Comparative Test of Cronbach's Alpha Reliability Coefficient, Kr-20, Kr-21, And Split-Half Method

Hari Sugiharto Setyaedhi[1*] ID

[1] Universitas Negeri Surabaya, Surabaya, Indonesia

ABSTRAK

Tes ujian akhir semester mempunyai peran yang penting dalam mengukur hasil belajar mahasiswa untuk itu tes harus mempunyai mutu yang baik. Permasalahannya adalah dosen tidak menggunakan alat ukur yang dapat diandalkan untuk mengukur hasil pembelajaran. Penelitian ini bertujuan untuk membandingkan berbagai koefisien reliabilitas sehingga diketahui metode reliabilitas manakah yang paling tinggi. Metode pengumpulan data dalam penelitian ini adalah observasi, wawancara, dan tes yang didasarkan pada respons mahasiswa terhadap ujian akhir semester. Teknik analisis data menggunakan reliabilitas rumus Alpha Cronbach, Kuder Richardson 20, 21, dan metode belah dua (split half method) yang terdiri dari Spearman Brown, Flanagan, Rulon, Hoyt. Jenis penelitian ini menggunakan deskriptif kuantitatif. Desain penelitian komparatif bersifat ex post facto. Subjek penelitian adalah mahasiswa S1 sebanyak 90 mahasiswa. Hasil penelitian memberikan beberapa kesimpulan, pertama, metode Rulon ganjil genap mempunyai koefisien reliabilitas paling tinggi yaitu 0,86, kedua, seluruh metode reliabel mempunyai koefisien diatas 0,7 sebagai syarat minimum penerimaan, dan ketiga. seluruh metode reliabilitas mempunyai koefisien reliabilitas yang relatif sama yaitu berkisar 0,8. keempat, semua metode reliabel cocok digunakan sebagai alat ukur dalam ujian akhir semester. Implikasi dari penelitian ini adalah dosen disarankan untuk menggunakan koefisien reliabilitas Rulon ganjil-genap ketika mengukur hasil pembelajaran.

ABSTRACT

Final semester exams have an essential role in measuring student learning outcomes. Therefore, the tests must be of good quality. The problem is that lecturers must use reliable measuring tools to measure learning outcomes. This research aims to compare various reliability coefficients to determine which method has the highest reliability. The data collection methods in this research are observation, interviews, and tests based on student responses to the final semester exams. The data analysis technique uses the reliability of the Alpha Cronbach formula, Kuder Richardson 20, 21, and the split-half method consisting of Spearman-Brown, Flanagan, Rulon, and Hoyt. This type of research uses quantitative description. Comparative research designs are ex post facto. The research subjects were 90 undergraduate students. The research results provide several conclusions: first, the odd-even Rulon method has the highest reliability coefficient, namely 0.86; second, all reliable methods have a coefficient above 0.7 as the minimum requirement for acceptance; and third. All reliability methods have relatively the same reliability coefficient, namely around 0.8. fourth, all reliable techniques are suitable for measuring tools in the final semester exam. This research implies that lecturers are advised to use the odd-even Rulon reliability coefficient when measuring learning outcomes.

1. INTRODUCTION

The success of the learning process can be measured through student learning outcomes. Student learning outcomes come from the assessment process carried out by the course lecturer (Indriasari et al., 2021). At the end of every semester, universities conduct end-of-semester assessments, or what are usually called final semester examinations (UAS) or final tests (Subarkah et al., 2020; Suradi, 2022). The aim of implementing the UAS is to test student learning outcomes based on what they have learned during one semester (Kadariah, 2018; Purba, 2021; Supriyadi, 2017; Suradi, 2022). Before doing the UAS, the lecturer makes a test instrument. Test instruments are used to measure learning outcomes (Magdalena, Syariah, et al., 2021). Learning outcome tests are used as an assessment tool to make appropriate decisions about whether the student has achieved the learning objectives (Tri Jampi Setiyorini, Zyah Rochmad Jaelani, 2022; Ulfah et al., 2020). This can be understood because tests are 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). Tests also function as a tool to inform students

Corresponding author.
E-mail addresses: harisetyaedhi@unesa.ac.id (Hari Sugiharto Setyaedhi)

about their mastery of the learning material (Wenno et al., 2021). Tests are a planned effort carried out by lecturers to show learning results to students (Kurniawati, 2019). Test instruments as measuring tools must meet the requirements, namely, they must be valid and reliable (Arifin, 2017; Çakir, 2022; Lia et al., 2020). What is worth paying attention to in a test instrument is that it meets the demands of validity and reliability, namely the accuracy of the measurement results and the consistency of the measurement results (Krieglstein et al., 2022; Nengsi & Efrina, 2019). An instrument (a measuring instrument) is said to be of good quality if it has been tested for validity and reliability (Dewi & Sudaryanto, 2020; Puspasari & Puspita, 2022). A test instrument is said to be good if it has evidence of validity and reliability and is suitable for use (Alfiatunnisa et al., 2022; Bashooir & Supahar, 2018; Budiantoro & Kurniawan, 2021).

