VALIDITY AND RELIABILITY OF TEST QUESTIONS TO MEASURE THE INFORMATION LITERACY SKILLS OF PROSPECTIVE TEACHER STUDENTS

NUR RAHMI AKBARINI[1], ARUN ANGGRAWAL[2]

[1] Program Studi PEndidikan Administrasi Perkantoran, Universitas Sebelas Maret, Surakarta, Indonesia

[2]Chitkara Business School, Chitkara University, Chandigarh, Punjab, India

nurrahmia@staff.uns.ac.id

Abstract

This study aims to describe the quality of information literacy instruments as seen from the aspects of validity, reliability, and difficulty level of questions. This research is a quantitative descriptive research. Data were obtained through documentation method in the form of 22 items of multiple choice questions with four answer options and tested on 224 students majoring in teacher education in a university. The data were analyzed using the classical test theory approach and Rasch model with JMetric software. Validity was assessed by content validity which relies on the use of a panel of experts to evaluate the elements of the instrument and rate them based on their relevance and representativeness to the content domain. To measure the experts' judgment, several indices have been discussed in this paper such as content validity ratio (CVR) and content validity index (CVI). From the CVR results, it is known that 20 questions are valid and 2 questions are considered invalid. From the results of the CVR value, the CVI index of 0.935 was calculated. The reliability value using the KR-20 formula through the classical test theory approach is 0.740 in the sufficient category and the reliability value in the Rasch model is 0.720 (sufficient) for Person Reliability Item Reliability of 0.97 (high). The results of the analysis of the level of difficulty based on the Rasch model show that there are 2 items with difficulty levels (10%), 7 items with moderate difficulty (35%), and 11 items with easy difficulty (55%).

Keywords: *validity; reliability, difficulty level; information literacy*

INTRODUCTION

The integration of digital learning and tools is a trend in education today, which requires educators and students to be more willing to develop technology. In this era of technological development, it is possible for students to become independent learners so that understanding and integration of students' information literacy skills is fundamental. In the educational context, information literacy is important for students in searching, finding, analyzing, evaluating and communicating information that functions to fulfill information needs that will solve various problems (Zaenab et al., 2017) . This leads to how students are able to select information in the era of the digital information explosion. Information literacy is the ability to determine the level of

information needed effectively and efficiently, as well as being able to evaluate information and utilize it for certain purposes (American Library Association, 2000) .

In higher education, students, namely students, are more required to be able to learn independently, so they must have information literacy skills (Julianti et al., 2022). To support academic success and ensure the quality of graduates so they are able to compete and adapt to developments in the world of work, students need to be equipped with information literacy skills (Muntashir, 2016) . Apart from that, information literacy is also needed to improve one's quality for lifelong learning. This is in line with (Ranaweera, 2020) who said that information literacy equips them with the critical skills needed to become independent lifelong learners. Therefore, information literacy skills among students are a necessity and will greatly support teaching and learning activities in higher education (Alam, 2013) .

There are several study programs in higher education that focus on preparing prospective teachers where the prospective teachers will teach knowledge to their students. A prospective educator is certainly expected to be information literate, as is the case with Kinengyere in (Zeeshan et al., 2020)who states that someone who is information literate needs to know how to clearly define information needs; subject or subject area; choose appropriate terminology that expresses the concept being investigated; formulate search strategies that take into account the various sources of information and the various ways that information is organized; analyze collected data for value, relevancy, relevance, quality and suitability; and then turn it into knowledge. (Beile, 2007) states that it is rare for teachers who have just entered the world of education to be well prepared to teach and provide examples of information literacy to their students. Traditionally, we assume that students will acquire information literacy skills automatically on their own. However, in reality, information literacy skills need to be instilled among students by teachers (Ranaweera, 2020) . (White, 2019) emphasizes that information literacy has a close relationship to fostering critical thinking skills in students' critical thinking.

Some of the problems that often arise regarding literacy skills in the era of technological development are that a lot of information in student assignments is not evaluated so that the authenticity, validity and reliability of the information becomes doubtful. Many students struggle with the information literacy skills of searching for appropriate information and evaluating the validity of sources when completing assignments (Hanbidge et al., 2018) . It is considered necessary for student teachers to be equipped with information literacy skills as early as possible, so that when

they enter to become professional teachers they can pass on information literacy skills to their students.

