations using the odd-even technique are grouping all the odd-numbered items into one group and give the name "odd" group, while all even-numbered items are grouped into one group and named the "even" group (Haq, 2022). The odd group consists of items numbered 1, 3, 5, 7, and so on, while the even group consists of items numbered 2, 4, 6, 8, and so on. Calculations with the first-last technique involve grouping all the items with the initial number into one group and giving it the name "first" group, while all the items with the final number are grouped into one group and given the name "last" group. The first group consists of items numbered 1, 2, 3, 4, 5, and so on, while the last even group consists of item numbers 6, 7, 8, 9, 10, and so on (depending on the number of items). It should be noted that the Spearman-Brown odd-even formula and the first-last Spearman-Brown formula have the same calculation formula, the odd-even Flanagan formula and the first-last Flanagan formula have the same calculation formula, as well as the odd-even Rulon formula and the first Rulon formula. The difference in calculations lies in odd-even and first-last data.

2.1 Spearman- Brown's Formula

Spearman Brown's reliability test uses the split-half technique. Odd-even groups and first-last groups use the same formula, namely the product moment, as follows:

$$r_{xy} = \frac{n.\sum XY - \sum X.\sum Y}{\sqrt{[(n.(\sum X^2) - (\sum X)^2][(n.(\sum Y^2) - (\sum Y)^2]}} \quad (1)$$

Information

r_{xy} = correlation coefficient between x and y

n = Number of test takers

$\sum XY$= total XY multiplication (Sum of X and Y multiplication results)

$\sum X$ = total score X

$\sum Y$ = sum of Y scores

$\sum X^2$ = sum of the squares of X's score

$\sum Y^2$ = sum of the squares of Y scores

Source: (Arosyadi & Suyantiningsih, 2020).

After the correlation value is found, the Spearman-Brown formula is used (Yusup, 2018).

$$r_{11} = \frac{2(r_{xy})}{(1 + r_{xy})} \quad (2)$$

Information

r_{11}= reliability test coefficient

r_{xy}= result of *Product Moment*

2.2 Flanagan's Formula

Flanagan's reliability test uses the split-half technique (odd-even groups and first-last groups) using the same formula, as follows:

$$r_{11} = 2\left(1 - \frac{s_1^2 + s_2^2}{s_t^2}\right) \quad (3)$$

Information

r_{11} = Test reliability coefficient

s_1^2= odd item variant
s_2^2= even variant of the item
s_t^2= variant total

2.3 Rulon's Formula

Rulon reliability test using the split-half technique odd-even groups and first-last groups use the same formula, as follows (Azwar, 2018):

$$r_{11} = 1 - \frac{s_d^2}{s_t^2} \quad (4)$$

Information
r_{11} = coefficient of reliability test
s_d^2 = variance difference between split half I and II
s_t^2 = variance total
d = difference, the difference between the score split-half I and II

2.4 KR-20 Formula

Kuder Richardson's (KR) reliability test uses single tes – single trial technique. The formula of KR that a commonly used are KR-20 and KR-21 (Yusup, 2018). Those two KR formulas have specific instrument criteria for using the formula. When the instrument obtained different levels of difficulty or heterogenous for each item the formula of KR-20 is preferably used to test the reliability (Fraenkel, Wallen, & Hyun, 2012).

$$r_{11} = \left[\frac{n}{n-1}\right]\left[\frac{S_1^2 - \sum pq}{S_t^2}\right] \quad (5)$$

Information
r_{11} = test reliability coefficient
n = number of items
S_t^2 = total variance
p = the proportion of test takers who answered correctly
q = proportion of test takers who answered incorrectly (q = 1 - p)
∑pq = sum of the multiplication results of p and q

2.5 KR-21 Formula

If the instrument has the same level of difficulty or is homogeneous for each item, then use the KR-21 formula to test its reliability (Yusup, 2018) The KR-21 formula is not suitable for use in tests where the level of difficulty of the items is not homogeneous or the variance of the items is unequal (Azwar, 2018).

$$r_{11} = \left[\frac{n}{n-1}\right]\left[1 - \frac{M_t\,(n-M_t)}{n.S_t^2}\right] \quad (6)$$

Information
r_{11} = reliability coefficient

n = number of items
M_t = mean/average score
S_t^2 = total variance

2.6 Hoyt Formula

Formula C. Hoyt analyzes the scores of learning achievement test items using the analysis of variance or ANOVA techniques (Magdalena, Fauziah, et al., 2021; Sudijono, 2016). The Hyot formula for estimating reliability uses the following formula:

$$r_{11} = 1 - \frac{Vs}{Vr} \tag{7}$$

Information
r_{11} = test reliability coefficient
V_r = Respondent variant
V_s = Residual variance

2.7 Alfa Cronbach's Formula

Cronbach's alpha formula can be used to test essays, questionnaires, or questionnaires (Yusup, 2018). Cronbach's Alpha Formula to estimate reliability uses the following formula:

$$r_{11} = \left[\frac{n}{n-1}\right]\left[1 - \frac{\sum \quad s_i^2}{s_t^2}\right] \text{(Azwar, 2018).} \tag{8}$$

Information
r_{11} = instrument reliability coefficient
n = number of items in the test
1 = constant number
$\sum \quad s_i^2$= number of variances in the score of each item
s_t^2= variant total variance

The interpretation of the reliability coefficient of the test (r11) can be done using the reliability categorization (Maulida & Hamama, 2021a) as follows:

Table 1. Reliability Categorization

Test Reliability Coefficient	Categorization
0.81 < 1.00	Very high reliability
0.61 < 0.80	High reliability
0.41 < 0.60	Moderate reliability
0.21 < 0.40	Low reliability
-1.00 <0.20	Very low reliability (not reliable

In general, the reliability coefficient of a test is described numerically with a range between $-1.00 \leq \rho \leq +1.00$ (Retnawati, 2017). The test reliability coefficient can be said to be good (the category is adjusted according to Table 1 above) if it has a coefficient above 0.70 (Alfiatunnisa et al., 2022). A high coefficient has high reliability, and vice versa, a low coefficient has low reliability. The higher the coefficient, the smaller the measurement error; otherwise, the lower the coefficient, the greater the measurement error (Retnawati, 2017). The explanation above can give the lecturer an idea of the need to compare the accuracy of estimates between reliability formulas so that the lecturer can determine the right reliability formula to use for measuring a test. The teacher will choose a high-reliability coefficient for the test gauge.

3 Findings and Discussion

3.1 Spearman-Brown Reliability Test of Odd and Even Items

The computational results using the Spearman-Brown formula for odd-even yield the reliability coefficient as follows:

3.1.1 Spearman-Brown with Product Moment's Formula

$$r_{xy} = \frac{n.\sum XY - \sum X.\sum Y}{\sqrt{[(N.(\sum X^2) - (\sum X)^2][(N.(\sum Y^2) - (\sum Y)^2]}}$$

Calculation results obtained: odd (∑X) = 784, even (∑Y) = 834, ∑XY = 7631, ∑X2 = 7208, ∑Y2 = 8302, with N = 90

$$r_{xy} = \frac{90\ x\ 7631 - 784\ x\ 834}{\sqrt{[(90x(7280) - (784)^2][(90\ x(8302) - (834)^2]}}$$

$r_{xy} = 0{,}71988$

Spearman Brown's Formula calculation

$$r_{11} = \frac{2(r_{xy})}{(1 + r_{xy})}$$

$$r_{11} = \frac{2(0{,}71988)}{(1 + 0{,}71988)}$$

$$r_{11} = 0{,}8371$$

3.1.2 Fist-Last Spearman-Brown Reliability Test

Computational results using the first-last Spearman-Brown formula produce the reliability coefficient as follows:

Calculation results obtained: Awal (∑X) = 884, Akhir (∑Y) = 734, ∑XY = 7551, ∑X2 = 9090, ∑Y2 = 6580, with N = 90

$$r_{xy} = \frac{90\ x\ 7551 - 884\ x\ 734}{\sqrt{[(90x(9090) - (884)^2][(90\ x(6580) - (734)^2]}}$$

$r_{xy} = 0{,}6963$

Spearman Brown's formula

$$r_{11} = \frac{2(0{,}6963)}{(1 + 0{,}6963)}$$

$$r_{11} = 0{,}8209$$

3.2 Odd-Even Flanagan Reliability Test

The computational results using the odd-even Flanagan formula produce the following reliability coefficients:

3.2.1 Flanagan's Formula

$$r_{11} = 2\left(1 - \frac{s_1^2 + s_2^2}{s_t^2}\right)$$

Calculation results obtained: odd (∑X) = 784, even (∑Y) = 834, ∑XY = 7631, ∑X2 = 7208, ∑Y2= 8302, ∑(X + Y) = 1618, ∑(X + Y)2 = 30772 with N = 90

Determining the Odd Variance

$$s_1^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}$$

$$s_1^2 = \frac{7208 - \frac{(784)^2}{90}}{90}$$

$$s_1^2 = 4{,}2054$$

Determining the Even Variance

$$s_2^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N}$$

$$s_2^2 = \frac{8302 - \frac{(834)^2}{90}}{90}$$

$$s_2^2 = 6{,}3733$$

Determining the Total Variance

$$s_t^2 = \frac{\sum (X+Y)^2 - \frac{(\sum (X+Y))^2}{N}}{N}$$

$$s_t^2 = 18{,}155$$

Enter the sum result into Flanagan's Formula

$$r_{11} = 2(1 - \frac{4{,}2054 + 6{,}3733}{18{,}710})$$

$r_{11} = 0{,}8692$

3.2.2 First-Last Flanagan Reliability Test

Computational results using the first-last Flanagan formula produce the reliability coefficient as follows:

Calculation results obtained: Ganjil (∑X) = 884, Genap (∑Y) = 734, ∑XY = 7551, ∑X2 = 9090, ∑Y2 = 6580, ∑(X + Y) = 1618, ∑(X + Y)2 = 30772 with N = 90

Determine First Variance

$$s_1^2 = \frac{407{,}16}{90}$$

$$s_1^2 = 4{,}524$$

Determining the Last Variance

$$s_2^2 = \frac{602{,}83}{90}$$
$$s_2^2 = 6{,}698$$

Determine the Total Variance

$$s_t^2 = \frac{1633{,}96}{90}$$
$$s_t^2 = 18{,}155$$

Enter the sum result into Flanagan's formula

$$r_{11} = 2(1 - \frac{4{,}524 + 6{,}698}{18{,}155})$$
$$r_{11} = 0{,}7638$$

3.3 Odd Even Rulon Reliability Test

The computational results using the first-last Rulon formula produce the following reliability coefficients:

3.3.1 Rulon’s Formula

$r_{11} = 1 - \frac{s_d^2}{s_t^2}$

Calculation results obtained: odd(∑X) = 784, even (∑Y) = 834, ∑d=-50, ∑d2=248 with N = 90

Determine the Total Variant

$$s_t^2 = \frac{\sum(X+Y)^2 - \frac{(\sum(X+Y))^2}{N}}{N}$$
$$s_t^2 = \frac{30772 - \frac{(1618)^2}{90}}{90}$$
$$s_t^2 = 18{,}155$$

The variance of difference between scores

$$s_d^2 = \frac{\sum d^2 - \frac{(\sum d)^2}{N}}{N}$$
$$s_d^2 = \frac{248 - \frac{(-50)^2}{90}}{90}$$
$$s_d^2 = 2{,}447$$

Plug it into the Rulon formula

$$r_{11} = 1 - \frac{2{,}447}{18{,}155}$$
$$r_{11} = 0{,}8653$$

3.3.2 Final Rulon Reliability Test

Computation results using the first-last Rulon formula produce the reliability coefficient as follows:

Calculation results obtained: odd (∑X) = 884, even (∑Y) = 734, ∑d=150, ∑d2=568 with N = 90

Determine the Total Variant

$$s_t^2 = \frac{30772 - \frac{(1618)^2}{90}}{90}$$

$$s_t^2 = 18{,}155$$

The variance of difference between scores

$$s_d^2 = \frac{568 - \frac{(150)^2}{90}}{90}$$

$s_d^2 = 3{,}5333$

Plug it into the Rulon formula

$$r_{11} = 1 - \frac{3{,}5333}{18{,}155}$$

$$r_{11} = 1 - 0{,}1946$$

$$r_{11} = 0{,}8054$$

3.4 KR-20 Reliability Test

The computational results using the KR-20 formula produce the reliability coefficient as follows:

KR-20 formula

$$r_{11} = \left[\frac{n}{n-1}\right]\left[\frac{S_1^2 - \sum pq}{S_1^2}\right]$$

Calculation results obtained $\sum$pq = 3,995, S_t^2 = 18,710 with n =24

$$r_{11} = \left[\frac{24}{24-1}\right]\left[\frac{18{,}71062 - 3{,}99555}{18{,}71062}\right]$$

$$r_{11} = 0{,}82065$$

3.5 KR-21 Reliability Test

The computational results using the KR-21 formula produce the reliability coefficient as follows:

KR-21 formula

$r_{11} = \left[\frac{n}{n-1}\right]\left[1 - \frac{M_t\,(n-M_t)}{n.S_t^2}\right]$

Calculation results obtained Mt = 17,9778, $\sum X_t^2$ = 1683,955, S_t^2 = 18,710 with n =24

$$r_{11} = \left[\frac{24}{24-1}\right]\left[1 - \frac{17{,}97778\,(24 - 17{,}97778)}{24\,x\,18{,}710}\right]$$

r_{11} = 0,79985 ≈ 0,8

3.6 Hoyt Reliability Test

The computational results using the KR 21 formula produce the reliability coefficient as follows:

Hoyt's formula

$$r_{11} = 1 - \frac{Vs}{Vr}$$

The calculation results

Table 2. ANAVA calculation

Source of Variance	Sum of Squares	db	Variance
Respondent ($JK_{(r)}$)	70,16482	89 (90-1)	$\frac{70,16482}{89} = 0,78837$
Item ($JK_{(i)}$)	46,39815	23 (24-1)	$\frac{46,39815}{23} = 2,01731$
Rest ($JK_{(s)}$)	289,43513	1934 (2046 - 89 – 23)	$\frac{289,43513}{1934} = 0,14966$
Total ($JK_{(t)}$)	405,9981	2046 (2047-1)	

$$r_{11} = 1 - \frac{0,14966}{0,78837}$$

$r_{11} = 0,8101$

The reliability coefficient of the test is 0.81 > 0.70, so the test instrument has high reliability (consistency with Table 1).

3.7 Alpha Cronbach Reliability Test

The results of processing using the Alpha Cronbach formula produce the reliability coefficient as follows:

Alpha Cronbach's formula is:

$$r_{11} = \left[\frac{n}{n-1}\right]\left[1 - \frac{\Sigma \quad s_i^2}{s_t^2}\right]$$

Calculation of the reliability coefficient using the SPSS application produces 0.821. To make it easier for researchers and readers to analyze the comparison (comparison) of the reliability coefficient, the researcher needs to make a recapitulation table. Recapitulation is presented in Table 3.

Table. 3. Reliability Coefficient Result

Reliability Type		**Result**	**Note**
Spearman-Brown	Odd- Even	$r_{11} = 0,837$	very high
	First - Last	$r_{11} = 0,820$	very high
Flanagan	Odd- Even	$r_{11} = 0,869$	very high
	First - Last	$r_{11} = 0,800$	very high
Rulon	Odd -Even	$r_{11} = 0,865$	very high
	First - Last	$r_{11} = 0,805$	very high
Kuder Richardson (KR)	20	$r_{11} = 0,820$	very high
	21	$r_{11} = 0,8$	very high
Hoyt		$r_{11} = 0,810$	very high
Alpha Cronbach		r11 = 0,821	very high

The result is consistent with the interpretation based on several results in Table 3.