A reliable test is one that will produce relatively the same data even though it is used by different people, times, and places with the same research objectives or research subjects (Lestari & Yudhanegara, 2017). Reliability is the accuracy or stability of the assessment results (Lia et al., 2020; Surucu & Maslakci, 2020). The reliability of a test aims to measure the consistency, precision, repeatability, trustworthiness, etc. of the test (Ajayi, 2013; Nath, 2013; Sarwiningsih, 2017). Reliability is needed to determine whether the measuring instrument, in the form of a test, is reliable or consistent over time (Arfah, 2021; Bahri, 2019; Erfan et al., 2020; Sanaky, 2021; Syahfitri et al., 2018). It is hoped that test instruments that have high reliability can serve as a guide or reference for measuring student abilities (Indriasari et al., 2021; Sarwiningsih, 2017). Test instruments as quality measuring tools can produce reliable or trustworthy information (Danni & Tauratiya, 2020). A quality instrument will produce accurate information so that measurement errors can be minimized (Faridah, 2021; Retnawati & Hadi, 2014). Quality instruments have good validity and reliability. This explanation proves that validity and reliability are absolute requirements that must be possessed by quality instruments, including sets of questions made by educators or lecturers (Danni & Tauratiya, 2020; Faridah, 2021). The reliability of assessments after teaching interventions is important to be carried out continuously (Lockhart, 2015). Reliability in quantitative research is very important (Campbell et al., 2013). Generally, before and after data collection, researchers need to consider the reliability of the data. Therefore, various methods are used to increase the validity and reliability of the data (Zohrabi, 2013). There are various types of reliability measurement methods, such as Cronbach's alpha, KR-20, and KR-21, and split-half methods such as the Spearman-Brown formula, Flanagan formula, Rulon formula, and Hoyt formula (Ajayi, 2013; Ekolu & Quainoo, 2019; Gunartha, 2022; Puspasari & Puspita, 2022).

Test reliability can be influenced by several factors, such as the characteristics of test takers, test conditions, variations in test administration, as well as errors and differences in scoring, the length of the test, the homogeneity of student abilities, and the level of difficulty of test items (Busnawir dalam (Putri & Nahadi, 2019). The research results state that test reliability is closely related to how the test is presented, the mood of the test taker, the test taker's attitude when facing the test, motivation, the condition of the test room, and so on (Putri & Nahadi, 2019). The procedures that must be carried out in developing a reliable test instrument are: 1) compiling test specifications; 2) writing test questions; 3) analyzing question items qualitatively; 4) conducting test trials; 5) analyzing question items quantitatively; 6) revising the test; 7) compiling the test; 8) implementing the test; and 9) interpreting the test results. (Ndiung & Jediut, 2020). The problem is, lecturers in graphic media development courses actually already know about the use of reliability as a prerequisite for test equipment to become a good measuring instrument, but they don't do it. Instructors must also understand various reliability coefficient formulas, such as Cronbach's alpha, 2). KR-20, 3). KR-21, 4). Spearman Brown, odd-even (5). Spearman Brown, beginning-end, 6). Flanagan odd-even, 7). Flanagan beginning-end (9). Rulon, odd-even, and 10). Rulon beginning-end. Which of the different types of reliability has the highest reliability coefficient? Because the reliability of an instrument is the consistency of an instrument in producing almost the same (relatively similar) data even though it is used by different people, times, and places with the same research objectives or research subjects (Lestari & Yudhanegara, 2017).

Gap analysis occurs between expectations and reality. It is hoped that all lecturers who teach graphic media courses can find out the high test reliability coefficient to determine student learning outcomes and improve the quality of test equipment. The reality in the field shows that: 1) most lecturers do not understand how to determine test reliability; 2) lecturers do not make maximum use of test reliability coefficients; and 3) lecturers do not know the various reliability formulas developed by experts. . Ernawati's research shows that the testing tools used so far have been tested qualitatively, but quantitative testing tools have never been implemented (Erawati, 2018). The urgency of this research is that lecturers need to calculate and understand how to determine the reliability coefficient of a test instrument, because accuracy in determining reliability is very necessary to determine the reliability of a test, which will then be used as a measuring tool. to evaluate the reliability of the test instrument. This is one of the expected improvements in the quality of education because reliable test instruments have a big

influence on analyzing student abilities as reference material for improving the quality of learning (Sarwiningsih, 2017). Testing device reliability is considered very important because it is used as a measuring tool to obtain data and information related to the problem being studied (Ayu & Rosli, 2020). It is important for educators and lecturers to know about reliability testing because it determines how reliable a measurement is because of its reliability (Yusup, 2018). Reliability is very important in measurement, especially to obtain consistent measurement results (Khumaedi, 2012). Some relevant previous research related to the reliability of a test instrument is: Arslan's research shows that a reliable TPACK (Technological Pedagogical Content Knowledge) instrument has an impact on the quality of language questions (Arslan, 2020). Research on the effectiveness of using Padlet in distance learning. As a result, the study recommended the use of padlet (Al Momani & Musa, 2022). Research shows that the professionalism mini-evaluation exercise (P-MEX) has adequate validity and reliability in assessing citizens' professionalism and has a positive impact on education (Taşçı et al., 2023). Research on critical thinking skills tests for high school students "produced reliable results in measuring the critical thinking skills of high school students (Orhan & Çeviker AY, 2022). Research aimed at self-assessment among Iranian EFL university students shows that while self-assessment is highly reliable (Manzari, 2023). Research on the Simulation Acceptance Scale (SAS) that was developed can be used because it is reliable (Sezer et al., 2020). Research to provide evidence of the validity and reliability of student assessments of teaching quality using assessments from TIMSS 2019 (Senden et al., 2023).