To determine the information literacy abilities of prospective teacher students, a valid and reliable instrument is needed. An instrument is said to be good if a feasibility and reliability test has been carried out. This test is carried out so that the instrument developed is in accordance with the objectives of the development so that the data obtained is correct (Nurfadillah et al., 2023) . In line with (Mardapi, 2012) that the instruments used in assessments have several requirements to be suitable for use, namely valid and reliable. Instruments that have internal validity such as tests must meet construct validity and content validity (Fitriyanto et al., 2019) . Reliability refers to the understanding that an instrument can be trusted enough to collect data because it is good, meaning that whenever it is used it will give the same results (Arikunto, 2010) .

Although reliability is important, it is not enough if it is not combined with validity. In other words, for a test to be reliable, the test must also be valid (Wilson in (Segal & Coolidge, 2018)). The two are often mentioned together, but they provide different information. Reliability and validity are the two most important properties a test score can have. Validity relates to what an instrument measures, and how well it measures it (Mohajan, 2017) . Validity shows whether the test score measures the right thing for the use of the test (Livingston, 2018). Meanwhile, reliability tells us how consistently the test scores measure something (Livingston, 2018) , this gives rise to someone's confidence in the data obtained from using the instrument (Mohajan, 2017) .

According to (Mok & Wright, 2004) the concept of objective measurement in the social sciences must have five categories, namely being able to provide a linear measure of value and with the same interval, carrying out a precise estimation process, being able to find items that are incorrect or not. general, overcomes missing data, produces independent measurements and assessments of the parameter data studied. Of the five categories that have been explained, so far only the Rasch model can fulfill these five categories (Ibnu et al., 2019). The Rasch model is a measurement model that allows the assessment of individual abilities and the characteristics of the test items separately. Thus, this model can help in identifying difficult questions and students who have high or low ability (Adawiyah et al., 2023). The way it works is by measuring individual abilities and the characteristics of the test items separately, making it possible to identify difficult test items and students who have high or low abilities. This is because using the Rasch model is a solution to the validity problem where the Rasch model is able to provide statistics and offers the

opportunity to investigate the validity of test instruments based on the responses of research subjects (Ibnu et al., 2019).

The aim of this research is to examine the validity of the instrument and estimate the reliability of the instrument for measuring the information literacy level of prospective teacher students. Next, the items are analyzed using Rasch modeling. The application of Rasch modeling measurements to raw test result data aims to produce a measurement scale with equal intervals which in the end can provide accurate information about the test taker's abilities and the quality of the questions worked on by the participants (Erfan et al., 2020) . Several previous studies regarding information literacy abilities used self-assessment via a Likert type scale (Prihandoko, 2021; Zeeshan et al., 2020) . In this research, a test instrument was used where each answer was given a score. Valid and reliable instruments will help in assessing students' information literacy skills and can later help institutions work on things that can help improve the information literacy skills of prospective teacher students.

RESEARCH METHODS

The method used in this research is part of development research. The instrument used in this research is a test instrument, namely a measuring tool to collect data regarding the information literacy of prospective teacher students. The type of instrument used is a closed instrument, namely an instrument that has the answers provided so that the respondent just has to provide answer choices that match the answer. Test instruments were given to validators and research respondents to collect data.

To produce an instrument blueprint that is appropriate and in line with the objectives, the instrument creation process must refer to information literacy indicators. The scale items were developed based on the National Educational Technology Standards (ISTE for Educators) and the Association of College & Research Libraries (ACRL) information literacy competency standards. This instrument consists of 22 objective multiple choice questions designed to measure prospective teacher students' cognitive knowledge about information literacy (Indrasvari et al., 2021). The standard (ACRL, 2019) provides an information literacy framework, then (Eynon, 2013) divides it into five standard indicators as follows:

a. Determine the scope and nature Information

Indicator This is used to know skills students in matter formulate information needed, identify types and varieties of information formats, as well as ability For evaluate the return nature and scope of information obtained.

b. Access Information with Effective and Efficient

Indicator This is used to know skills students in matter choose the method to search, using search strategies such as URL and type document as well as quoting, noting, and managing source information.

c. Evaluate Information Based on Source

Indicator This is used to know Skills student in matter summarize the main idea quoted, using the main idea from information obtained For construct draft new, as well compare and analyze information obtained with knowledge that has been There is.

d. Use Information for Specific Purposes

Indicator This is used for know Skills Students in matter use information new and knowledge previous to produce work, communicate results with appropriate media, and use a bibliography in making work.

e. Use Information By Ethical

Indicator This is used For know Skills students in use containing information right create and recognize necessary information accessed with permission special.