The results of the comparison of the reliability coefficient of UAS scores for graphic media courses using various reliability test methods, as shown in Table 3, turned out to produce a reliability coefficient with an average above 0.80. This is in accordance with Retnawati's view, which states that the reliability coefficient range of a test numerically has a range between $-1.00 \leq \rho \leq +1.00$ (Retnawati, 2017). The reliability coefficient of the test can be said to be good if it has a coefficient above 0.70 (Alfiatunnisa et al., 2022; Yusup, 2018). While the reliability coefficient with a value of 0.81 <1.00 can be categorized as having very high reliability (Maulida & Hamama, 2021b).

The Spearman-Brown reliability coefficient odd-even is 0.837, so the test instrument has very high reliability. The first-last reliability coefficient is 0.820, so the test instrument has very high reliability. The odd-even Flanagan reliability coefficient is 0.8692, so the test instrument has a very high reliability. The first-last Flanagan reliability coefficient is 0.80, so the test instrument has very high reliability. The reliability coefficient of odd-even Rulon is 0.8653, so the test instrument has very high reliability. The first-last Rulon reliability coefficient is 0.8054, so the test instrument has very high reliability. The KR-20 reliability coefficient is 0.820, so the test instrument has very high reliability. The KR-21 reliability coefficient is 0.80, so the test instrument has very high reliability. The results of calculations using the KR-20 and KR-21 reliability formulas produce calculations that are not much different. The formula with KR-21 produces a slightly lower reliability coefficient than the KR-20 formula (Gunartha, 2022; Sarwiningsih, 2017). This is because the value of p or the proportion on KR-20 has a very varied (heterogeneous) level of difficulty among the items in the test concerned. In the KR-21 formula, the instruments have the same level of difficulty or are homogeneous (Gunartha, 2022). Hoyt's reliability coefficient is 0.81, so the test instrument has very high reliability. The Hoyt reliability coefficient of 0.810 is included in the very high category (Soleh, A, M. Khumaedi, Pramono, 2017). The Cronbach's alpha reliability coefficient is 0.821, so the test instrument has very high reliability. Alpha Cronbach is a mathematical formula used to test the level of reliability. An instrument can be said to be reliable if it has a coefficient of 0.6 or more (Zahra & Rina, 2018). The reliability coefficient of the test is 0.821, so the test instrument has very high reliability. Crobach's alpha value of more than 0.6 means that reliability is very good (Aprillia & Magdalena, 2018).

The results of the reliability test have an average value of 0.7 and above, which is classified as good, meaning that the test has high reliability (Yusup, 2018). Theoretically, the measurement results can be trusted if, in several measurements of the same group of subjects, relatively the same results are obtained, as long as the subject being measured has not changed (Syamsuddin, 2017). Test reliability is closely related to validity because a valid measuring instrument can be ascertained to be reliable, but a reliable measuring instrument is not necessarily valid (Sugiyono, 2019). Arikunto menyatakan penting bagi suatu tes memiliki persyaratan validitas dan reliabilitas, (Arikunto, 2018). Dalam tes mungkin reliabel, tetapi tidak valid. Sebaliknya suatu tes yang valid sudah pasti reliabel (Suci Mitra & Helendra, 2022). The quality of the items used to measure the ability of the test takers needs to be considered, including whether the items are valid and reliable; besides that, the items are said to be good if they are not too easy or too difficult. Items must be able to distinguish between test takers who are clever and who are not clever, and the effectiveness of the distractor must be functioning properly (Friatma & Anhar, 2019).

4 Conclusion

Spearman Brown's odd-even formula yields a reliability coefficient of 0.837, first-last Spearman Brown of 0.820. The Odd-Even Flanagan formula is 0.835, the first-last Flanagan formula is 0.8. The odd-even Rulon formula is 0.865, the first-last Rulon formula is 0.805, Kuder Richardson (KR-20) is 0.820, and KR-21 is 0.8. Hoyt's formula is 0.810, while Cronbach's Alpha is 0.821. When viewed from the average reliability coefficient of all formulas, it shows a figure above 0.7, meaning that all reliability coefficients are included in the very good category, which is above 0.7. The results of the comparison (comparison) of all the reliability results obtained the results of a reliability coefficient above 0.7, which means that all reliability meets the requirements and is suitable to be used as a measuring tool, but if using reliability with the Rulon odd-even formula, the results are higher, so it would be better to calculate the UAS eye test Graphic media development lectures use the Rulon odd-even formula.