Research that develops valid and reliable games, good for educators through game-based assessment in physical education (Wilkie et al., 2023). Research aimed at evaluating POMA (Tinetti Performance Oriented Mobility Assessment) stated that POMA has satisfactory reliability and validity among Chinese elderly (Yang et al., 2023). The Canadian Agility Movement Skills Assessment (CAMSA) has instrument reliability and validity so it can be used to assess motor competence in Spanish children (Menescardi et al., 2022). Research that aims to develop a valid and reliable scale to measure students' expectations from the discipline program implemented in their school (Özcan et al., 2020). Research on the Analysis of the National Middle School English Final Examination in View of Validity and Reliability" calculates the reliability results of the Middle School English National Examination using the Kuder-Richardson Formula (KR-20), showing a reliability coefficient value of 0.89 (Sugianto, 2016). The research entitled "Validity and reliability of educational family life measurement instruments" produced a Cronbach's alpha value of 0.950, indicating that the structure of the child measuring instrument (CMI) and the structural dimensions of the mother's measuring instrument MMI are very reliabel. The novelty of this research is that previous research did not widely compare reliability coefficients such as: Cronbach's alpha, 2). KR-20, 3). KR-21, 4). Spearman Brown, odd-even (5). Spearman Brown, beginning-end, 6). Flanagan odd-even, 7). Flanagan beginning-end (9). Rulon, odd-even, and 10). Rulon beginning-end to identify the highest reliability coefficient so that it is appropriate to use as a measuring tool for graphic media UAS. Based on the explanation of the problems above, this research was carried out with the aims of 1) describing several methods for testing the reliability of coefficients, 2) describing several methods for determining reliability coefficients, 3) determining reliability coefficient tests, and 4) determining the reliability coefficients of tests from UAS for media courses graphic. The reliability coefficient compared is 1). Cronbach's alpha, 2). KR-20, 3). KR-21, 4). Spearman Brown, odd-even 5). Spearman Brown, beginning-end, 6). Flanagan odd-even, 7). Flanagan beginning-end (9). Rulon, odd-even, and 10). Rulon beginning-end.

2. METHODS

This type of research is quantitative, using a quantitative descriptive approach. This research includes comparative research, which is a type of research that aims to compare two or more groups, variables, or conditions to identify differences or similarities between them. Data collection methods include: a) observation. Data was obtained through direct observation according to reality (Dwiqi et al., 2020). Observations were carried out in the Unesa Educational Technology department, with the research object being students who were taking graphic media courses. b) interviews aimed at obtaining accurate information from students and lecturers from various questions (Khoridah et al., 2019); c) tests in the form of final exams that aim to obtain data about student learning outcomes. The subjects of this research were 90 undergraduate students in the Department of Educational Curriculum and Technology, class of 2021–2022. The test instrument used is a multiple-choice objective test with 5 answer choices totaling 24. Data analysis uses various reliability formulas assisted by SPSS version 22 and Microsoft Excel. The instrument used in this research is an instrument in the form of a test with graphic design material. The reliability analysis was carried out based on data obtained from the UAS (Cevik & Senturk, 2019). The graphic media Final Semester Examination test instrument grid is presented in Table 1.

Table 1. Final Semester Examination Grid

No	Aspect	Indicator	Item
1	Get to know bitmap and vector type graphic design	- Students know the characteristics of bitmap and vector graphic design.	1, 2
		- Students know various bitmap and vector-type graphic designs.	3, 4, 5
		- Students know the characteristics of bitmap and vector graphic design.	6, 7
2	Shows menus and icons contained in Corel Draw graphic design software	- Students can name color groups in Corel Draw.	8, 9. 10
		- Students can mention the standard Corel Draw menu.	11, 12, 13
		- Students can mention terms in the Corel Draw program.	14, 15
3	Distinguish between bitmap and vector	- Students can differentiate bitmap and vector image results.	16, 17
		- Students can differentiate between the purposes of using bitmaps and vectors.	18, 19
4	Shows menus and icons contained in Adobe Photoshop graphic design software	- Students can mention the standard Adobe Photoshop menu.	20, 21
		- Students can mention the facilities in Adobe Photoshop.	22, 23
5	Determines applications for creating vector-based graphics	- Students can differentiate between bitmaps and vectors from the software they are made from.	24

To estimate reliability using the split-half method, namely Spearman-Brown, Flanagan, and Rulon (Retnawati, 2017; Steinke & Kopp, 2020). Determining the reliability of learning outcomes tests in the form of objective tests is carried out by "splitting in two," or the Split-Half Technique, or single test, single trial question items, namely odd-even and early-end numbered questions (Betul & Turan, 2019; Jago, 2019; Ossai & Chiekem, 2022; Pronk et al., 2022). The language for reliability is r (Fu et al., 2022). Calculations using the odd-even technique are grouping all the odd-numbered items into one group and giving the name "odd" group, while all even-numbered items are grouped into one group and naming 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 (Ajayi, 2013; Van Norman & Parker, 2018). The Spearman Brown reliability test uses the product moment correlation formula and after the correlation is found, the Spearman Brown formula is used (Ekolu & Quainoo, 2019; Nath, 2013; Yusup, 2018). Flanagan's reliability test uses the Split Half technique. Odd – even groups and early – late groups use the same formula as follows (Cho & Chun, 2018). Rulon's reliability test uses the Split Half technique. Odd – even groups and early – late groups use the same formula as follows (Cho & Chun, 2018). Formulas 20 and 21 from Kuder and Richardson are abbreviated as KR-20 and KR-21(Foster, 2021). The Kuder Richardson (KR) reliability test uses a single test technique - single trial. Both KR techniques have specific instrument criteria for using the formula. If the instrument has unequal or heterogeneous levels of difficulty for each question item then use the KR-20 formula to test its reliability (Fraenkel et al., 2013). If the instrument has the same or homogeneous level of difficulty for each question item then use the KR-21 formula to test its reliability (Cho & Chun, 2018; Ekolu & Quainoo, 2019; Foster, 2021; Yusup, 2018). C. Hoyt's formula analyzes scores on learning outcomes test items using variance analysis or ANOVA techniques (Magdalena, Fauziah, et al., 2021; Sudijono, 2006). Cronbach's alpha formula can be used to test essays, questionnaires or questionnaires (Yusup, 2018). Cronbach's alpha shows a tendency to measure consistency rather than reliability (Rosli et al., 2021). Cronbach's Alpha Formula to estimate reliability uses the following formula (Rosli et al., 2021). As for the interpretation of the test reliability coefficient (r_{11}) can use reliability categorization showed in Table 2.

Table. 2. Reliability Categorization

Test Reliability Coefficient	Categorization
1,00 > 0,90	Very high reliability
0,75 - 0,90	Good reliability
0,50 - 0,75	Medium reliability
-1,00 < 0,50	Poor reliability

Source: (Chen et al., 2021; Grgic et al., 2021; Ursavaş & Bayrak, 2021).