Instrument This arranged based on the existing blueprint made Previously , points from the blue print developed become A question short , so can presented and easy understood by respondents . Types of research data is qualitative and quantitative data, data is analyzed descriptive statistics . Qualitative data form comments and input For repair product from media experts , languages, and practitioners . Furthermore analyzed and described in a way descriptive qualitative For revise developed products . Quantitative data obtained from score assessment media, language, practitioner and score experts test instrument response students on a readability test . Reviewers expert evaluate details question For accuracy, clarity, difficulty, and objectivit. After survey and item revision, testing more scale big done. Test instrument spread to student undergraduate level major teacher education at a university, and obtained 224 answers respondents. Data analysis is activity For research, study, study, compare existing data and create necessary interpretation

(Fitriyanto et al., 2019). In research this is an existing instrument So Then tested its validity and reliability.

Content Validity

Validity fill something item can proven with using Content validity ratio (CVR) and Content validity index (CVI) or V aiken coefficient (Bashooir & Supahar, 2018) A number expert requested To identify items with a response provided. (Lawshe, 1975) suggests two assumptions are made and each is consistent with principles psychophysical have set namely (1) Every item a matter of performance considered "essential" by more than half of panelists, has a level of validity fill certain; (2) Increasing Lots panelists (more of the 50%) who consider the item "essential", the more big level validity its contents. In (Lawshe, 1975) there are three rating scales viz Essential, Useful However No essential, or Not required. Validity This involving 7 experts: three lecturers of expert technology information, two lecturers Linguist, and two practitioners. The CVR value can be calculated with the use CVR formula = (ne - N/2) / (N/2), where ne is amount experts stating the item is relevant and N is the total number of experts involved in validation (Puspitasari & Febrinita, 2021) . CVR value has a range between -1 to with 1. If half from SME stated essential so the CVR value will be value 0. CVR will value 1 if all SMEs stated essential For something item (Bashooir & Supahar, 2018) .

Table 1. CVR Minimum Standards based on the number of SMEs

Number of SMEs	Minimum CVR value
5	0.99
6	0.99
7	0.99
8	0.75
9	0.78
10	0.62
11	0.59
12	0.56
13	0.54
14	0.51
15	0.49
20	0.42
25	0.37
30	0.33
35	0.31
40	0.29

(Lawshe, 1975)

Items that obtain a negative CVR value are invalid and must be eliminated. The overall test validity value can be determined using the CVI (Content Validity Index). Determination of CVI (Lawshe, 1975, p. 568; Azwar, 2012, p. 115) is carried out using the formula:

$$CVI: \frac{(\sum CVR)}{k}$$

CVI can calculated with count validity filled in individual items (i-CVI) (Puspitasari & Febrinita, 2021). On proof validity content (Azwar, 2012, p. 113), researcher can determine many desired rating category.

Validity empirical

Validity empirical obtained from analysis response to given test to respondents. The response obtained from the testing test to the respondents. Validity empirical can determined by using Classical Test Theory (CTT) or Item Response Theory (IRT) (Retnawati, 2016). However, a number of research can use the Rasch model when Already getting results trial field (Wati & Mahtari, 2017). In this research, Rasch modeling is carried out with JMetrik program help. Criteria used is, the increasingly smaller the index obtained, the more difficult the question. On the contrary increasingly big the index obtained, the easier question. Criteria Index difficulty question. The details are there in the range under This can used for needs measurement Skills literacy information student.

Mark	Interpretation
Less than 0.3	Too hard
0.30 – 0.70	Fair (Medium)
More of 0.70	Too easy

(Sudijono, 2011)

Reliability

One way to determine the level of instrument reliability is to use the KR-20 calculation. In this study, instrument reliability estimation using the KR-20 was used to estimate the reliability of the test instrument. There is no absolute rule for internal consistency, but most agree with a minimum internal consistency coefficient of 0.70 (Whitley, 2002). Meanwhile, for exploratory studies or pilot studies, it is recommended that reliability be equal to or above 0.60 (Taherdoos, 2016). The KR-20 formula can be used for dichotomous item analysis. For instrument items with

dichotomous scoring, for example 1-0, true-false, yes-no, on-dead, etc., reliability estimation can be done using the KR-20 formula (Retnawati, 2016). The KR-20 formula is as follows:

$$r_{ii} = \frac{k}{(k-1)}\left\{\frac{s_t^2 \sum p_i q_i}{s_t^2}\right\}$$

With

r_{ii} = reliability of instrument scores;

k = number of questions or number of questions;

s_t^2 = total score variance;

p_i = proportion of subjects who answered correctly on an item (proportion of subjects who got a score of 1) which is calculated by

$$p_i = \frac{banyak\ subjek\ yang\ skornya\ 1}{N}; dan\ q_i = 1 - p_i$$

(Retnawati, 2016)