References

1. Akbar, M. R., Mulyadi, M., & Shandi, S. A. (2021). Kajian Literatur Media Pembelajaran Grafis dalam Pembelajaran Bahasa. *Jurnal Pendidikan Bahasa*, *11*(2), 46–56. https://ejournal.tsb.ac.id/index.php/jpb/article/view/527
2. Alfiatunnisa, E., Khairunnisa, H. Z., Hayati, S., & Maulida, V. L. (2022). Uji Validitas dan Reliabilitas Terhadap Kemandirian Siswa Sekolah Dasar Kelas 1. *Jurnal Hurriah: Jurnal Evaluasi Pendidikan Dan Penelitian*, *3*(2), 29–36. https://academicareview.com/index.php/jh/article/view/81
3. Anastasi. Anne and Urbina, S. (2017). *Psichological Testing* (Seventh Ed). New Jersey: Prentice-Hall, Inc.
4. Aprillia, A., & Magdalena, N. (2018). Pengukuran Kualitas Layanan Dalam Bidang Pendidikan Tinggi. *Jurnal Manajemen Maranatha*, *18*(1), 11–22. https://doi.org/10.28932/jmm.v18i1.1092
5. Apsari, Y., & Acep Haryudin. (2017). The Analysis Of English Lecturers ' Classroom-Based Reading Assessments To Improve Students ' Reading Comprehension. *Journal Of English Language Teaching In Indonersia*, *5/1*, 35–44. http://e-journal.stkipsiliwangi.ac.id/index.php/eltin/article/view/384
6. Arfah, M. (2021). Evaluasi Hasil Belajar Pendidikan Agama Islam (PAI). *Manufacturing Chemist*, *88*(10), 48–49. https://doi.org/10.12968/nuwa.2018.17.30
7. Arifin, Z. (2017). Kriteria Instrumen dalam suatu Penelitian. *THEOREMS (The Original Research of Mathematics)*, *2*(1), 28–36.

8. Arikunto, S. (2018). *Dasar - Dasar Evaluasi Pendidikan*. Bumi Aksara.
9. Arosyadi, M. I., & Suyantiningsih, S. (2020). Korelasi antara persepsi pengelolaan dan layanan pustaka dengan motivasi belajar di digital library UNY. *Epistema*, *1*(1), 59–67. https://doi.org/10.21831/ep.v1i1.32489
10. Ayu, S., & Rosli, M. S. Bin. (2020). Uji reliabilitas instrumen penggunaan SPADA (Sistem Pembelajaran dalam Jaringan). *Biormatika*, *6*(1), 145–155. http://ejournal.unsub.ac.id/index.php/FKIP/
11. Azwar, S. (2016). *Konstruksi Tes Kemampuan Kognitif*. Pustaka Pelajar.
12. Azwar, S. (2018). Reliabilitas dan Validitas.
13. Bahri, M. F. (2019). Content Validity and Reliability Analisys of Integrated Islamic - Science Test Instrument to Measure The Student's Critical Thinking Ability. *ISLAM Realitas: Journal of Islamic & Social Studies*, *5*(1), 45–51. http://dx.doi.org/10.30983/islam_realitas.v5i1.894
14. Bashooir, K., & Supahar. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 168–181. https://doi.org/10.21831/pep.v22i2.20270
15. Budiantoro, T., Kurniawan, B., Negeri, P., Laut, T., Negeri, P., & Laut, T. (2021). *Validitas Dan Reliabilitas Instrumen Keterampilan Komunikasi Dan Keterampilan Kolaborasi Pada Mata Kuliah Bahasa Indonesia*. *7*. https://jht.politala.ac.id/index.php/jht/article/view/89
16. Dewi, S. K., & Sudaryanto, A. (2020). Validitas dan Reliabilitas Kuisioner Pengetahuan , Sikap dan Perilaku Pencegahan Demam Berdarah. *Prosiding Seminar Nasional Keperawatan Universitas Muhammadiyah Surakarta*, 73–79. https://publikasiilmiah.ums.ac.id/handle/11617/11916
17. Elviana. (2020). Analisis Butir Soal Evaluasi Pembelajaran Pendidikan Agama Islam Menggunakan Program Anates. *Jurnal Mudarrisuna*, *10*(2), 58–74. http://dx.doi.org/10.22373/jm.v10i2.7839
18. Erawati, N. K. (2018). Analisis Tes Penilaian Pencapaian Kompetensi Pada Mahasiswa Kebidanan. *Jurnal Penjakora*, *5*(2), 111–120. https://ejournal.undiksha.ac.id/index.php/PENJAKORA/article/view/17287/10378
19. Erfan, M., Maulyda, M. A., Ermiana, I., Hidayati, V. R., & Widodo, A. (2020). Validity and Reliability of Cognitive Tests Study and Development of Elementary Curriculum Using Rasch Model. *Psychology, Evaluation, and Technology in Educational Research*, *3*(1), 26–33. https://doi.org/10.33292/petier.v3i1.51
20. Farida, & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*, *1*(1), 34–44. https://jurnal.lp2msasbabel.ac.id/index.php/AL-MUARRIB
21. Fraenkel, J. L., Wallen, N. E., & Hyun, H. H. (2012). *How to Design and Evaluate Research in Education* (eighth edi). New York : Mc Graw Hill.
22. Friatma, & Anhar. (2019). Analysis of validity, reliability, discrimination, difficulty and distraction effectiveness in learning assessment. *Journal of Physics: Conference Series*, *1387*(1). https://doi.org/10.1088/1742-6596/1387/1/012063
23. Gunartha, I. W. (2022). Estimasi Kesalahan Pengukuran Dalam Bidang Pendidikan Berdasarkan Teori Tes Klasik. *Widyadari*, *23*(1), 34–47. https://doi.org/10.5281/zenodo.6390889
24. Halimatus, H., Fridani, L., & Meilani, S. M. (2019). Pengembangan Media Grafis untuk Pengenalan Life Science pada Anak Usia Dini. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *4*(1), 395. https://doi.org/10.31004/obsesi.v4i1.318
25. Hanifah, N. (2014). Perbandingan Tingkat Kesukaran, Daya Pembeda Butir Soal Dan Reliabilitas Tes Bentuk Pilihan Ganda Biasa Dan Pilihan Ganda Asosiasi Mata Pelajaran