In general, the reliability coefficient of a test is described numerically in the range between $-1.00 \leq \rho \leq +1.00$. If this value is close to zero, then the item does not differentiate sufficiently between the upper and lower groups (Çakir, 2022; Retnawati, 2017). The test reliability coefficient can be said to be good if it has a coefficient above 0.70(Alavi & Ghaemi, 2013; Alfiatunnisa et al., 2022; Isa & Azid, 2022; Tobón & Luna, 2021). The higher the coefficient, the smaller the level of measurement error, whereas the lower the coefficient, the greater the measurement error (Retnawati, 2017). The explanation above can give lecturers an idea of the need to compare the accuracy of estimates between reliability formulas so that lecturers can determine the appropriate reliability formula to use for measuring a test. The lecturer will choose a high reliability coefficient for the test meter.

3. RESULT AND DISCUSSION

Results

The results of the split-half calculation using the Spearman Brown odd-even formula produce a reliability coefficient of $r_{11}= 0.84$ (very high category). The results of calculations using the Spearman Brown beginning-end formula produce the following reliability coefficient: $r_{11}= 0.82$ (very high category). The results of calculations using the odd-even Flanagan formula produce a reliability coefficient of $r_{11}= 0.83$ (very high category). The results of calculations using the beginning-end Flanagan formula produce a reliability coefficient of $r_{11}= 0.8$ (very high category). The results of calculations using the odd-even Rulon formula produce a reliability coefficient of $r_{11}= 0.86$ (very high category). The results of calculations using the beginning-end Rulon formula produce a reliability coefficient of $r_{11}= 0.8$ (very high category). The results of calculations using the KR-20 formula produce the following reliability coefficient: $r_{11}= 0.82$ (very high category). The results of calculations using the KR-21 formula produce a reliability coefficient of $r_{11}= 0.8$ (very high category). The results of calculations using the Hoyt formula produce a reliability coefficient of $r_{11}= 0.81$ (very high category). The results of calculations using the Cronbach's alpha formula produce a reliability coefficient of $r_{11}= 0.821$ (very high category). To make it easier for researchers and readers to analyze comparisons of reliability coefficients, researchers need to create a recapitulation table. The recapitulation is presented in Table 3.

Table. 3. Reliability Coefficient Recapitulation

Reliability Type		Results	Note
Spearman Brown	Odd – Even	$r_{11} = 0,84 > 0,7$	very high
	Beginning - End	$r_{11} = 0,82 > 0,7$	very high
Flanagan	Odd – Even	$r_{11} = 0,83 > 0,7$	very high
	Beginning - End	$r_{11} = 0,8 > 0,7$	very high
Rulon	Odd – Even	$r_{11} = 0,86 > 0,7$	very high
	Beginning - End	$r_{11} = 0,8 > 0,7$	very high
Kuder Richardson	20	$r_{11} = 0,82 > 0,7$	very high
(KD)	21	$r_{11} = 0,8 > 0,7$	very high
Hoyt		$r_{11} = 0,81 > 0,7$	very high
Alpha Cronbach		$r_{11} = 0,82 > 0,7$	very high

In Table 3, the reliability coefficients use different reliability formulas. The results of all calculations using various reliabilities are included in the very high category. There are only slight differences in the calculation results between the reliability coefficient formulas used. Research analysis shows that; a) the reliability coefficient in classical test theory has relatively almost the same estimation accuracy, namely 0.8. b) The reliability coefficient has an estimated average of around 0.8; c) All reliability coefficients are above 0.7, which means they are in the very high category; d) All reliability coefficients are suitable for use as a measuring tool for graphic media UAS; e) Reliability coefficient The odd-odd rule has the highest value, namely 0.86. The conclusion is that all reliability coefficients can be used as a tool to test the graphic design UAS test instrument. However, the highest reliability coefficient is obtained by using the odd-even Rulon formula to estimate the reliability coefficient.

Discussion

Reliability is an important psychometric characteristic for all measurements (Steinke & Kopp, 2020). Reliability with the split-half method is the division of question items with odd-even numbers or first and last-numbered questions (Steinke & Kopp, 2020). The comparison test results for all average reliability coefficients are around 0.8 or above 0.7. Research states that the range of reliability coefficients

for a test numerically ranges between -1.00 ≤ ρ ≤ +1.00 (Çakir, 2022; Retnawati, 2017). Correlation coefficient values < 0.50 (poor reliability), values between 0.50 and 0.75 (medium reliability), values between 0.75 and 0.90 (good reliability), and correlation coefficient values > 0.90 (very good reliability) (Chen et al., 2021; Grgic et al., 2021; Ursavaş & Bayrak, 2021). Very good (> 0.90), good (0.75 – 0.90) and poor (< 0.75) (Moya-ramon et al., 2022). Other research states that the correlation coefficient is > 0.60, sufficient reliability), 0.61 – 0.80 (good reliability), and > 0.81 (very good reliability)(Menescardi et al., 2022). Other research states that the correlation coefficient value is between 0 - 0.39 (poor reliability), 0.4 - 0.59 (medium reliability), 0.6 - 0.74 (good reliability), and 0.75 - 1.0 (very good reliability) (Keogh et al., 2019; Lee et al., 2021; Matheson, 2019). Other research states that the correlation coefficient value is between 0.00 – 0.25 (poor reliability), 0.26 – 0.50, moderate reliability (0.51 – 0.75), and good reliability (0.76 – 1, 00)(Paraskevopoulos et al., 2023). Cronbach's Alpha is used to measure reliability capabilities (Ford et al., 2015). Cronbach's alpha (α) assesses the internal consistency and reliability of the test instrument (Wafudu et al., 2022). Cronbach's alpha is recommended to have very good reliability if it ranges between 0.7 and 0.8 (Cheung et al., 2023; Derakhshan et al., 2023; Filho et al., 2021; Kullan et al., 2022). Other research says Cronbach's Alpha is accepted at a value of 0.7 or 0.6" (Taber, 2018). Other research says the Alpha value is > 0.60 for each variable (Ngaliman et al., 2019). Cronbach's alpha $0.5 < \alpha < 0.8$ acceptable reliability (Ekolu & Quainoo, 2019). Thus, the results of the recap of the reliability coefficient calculations can be concluded that all reliability calculations are > 0.70, which means the reliability coefficient is in the very high category.