Ekonomi. *Jurnal SOSIO E-KONS*, *6*(1), 41–55. https://journal.lppmunindra.ac.id/index.php/sosio_ekons/article/view/1715

26. Haq, V. A. (2022). Menguji Validitas dan Reliabilitas pada Mata Pelajaran Al Qur'an Hadits Menggunakan Korelasi Produk Momenspearman Brown. *Jurnal Studi Islam*, *04*(01), 11–24. https://jurnal.staiannawawi.com/index.php/annawa/article/view/419
27. Hikmah, & Muslimah. (2021). Validitas dan Reliabilitas Tes dalam Menunjang Hasil Belajar PAI. *Proceedings*, *1*, 345–356. https://e-proceedings.iain-palangkaraya.ac.id/index.php/PICIS/article/view/520
28. Huda, U., Haryanto, T., & Haryanto, B. S. (2018). Strategi Penanggulangan Radikalisme Di Perguruan Tinggi Kabupaten Banyumas. *An-Nidzam*, *5*(1), 541–552.
29. Idrus. (2019). Evaluasi Dalam Proses Pembelajaran. *Evaluasi Dalam Proses Pembelajaran*, *2*, 920–935. http://dx.doi.org/10.35673/ajmpi.v9i2.427
30. Isnaeni, F., Widagdo, S., & Supardi, S. (2019). Quality Reliability Service Towards Student Satisfaction. *International Journal of Social Science and Business*, *3*(4), 386. https://doi.org/10.23887/ijssb.v3i4.21401
31. Jago, N. (2019). Uji Reliabilitas Instrumen Tingkat Kegemaran Membaca Masyarakat Umum di Provinsi Bali Tahun 2018 dengan Teknik Belah Dua (Split Half). *Acarya Pustaka*, *6*(1), 1. https://doi.org/10.23887/ap.v6i1.20766
32. Khumaedi, M. (2012). The Reliability of Education Research Instrumens. *JPTM: Jurnal Pendidikan Teknik Mesin*, *12*(1), 25–30. http://jurnal.fkip.uns.ac.id
33. Kurniawan, U., & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, *6*(1), 34–39. https://jurnal.unimus.ac.id/index.php/statistik/article/view/3421
34. Kurniawati, A. (2019). Analisis Hasil Tes Evaluasi Pendidikan Pada Mahasiswa Ditinjau Dari Perbedaan Gender. *Jurnal Ilmiah DIDAKTIKA*, *19*(1), 89–106. https://jurnal.ar-raniry.ac.id/index.php/didaktika/article/view/4196
35. Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas III SDN Karet 1 Sepatan. *BINTANG : Jurnal Pendidikan Dan Sains*, *3*(2), 198–214. https://doi.org/10.36088/bintang.v3i2.1291
36. Magdalena, I., Syariah, E. N., Mahromiyati, M., & Nurkamilah, S. (2021). Analisis Instrumen Tes Sebagai Alat Evaluasi Pada Mata Pelajaran SBDP Siswa Kelas II SDN Duri Kosambi 06 Pagi. *Jurnal Pendidikan Dan Ilmu Sosial*, *3*, 276–287. http://jurnal.fkip.unila.ac.id/index.php/JPPPI/article/view/22206
37. Manshur, A., & Rodhi, A. (2020a). Pengembangan Media Grafis Dalam Pembelajaran. *Al-Aufa: Jurnal Pendidikan Dan Kajian Keislaman*, *2*(2), 1–13. https://doi.org/10.36840/alaufa.v2i2.313
38. Manshur, A., & Rodhi, A. (2020b). Pengembangan Media Grafis Dalam Pembelajaran. *Al-Aufa: Jurnal Pendidikan Dan Kajian Keislaman*, *2*(2), 1–13. https://doi.org/10.36840/alaufa.v2i2.313
39. Mardapi, D., & Kertowagiran, B. (2011). Instrumens for Student Achievements. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *2*(4), 326 – 341.
40. Maulida, & Hamama, S. F. (2021a). Pengembangan Instrumen Tes Tipe Pilihan Ganda Dalam Evaluasi Hasil Belajar Siswa Pada Konsep Sel Tingkat Sekolah Menengah Atas. *Jurnal Dedikasi Pendidikan*, *8848*(1), 171–178.
41. Maulida, & Hamama, S. F. (2021b). Pengembangan Instrumen Tes Tipe Pilihan Ganda Dalam Evaluasi Hasil Belajar Siswa Pada Konsep Sel Tingkat Sekolah Menengah Atas. *Jurnal Dedikasi Pendidikan*, *5*(1), 171–178. http://jurnal.abulyatama.ac.id/index.php/dedikasi/article/downloadSuppFile/1498/209

42. Murti, Wiyanto, & Hartono. (2018). Studi Komparasi antara Tes Testlet dan Uraian dalam Mengukur Hasil Belajar Kognitif Siswa Kelas XI SMA Negeri 1 Gombong. *Unnes Physics Education Journal*, *7*(1), 23–31. http://journal.unnes.ac.id/sju/index.php/upej
43. Ndiung, S., & Jediut, M. (2020). Pengembangan Instrumen Tes Hasil Belajar Matematika Peserta Didik Sekolah Dasar Berorientasi Pada Berpikir Tingkat Tinggi. *Premiere Educandum : Jurnal Pendidikan Dasar Dan Pembelajaran*, *10*(1), 94. https://doi.org/10.25273/pe.v10i1.6274
44. Nengsi, A. R., & Efrina, G. (2019). Optimasi validitas dan reliabilitas tes pilihan ganda buatan guru mata pelajaran ips sd. *Journal Innovation in Islamic Education: Challenges and Readiness in Society 5.0, 4th International Conference on Education*, 43–48. https://ojs.iainbatusangkar.ac.id/ojs/index.php/proceedings/article/view/2138
45. Neti. (2020). *Akhir Semester Genap Pelajaran Fisika Kelas Xi MIPA 1 SMA Negeri 1 Kepahiang*. *14*(2), 89–98. https://ojs.stkippgri-lubuklinggau.ac.id/index.php/JPP/article/view/1049/523
46. Nusantari, D. O. (2016). Meningkatkan Estimasi Reliabilitas Instrumen Pilihan Ganda dengan Memperhatikan Probabilitas Guessing. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, *6*(2), 100–109. https://doi.org/10.30998/formatif.v6i2.943
47. Pangalila, T. (2017). Interaksi Sosial Dosen Dan Mahasiswa Dalam Proses Perkuliahan Di Jurusan PPKN FIS UNIMA. *PKn Progresif*, *12*(2), 699–706. https://jurnal.fkip.uns.ac.id/index.php/progresif/article/view/11274
48. Purniasari, L., Masykuri, M., & Ariani, S. R. D. (2021). Analisis Butir Soal Ujian Sekolah Mata Pelajaran Kimia SMA N 1 Kutowinangun Tahun Pelajaran 2019/2022 Menggunakan Model Iteman dan Rasch. *Jurnal Pendidikan Kimia*, *10*(2), 205–214. https://jurnal.uns.ac.id/JPKim/article/view/48244
49. Purwani, A., Fridani, L., & Fahrurrozi, F. (2019). Pengembangan Media Grafis untuk Meningkatkan Siaga Bencana Banjir. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, *3*(1), 55. https://doi.org/10.31004/obsesi.v3i1.142
50. Puspasari, H., & Puspita, W. (2022). Uji Validitas dan Reliabilitas Instrumen Penelitian Tingkat Pengetahuan dan Sikap Mahasiswa terhadap Pemilihan Suplemen Kesehatan dalam Menghadapi Covid-19 Validity Test and Reliability Instrument Research Level Knowledge and Attitude of Students Towards. *Jurnal Kesehatan*, *13*, 65–71. http://www.ejurnal.poltekkes-tjk.ac.id/index.php/JK/article/view/2814
51. Putri, D., & Nahadi. (2019). Perbandingan Reliabilitas Tes Hasil Belajar Matematika Sma Berdasarkan Teknik Penskoran Dan Ukuran Sampel. *Journal Education and Chemistry (JEDCHEM)*, *1*(1), 10–24.
52. Putri, R. Z., & Sari, R. (2020). Pengembangan dan Validasi Instrumen Tes untuk Mengukur Keterampilan Menyelesaikan Masalah Peserta Didik SMA pada Pelajaran Fisika. *Jurnal Penelitian Pembelajaran Fisika*, *11*(1), 17–25. https://doi.org/10.26877/jp2f.v11i1.3993
53. Retnawati, H. (2017). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin Unnes*, *12*(1), 129541.
54. Sanaky, M. M. (2021). Analisis Faktor-Faktor Keterlambatan Pada Proyek Pembangunan Gedung Asrama Man 1 Tulehu Maluku Tengah. *Jurnal Simetrik*, *11*(1), 432. https://doi.org/10.31959/js.v11i1.615
55. Sarwiningsih, R. (2017). The Comparison Accuracy Estimation of Test Reliability Coefficients for National Chemistry Examination in Jambi Province on Academic Year 2014/2015. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *2*(1), 34. https://doi.org/10.20961/jkpk.v2i1.8740
56. Soleh, A, M. Khumaedi, Pramono, E. (2017). Pengembangan Instrumen Penilaian Mata Pelajaran PKn Standar Kompetensi Memahami Kedaulatan Rakyat Dalam Sistem