Theoretically, measurement results can be trusted if in several measurements on the same group of subjects relatively similar results are obtained, as long as the subjects being measured do not experience changes (Lestari & Yudhanegara, 2017; Syamsuddin, 2017). Stable construction over time when inspected (Krieglstein et al., 2022). Test reliability is closely related to validity because a valid measuring instrument can be guaranteed to be reliable, but a reliable measuring instrument is not necessarily valid (Sugiyono, 2013). Arikunto stated that it is important for a test to have validity and reliability requirements, (Arikunto, 2018). In that test it may be reliable, but it is not valid. On the other hand, a valid test is definitely reliable (Suci Mitra & Helendra, 2022). The quality of the questions used to measure test takers' abilities needs to be considered, including whether the questions are valid and reliable. Apart from that, a question is said to be good if it is not too easy or too difficult. The questions must be able to differentiate between smart and not-smart test takers, and the effectiveness of the distractors must function well (Friatma & Anhar, 2019).

The results of the comparison test of the reliability coefficient of UAS scores for graphic media courses using various reliability test methods, as shown in Table 2, apparently produced a reliability coefficient with an average of above 0.80. This is in accordance with several relevant previous studies regarding reliability comparisons. Sarwiningsih's research concluded that various types of reliability coefficients have almost the same accuracy (Sarwiningsih, 2017). Research conducted by Widhiarso and Mardapi (2010) concluded that the split half, KR-20, KR-21, and Cronbach's alpha tests obtained relatively similar results (Widhiarso & Mardapi, 2010). Studies on the Turkish adaptation, validity, and reliability of the Early Childhood Teacher Academic Competency Assessment Scale produced Spearman Brown reliability coefficients and the Split Half Test. Reliability produced figures that were not much different (Betul & Turan, 2019). The advantage of this research is that the test instrument developed for the final semester exam in graphic media courses has very high reliability. This is proven by the average reliability coefficient, which is above 0.7. This research aims to compare various reliability coefficients so that it is known which method has the highest reliability. The limitation of this research is that the researcher only tested the test instrument in terms of its reliability, even though a good test instrument must meet the validity and reliability test requirements. The solution to overcome the limitations of this research is that for future research, researchers will use validity and reliability tests with the same samples and subjects.

4. CONCLUSION

Based on the results of calculating the average reliability coefficient, the results show that all reliability formulas show coefficients above the predetermined average, meaning that all reliability coefficients are included in the very high category; thus, all types of reliability meet the requirements and are suitable to be used as measuring instruments. However, if you use reliability with the odd-even Rulon formula, the results are slightly higher compared to other reliabilities, so it would be better to calculate the UAS test for graphic media development courses using the odd-even Rulon formula.

5. REFERENCES

Ajayi. (2013). a Comparative Analysis of Test Re-Test and Equivalent Reliability Methods. *International Journal of Education and Research*, *1*(6), 209–216.

Al Momani, J. A., & Musa, M. A. A. (2022). A Comparative Study of the Effectiveness of Using Padlet in Distance Learning: Viewpoint of Postgraduate Students. *Journal of Education and e-Learning Research*, *9*(2), 95–102. https://doi.org/10.20448/jeelr.v9i2.3954.

Alavi, S. M., & Ghaemi, H. (2013). Reliability Assessment and Construct Validation of Translation Competence Questionnaire (TCQ) in Iran. *Language Testing in Asia*, *3*(1), 1–10. https://doi.org/10.1186/2229-0443-3-18.

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://doi.org/10.56806/jh.v3i2.81.

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. https://doi.org/10.22460/eltin.v5i1.p35-44.

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.

Arifin, Z. (2017). Kriteria Instrumen dalam suatu Penelitian. *The Original Research of Mathematics*, *2*(1). https://doi.org/10.31949/th.v2i1.571.

Arikunto, S. (2018). *Dasar - Dasar Evaluasi Pendidikan*. Bumi Aksara.

Arslan, A. (2020). Reliability and Validity of TPACK Instruments in EFL. *International Journal of Assessment Tools in Education*, *7*(3), 343–360.

Ayu, S., & Rosli, M. S. Bin. (2020). Uji reliabilitas instrumen penggunaan SPADA (Sistem Pembelajaran dalam Jaringan). *Biormatika*, *6*(1), 145–155. https://doi.org/10.35569/biormatika.v6i1.706.

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. https://doi.org/10.30983/islam_realitas.v5i1.894.

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.

Betul, F., & Turan, F. (2019). Teacher Rating Scales of Early Academic Competence (TRS-EAC): Adaptation to Turkish, Validity and Reliability. *International Journal of Educational Methodology*, *5*(1), 43–57. https://doi.org/10.12973/ijem.5.1.43.

Budiantoro, T., & Kurniawan, B. (2021). Validitas Dan Reliabilitas Instrumen Keterampilan Komunikasi Dan Keterampilan Kolaborasi Pada Mata Kuliah Bahasa Indonesia. *Jurnal Humaniora Teknologi*, *7*(2).

Çakir, H. (2022). Development and Validation of an Achievement Test in Biology. *Journal of Social Sciences and Education*, *5*(1), 64–75. https://doi.org/10.53047/josse.1102697.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods and Research*, *42*(3), 294–320. https://doi.org/10.1177/0049124113500475.

Cevik, M., & Senturk, C. (2019). Multidimensional 21th century skills scale: Validity and reliability study. *Cypriot Journal of Educational Sciences*, *14*(1), 11–28. https://doi.org/10.18844/cjes.v14i1.3506.

Chen, C. H., Yang, S. J. H., Weng, J. X., Ogata, H., & Su, C. Y. (2021). Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers. *Australasian Journal of Educational Technology*, *37*(4), 130–144. https://doi.org/10.14742/ajet.6116.

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2023). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. In *Asia Pacific Journal of Management* (Number 0123456789). Springer US. https://doi.org/10.1007/s10490-023-09871-y.

Cho, E., & Chun, S. (2018). Originators of Reliability Coefficients : A Historical Review of the Originators of Reliability Coefficients Including Cronbach's Alpha. *Survey Research*, *19*(2), 73–104. https://doi.org/10.20997/sr.19.2.4.

Danni, R., & Tauratiya, T. (2020). Analisis Kemampuan Berpikir Kritis Mahasiswa Program Studi Hukum Keluarga Islam IAIN Syaikh Abdurrahman Siddik Bangka Belitung. *Tarbawy : Jurnal Pendidikan Islam*, *7*(1), 17–22. https://doi.org/10.32923/tarbawy.v7i1.1191.

Derakhshan, L., Maarefvand, M., Ebadi, A., & Mousvai, M. T. (2023). Evaluation of the validity and reliability of Children's Worlds Subjective Well-Being Scale (CW-SWBS) in Tehran, Iran. *Children and Youth Services Review*, *155*(October), 107185. https://doi.org/10.1016/j.childyouth.2023.107185.

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.

Dwiqi, G. C. S., Sudatha, I. G. W., & Sukmana, A. I. W. I. Y. (2020). Pengembangan Multimedia Pembelajaran Interaktif Mata Pelajaran IPA Untuk Siswa SD Kelas V. *Jurnal Edutech Undiksha*, *8*(2), 33. https://doi.org/10.23887/jeu.v8i2.28934.

Ekolu, S. O., & Quainoo, H. (2019). Reliability of Assessments in Engineering Education using Cronbach's Alpha, KR and Split-Half Methods. *Global Journal of Engineering Education*, *21*(1), 24–29.

Erawati, N. K. (2018). Analisis Tes Penilaian Pencapaian Kompetensi Pada Mahasiswa Kebidanan. *Jurnal Penjakora*, *5*(2), 111–120.

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.

Faridah, A. (2021). Karakteristik Butir Soal Ujian Akhir Semester Mata Pelajaran Sejarah. *Ekspose: Jrnal Penelitian Hukum dan Pendidikan*, *20*(2), 1281–1288. https://doi.org/10.30863/ekspose.v20i2.1819.

Filho, V. C. B., Bandeira, A. S., Rech, C. R., Lopes, A. S., Mota, J., & Silva, K. S. (2021). Validity and reliability of an instrument to measure factors associated with screen time in brazilian students. *Ciencia e Saude Coletiva*, *26*(3), 1047–1061. https://doi.org/10.1590/1413-81232021263.31322018.

Ford, M. E., Wei, W., Moore, L. A., Burshell, D. R., Cannady, K., Mack, F., Ezerioha, N., Ercole, K., & Garrett-Mayer, E. (2015). Evaluating the reliability of the Attitudes to Randomized Trial Questionnaire (ARTQ) in a predominantly African American sample. *SpringerPlus*, *4*(1). https://doi.org/10.1186/s40064-015-1208-z.

Foster, R. C. (2021). KR20 and KR21 for Some Nondichotomous Data (It's Not Just Cronbach's Alpha). *Educational and Psychological Measurement*, *81*(6), 1172–1202. https://doi.org/10.1177/0013164421992535.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2013). How to Design and Evaluate Research in Education. *In McGraw-Hil*, *53*. https://doi.org/10.1017/CBO9781107415324.004.

Fu, X., Zhang, Z., Zhou, Y., Chen, Q., Yang, L. Z., & Li, H. (2022). The Split-Half Reliability and Construct Validity of the Virtual Reality-Based Path Integration Task in the Healthy Population. *Brain Sciences*, *12*(12). https://doi.org/10.3390/brainsci12121635.

Grgic, J., Lazinica, B., & Pedisic, Z. (2021). Test–Retest Reliability of the 30–15 Intermittent Fitness Test: A Systematic Review. *Journal of Sport and Health Science*, *10*(4), 413–418. https://doi.org/10.1016/j.jshs.2020.04.010.

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.

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://doi.org/10.37758/annawa.v4i1.419.

Hikmah, & Muslimah. (2021). Validitas dan Reliabilitas Tes dalam Menunjang Hasil Belajar PAI. *Proceedings*, *1*, 345–356.

Indriasari, B. A., Efila, M. R., & Korespondensi, A. (2021). Analisis Validitas Soal Ujian Akhir Semester Mata Kuliah Wajib Umum (Mkwu) Bahasa Indonesia Universitas Katolik Musi Charitas. *Jurnal Evaluasi Pendidikan*, *12*(1), 20–26.

Isa, Z. C., & Azid, N. (2022). Multimedia constructivism instrument: Validity and reliability analysis. *International Journal of Evaluation and Research in Education*, *11*(4), 1818–1824. https://doi.org/10.11591/ijere.v11i4.22730.

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.

Kadariah, K. (2018). Evaluasi Penyelenggaraan Ujian Akhir Semester Mahasiswa Program Non Pendidikan Dasar Di UPBJJ-UT Makassar. *Jekpend: Jurnal Ekonomi dan Pendidikan*, *1*(1), 97. https://doi.org/10.26858/jekpend.v1i1.5075.

Keogh, J. W. L., Alistair, Anderson, S., Liew, B., Olsen, A., Schram, B., & Furness, J. (2019). Reliability and Validity of Clinically Accessible Smartphone Applications to Measure Joint Range of Motion : A Systematic Review. *PLoS ONE*, *14*(5), 1–24. https://doi.org/10.1371/journal.pone.0215806.