Pemerintahan di Indonesia. *Journal of Research and Educational Research Evaluation*, *6*(1), 71–80. https://journal.unnes.ac.id/sju/index.php/jere/article/view/16210
57. Subhi, M., Lubis, K., Harahap, H. S., Al, U., Medan, W., & Belajar, H. (2022). Penggunaan Media Grafis Untuk Meningkatkan Hasil Akidah Akhlak. *Jurnal Pendidikan Agama Islam*, *1*(1), 73–82. https://ejurnal.univamedan.ac.id/index.php/Tajribiyah/article/view/209
58. Suci Mitra, P., & Helendra. (2022). Analisis Kualitas Butir Soal Ujian Akhir Semester Ganjil Tahun Pelajaran 2020 / 2021 Mata Pelajaran Biologi Kelas X SMA Negeri 1 Teluk Sebong Analysis of Quality The Question Final Exam Odd Semester 2020 / 2021 Biology Class X SMA Negeri 1 Teluk Sebong. *Biodidaktika: Jurnal Biologi Dan Pembelajarannya*, *17*(2). http://dx.doi.org/10.30870/biodidaktika.v17i2.16493
59. Sudiapermana, E., & Setiawan, B. (2022). Validity and reliability of measurement instruments of educative family life. *Cypriot Journal of Educational Sciences*, *17*(5), 1727–1741. https://doi.org/10.18844/cjes.v17i5.7275
60. Sudijono, A. (2016). *Pengantar Evaluasi Pendidikan*. PT RajaGrafindo.
61. Sugianto, A. (2016). Journal on English as a Foreign Language An Analysis of English National Final Examination for Junior High School in Terms of Validity and Reliability. *Journal on English as a Foreign Language*, *6*, 31–42. https://e-journal.hamzanwadi.ac.id/index.php/veles/article/view/1551
62. Sugiono, Noerdjanah, & Wahyu, A. (2020). Uji Validitas dan Reliabilitas alat Ukur SG Posture Evaluation. *Jurnal Keterapian Fisik*, *5 no 1*, 55–61. https://jurnalketerapian-fisik.com/index.php/jpt/article/view/167
63. Sugiyono. (2019). *Metode Penelitian Pendidikan*. CV. Alfabeta.
64. Sulistianingsih. (2020). Pengetahuan Guru tentang Konstruksi Tes , Penguasaan Materi Pelajaran Sains dengan Reliabilitas Tes Buatan Guru. *Jurnal Ilmu Pendidikan (JIP) STKIP Kusuma Negara*, *11*(2), 145–153. https://jurnal.stkipkusumanegara.ac.id/index.php/jip/article/view/120
65. Supriyadi, M.Pd, E. (2017). Pengembangan Model Evaluasi Untuk Meningkatkan Mutu Pendidikan Teknik Elektro. *Jurnal Edukasi Elektro*, *1*(1), 25–35. https://doi.org/10.21831/jee.v1i1.13254
66. Surucu, L., & Maslakci, A. (2020). Validity and Reliability in Quantitative Research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726.
67. Suwartono, C., Amiseso, C. P., & Handoyo, R. T. (2017). Uji Reliabilitas Dan Validitas Eksternal The Raven ' S Standard Progressive Matrices. *Jurnal HUMANITA*, *14*(1), 1–9. http://journal.uad.ac.id/index.php/HUMANITAS/article/view/5772
68. Syahfitri, J., Firman, H., Redjeki, S., & Sriyati, S. (2018). Validitas Dan Reliabilitas Tes Disposisi Berpikir Kritis Dalam Biologi Perguruan Tinggi. *Seminar Nasional Edusainstek FMIPA UNIMUS 2018*, *2014*, 82–86. https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/4109
69. Syamsuddin, M. (2017). Pengembangan Instrumen Hasil Belajar Fisika Peserta Didik Kelas XI IPA SMA Negeri 1 Donri-Donri Kabupaten Soppeng. *PEP Educational Assessment*, *1*(1), 43–50. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=_kWt4kUAAAAJ&pagesize=100&citation_for_view=_kWt4kUAAAAJ:9yKSN-GCB0IC
70. Tri Jampi Setiyorini, Zyah Rochmad Jaelani, A. N. (2022). Analisis Faktor-Faktor Yang Mempengaruhi Reliabilitas. *Didaktis: Jurnal Pendidikan Dan Ilmu Pengetahuan*, *22*(3). http://journal.um-surabaya.ac.id/index.php/didaktis/article/viewFile/11286/5537
71. Ulfah, A. A., Kartono, K., & Susilaningsih, E. (2020). *Validity of Content and Reliability of Inter-Rater Instruments Assessing Ability of Problem Solving*. *9*(1), 1–7. https://journal.unnes.ac.id/sju/index.php/jere/article/view/40423

72. Wenno, I. H., Tuhurima, D., & Manoppo, Y. (2021). How to Create a Good Test. *Jurnal Pendidikan Profesi Guru Indonesia*, *1*(1), 11–20. https://ojs3.unpatti.ac.id/index.php/jppgi2019/article/view/3093
73. Widhiarso, W., & Mardapi, D. (2010). The Comparison Of Reliability Coefficients Estimation Among Classical Test Theory. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *14*(1), 1–16. https://doi.org/10.21831/pep.v14i1.1973
74. Yulia Pramusinta. (2017). Penggunaan Media Grafis (Peta Konsep) Dalam Pembelajaran Sejarah Untuk Mempermudah Pemahaman Siswa Kelas V MI Darul Ulum Bojonegoro. *Jurnal Pendidikan Guru Madrasah Ibtidaiyah Volume*, *1*(3), 248–253. https://jurnalfai.unisla.ac.id/index.php/at-thulab/article/view/77%0A%0A%0A
75. Yusup, F. (2018). Uji Validitas Dan Reliabilitas Instrumen Penelitian Kuantutatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, *7*(1), 17–23. https://doi.org/10.21831/jorpres.v13i1.12884
76. Zahra, R. R., & Rina, N. (2018). Pengaruh Celebrity Endorser Hamidah Rachmayanti Terhadap Keputusan Pembelian Produk Online Shop Mayoutfit Di Kota Bandung. 6(5), 613–616. https://doi.org/10.7868/s0869565218050249
77. Zainal Arifin. (2019). Evaluasi Pembelajaran, Prinsip, Teknik, dan Prosedur. PT Remaja Rosdakarya.
78. Zainal, N. F. (2020). Pengukuran, Assessment dan Evaluasi dalam Pembelajaran Matematika. *Laplace : Jurnal Pendidikan Matematika*, *3*(1), 8–26. https://doi.org/10.31537/laplace.v3i1.310