Khoridah, F., Prasetiyawati, D., & Baedowi, S. (2019). Analisis Penerapan Metode SAS (Struktural Analitik Sintetik) dalam Kemampuan Menulis Permulaan. *Journal for Lesson and Learning Studies*, *2*(3), 396–403. https://doi.org/10.23887/jlls.v2i3.19899.

Khumaedi, M. (2012). The Reliability of Education Research Instrumens. *JPTM: Jurnal Pendidikan Teknik Mesin*, *12*(1), 25–30.

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. In *Educational Psychology Review* (Vol 34, Number 4). Springer US. https://doi.org/10.1007/s10648-022-09683-4.

Kullan, S., Mansor, M., & Ishak, R. (2022). The validity and reliability of an instrument to evaluate the practices of learning organization. *International Journal of Evaluation and Research in Education*, *11*(4), 1725–1733. https://doi.org/10.11591/ijere.v11i4.22974.

Kurniawati, A. (2019). Analisis Hasil Tes Evaluasi Pendidikan Pada Mahasiswa Ditinjau Dari Perbedaan Gender. *Jurnal Ilmiah DIDAKTIKA*, *19*(1), 89–106.

Lee, D., Lim, W. Y., Park, S., Jin, Y. W., Lee, W. J., Park, S., & Seo, S. (2021). Reliability and Validity of a Nationwide Survey (the Korean Radiation Workers Study). *Safety and Health at Work*, *12*(4), 445–451. https://doi.org/10.1016/j.shaw.2021.07.012.

Lestari, & Yudhanegara. (2017). *Penelitian Pendidikan Matematika*. PT Refika Aditama.

Lia, R. M., Rusilowati, A., & Isnaeni, W. (2020). NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model. *Research and Evaluation in Education*, *6*(1), 41–50. https://doi.org/10.21831/reid.v6i1.30112.

Lockhart, J. (2015). Using Item Analysis to Evaluate the Validity and Reliability of an Existing Online Information Literacy Skills Assessment Instrument. *South African Journal of Libraries and Information Science*, *80*(2), 36–46. https://doi.org/10.7553/80-2-1515.

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.

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.

Manzari, A. (2023). Reliability and Validity of Self-Assessments Among Iranian EFL University Students. *International Journal of Language Testing*, *13*(1), 225–235. https://doi.org/10.22034/IJLT.2023.363420.1199.

Matheson, G. J. (2019). We Need to Talk about Reliability: Making better Use of Test-Retest Studies for Study Design and Interpretation. *PeerJ*, *2019*(5), 1–25. https://doi.org/10.7717/peerj.6918.

Menescardi, C., Villarrasa-Sapiña, I., Lander, N., & Estevan, I. (2022). Canadian Agility Movement Skill Assessment (CAMSA) in a Spanish Context: Evidences of Reliability and Validity. *Measurement in Physical Education and Exercise Science*, *26*(3), 245–255. https://doi.org/10.1080/1091367X.2021.2020794.

Moya-ramon, M., Mateo-march, M., Peña-gonzález, I., Zabala, M., & Javaloyes, A. (2022). Computer Methods and Programs in Biomedicine Validity and Reliability of Different Smartphones Applications to Measure HRV During Short and Ultra-Short Measurements in Elite Athletes. *Computer Methods and Programs in Biomedicine*, *217*, 106696. https://doi.org/10.1016/j.cmpb.2022.106696.

Nath, S. (2013). Best Split-Half and Maximum Reliability. *IOSR Journal of Research & Method in Education (IOSRJRME)*, *3*(1), 01–08. https://doi.org/10.9790/7388-0310108.

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..

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.

Ngaliman, Eka J, M. G., & Suharto. (2019). The Effect Of Tangibles, Responsiveness, And Reliability On Customer Satisfaction Of Delivery Services. *International Journal of Economics and Management Studies*, *6*(5), 86–92. https://doi.org/10.14445/23939125/ijems-v6i5p113.

Orhan, A., & Çeviker AY, Ş. (2022). Developing the Critical Thinking Skill Test for High School Students: A Validity and Reliability Study. *International Journal of Psychology and Educational Studies*, *9*(1), 130–142. https://doi.org/10.52380/ijpes.2022.9.1.561.

Ossai, P. A. U., & Chiekem, E. (2022). Split-half method of estimating reliability index : a comparison of the odd-even and first-and second-half ways of splitting test items . *Psychology and Education*, *59*(1), 612–624.

Özcan, G., Aktağ, I., & Gülözer, K. (2020). Developing the Scale on Discipline Expectations of Students: A Validity and Reliability Study. *International Journal of Evaluation and Research in Education*, *9*(4), 840–846. https://doi.org/10.11591/ijere.v9i4.20585.

Paraskevopoulos, E., Pamboris, G. M., Plakoutsis, G., & Papandreou, M. (2023). Reliability and measurement error of tests used for the assessment of throwing performance in overhead athletes: A systematic review. *Journal of Bodywork and Movement Therapies*, *35*(April), 284–297. https://doi.org/10.1016/j.jbmt.2023.04.042.

Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (2022). Methods to Split Cognitive Task Data for Estimating Split-Half Reliability: A Comprehensive Review and Systematic Assessment. *Psychonomic Bulletin and Review*, *29*(1), 44–54. https://doi.org/10.3758/s13423-021-01948-3.

Purba, R. A. (2021). Efektivitas Aplikasi UASOnline Dalam Menjaga Mutu Perkuliahan Di Masa Pendemi Covid-19. *Jurnal Pendidikan Teknologi dan Kejuruan*, *18*(2), 185. https://doi.org/10.23887/jptk-undiksha.v18i2.34101.

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. https://doi.org/10.26630/jk.v13i1.2814.

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.

Retnawati, H. (2017). Reliabilitas Instrumen Penelitian Pendidikan. *Jurnal Pendidikan Teknik Mesin Unnes*, *12*(1), 129541. https://doi.org/10.17977/jip.v20i2.4615.

Retnawati, H., & Hadi, S. (2014). Sistem Bank Soal Daerah Terkalibrasi untuk Menyongsong Era Desentralisasi. *Jurnal Ilmu Pendidikan*, *20*(2), 183–193. https://doi.org/10.17977/jip.v20i2.4615.

Rosli, M. S., Saleh, N. S., Alshammari, S. H., Ibrahim, M. M., Atan, A. S., & Atan, N. A. (2021). Improving Questionnaire Reliability using Construct Reliability for Researches in Educational Technology. *International Journal of Interactive Mobile Technologies*, *15*(4), 109–116. https://doi.org/10.3991/IJIM.V15I04.20199.

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.

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.

Senden, B., Nilsen, T., & Teig, N. (2023). The Validity of Student Ratings of Teaching Quality: Factorial Structure, Comparability, and the Relation to Achievement. *Studies in Educational Evaluation*, *78*(May), 101274. https://doi.org/10.1016/j.stueduc.2023.101274.

Sezer, B., Teker, G. T., Sezer, T. A., & Elcin, M. (2020). Simulation Acceptance Scale (SAS): A Validity and Reliability Study. *Australasian Journal of Educational Technology*, *36*(3), 56–68. https://doi.org/10.14742/AJET.4950.

Steinke, A., & Kopp, B. (2020). Relex: An Excel-Based Software Tool for Sampling Split-Half Reliability Coefficients. *Methods in Psychology*, *2*(February), 0–6. https://doi.org/10.1016/j.metip.2020.100023.

Subarkah, I., Rinawati, A., & Munawaroh, I. (2020). Evaluasi Program Ujian Akhir Semester (UAS) Virtual Model CIPP. *Jurnal Cakrawala: Manajemen Pendidikan Islam dan Studi Sosial*, *4*(2), 14–29. https://doi.org/10.33507/cakrawala.v4i2.244.

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). https://doi.org/10.30870/biodidaktika.v17i2.16493.

Sudijono, A. (2006). *Pengantar Evaluasi Pendidikan*. PT. RajaGrafindo Persada.

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://doi.org/10.23971/jefl.v6i1.427.

Sugiyono. (2013). Metode Penelitian Pendidikan. In *Alfabeta*.

Supriyadi, 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.

Suradi, A. A. M. (2022). Perancangan Sistem Informasi Ujian Akhir Semester Berbasis Komputer Pada Perguruan Tinggi. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, *12*(1), 30. https://doi.org/10.35585/inspir.v12i1.2679.

Surucu, L., & Maslakci, A. (2020). Validity and Reliability in Quantitative Research. *Business & Management Studies: An International Journal*, *8*(3), 2694–2726.

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.

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.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Taşçı, A. İ., Akdeniz, E., Gülpınar, M. A., Danacıoğlu, Y. O., Sarı, E. E., Yaşar, L., Karandere, F., & Ferahman, S. (2023). Adaptation of The Professionalism Mini-Evaluation Exercise Instrument into Turkish: a Validity and Reliability Study. *BMC Medical Education*, *23*(1), 1–13. https://doi.org/10.1186/s12909-023-04675-6.

Tobón, S., & Luna, J. (2021). Complex thinking and sustainable social development: Validity and reliability of the complex-21 scale. *Sustainability (Switzerland)*, *13*(12), 1–19. https://doi.org/10.3390/su13126591.

Tri Jampi Setiyorini, Zyah Rochmad Jaelani, A. N. (2022). Analisis Faktor-Faktor Yang Mempengaruhi Reliabilitas. *Didaktis: Jurnal Pendidikan dan Ilmu Pengetahuan*, *22*(3).

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.

Ursavaş, F. E., & Bayrak, D. (2021). Validity and reliability of the Turkish version of the pressure injury knowledge assessment tool 2.0. *Journal of Tissue Viability*, *30*(4), 582–587. https://doi.org/10.1016/j.jtv.2021.05.002.

Van Norman, E. R., & Parker, D. C. (2018). A Comparison of Split-Half and Multilevel Methods to Assess the Reliability of Progress Monitoring Outcomes. *Journal of Psychoeducational Assessment*, *36*(6), 616–627. https://doi.org/10.1177/0734282917696936.

Wafudu, S. J., Kamin, Y. Bin, & Marcel, D. (2022). Validity and reliability of a questionnaire developed to explore quality assurance components for teaching and learning in vocational and technical education. *Humanities and Social Sciences Communications*, *9*(1), 1–10. https://doi.org/10.1057/s41599-022-01306-1.

Wenno, I. H., Tuhurima, D., & Manoppo, Y. (2021). How to Create a Good Test. *Jurnal Pendidikan Profesi Guru Indonesia*, *1*(1), 11–20.

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.

Wilkie, B., Jordan, A., Foulkes, J., Woods, C. T., Davids, K., & Rudd, J. (2023). Examining the Validity, Reliability and Feasibility of Capturing Children's Physical Literacy Through Games-Based Assessment in Physical Education. *Frontiers in Sports and Active Living*, *5*(June), 1–13. https://doi.org/10.3389/fspor.2023.1188364.

Yang, C., Mo, Y., Cao, X., Zhu, S., Wang, X., & Wang, X. (2023). Reliability and Validity of the Tinetti Performance Oriented Mobility Assessment in Chinese Community-Dwelling Older Adults. *Geriatric Nursing*, *53*, 85–89. https://doi.org/10.1016/j.gerinurse.2023.06.020.

Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, *7*(1). https://doi.org/10.18592/tarbiyah.v7i1.2100.

Zohrabi, M. (2013). Mixed Method Research: Instruments, Validity, Reliability and Reporting Findings. *Theory and Practice in Language Studies*, *3*(2), 254–262. https://doi.org/10.4304/tpls.3.2.254-262.