RESULTS AND DISCUSSION

The results of this research are in the form of a blueprint of the research instrument including the level of validity and reliability estimates. The written test assessment tools were analyzed using content validity. The content validation process and content validity index are important factors in the instrument development process, they must be treated and reported as important as other types of construct validation (Almanasreh et al., 2019). Content validity is different from other types of validity because content validity refers to test-based validity and not score-based. This validity describes the necessary elements of content for the instrument and is not related to the scores obtained from the construct (Almanasreh et al., 2019). Content validity ratio (CVR), an item statistic originally suggested by Lawshe (1975), is one of the most widely used methods for measuring content validity (Almanasreh et al., 2019) . The written test assessment tool in this study was analyzed using Lawshe content validity where the CVR validity standard depends on the number of SMEs. The number of SMEs in this study was 7, so the CVR value must meet 0.99 for the item to be declared valid (Lawshe, 1975) . Based on the CVR value exceeding 0.99, all items are declared valid and suitable for use in further research (Bashooir & Supahar, 2018) .

Table 2. Data from expert evaluation results and CVR value

Query Item	Member							CVR	Results
	Member 1	Member 2	Member 3	Member 4	Member 5	Member 6	Member 7		
item 1	3	3	3	3	3	3	3	1	Valid
item 2	3	3	3	3	3	3	3	1	Valid
item 3	3	3	3	3	3	3	3	1	Valid
item 4	3	3	3	3	3	3	3	1	Valid
item 5	3	3	3	3	3	3	3	1	Valid
item 6	3	3	3	3	3	3	3	1	Valid
item 7	3	3	3	3	3	3	3	1	Valid
item 8	3	3	3	3	3	3	3	1	Valid
item 9	3	3	3	3	3	3	3	1	Valid
item 10	3	3	3	3	3	3	3	1	Valid
item 11	3	2	3	3	2	3	3	0,428571	Invalid
item 12	3	3	3	3	3	3	3	1	Valid
item 13	3	3	3	3	3	3	3	1	Valid
item 14	3	3	3	3	3	3	3	1	Valid
item 15	3	3	3	3	3	3	3	1	Valid
item 16	3	3	3	3	3	3	3	1	Valid
item 17	3	3	3	3	3	3	3	1	Valid
item 18	3	3	3	3	3	3	3	1	Valid
item 19	3	3	3	3	3	3	3	1	Valid
item 20	3	3	3	3	3	3	3	1	Valid
item 21	3	3	3	3	3	3	3	1	Valid
item 22	3	2	2	3	2	3	3	0,142857	Invalid

From the table above, it can be seen that of the 22 questions that have been validated by the validator, the CVR value is above 0.99 for 20 items, so from the table it is stated that 20 questions are valid, while 2 questions are invalid. Question 2 is invalid, items not used and not given to respondents. In testing content validity, if you use a group of experts you can usually determine the content validity ratio (CVR) and content validity index (CVI). CVI calculations should be carried out only on selected items, namely items that have been declared to have a valid CVR (Puger, 2021) . The CVR coefficient shows whether or not an item can be used as part of an information literacy test. Meanwhile, CVI is a direct indication of the average number of items that have the ability to compose an instrument (Puger, 2021) . CVI represents the average CVR value of the retained items. Based on existing psychophysical principles, Lawshe proposed that a 50% agreement level ensures content validity (Almanasreh et al., 2019) . The CVI value obtained can be displayed as follows:

$$CVI: \frac{(\sum CVR)}{k} = \frac{0,935065}{22} = 0.935065$$

Next, the instrument was tested on 224 prospective teacher students. Student responses were analyzed using the IRT Rasch model assisted by the JMetrik program. The results of research (Aksu et al., 2019) found that the jMetric program was able to carry out Item Response Theory (IRT) analysis for two-category and multi-category questions.

Table of characteristics of the questions.

```
                        RASCH ANALYSIS
                     tes1i.JAWABANMAHASISWA
                  November 21, 2023  11:37:17


                  FINAL JMLE ITEM STATISTICS
=========================================================================
Item      Difficulty    Std. Error     WMS    Std. WMS    UMS    Std. UMS
-------------------------------------------------------------------------
var1        -1.26         0.17        1.10      1.13      1.20      1.13
var2        -0.38         0.15        1.06      0.95      1.10      1.03
var3        -0.54         0.16        1.23      3.38      1.35      2.94
var4        -0.24         0.15        1.21      3.15      1.22      2.29
var5         1.36         0.16        0.94     -0.79      0.91     -0.73
var6        -0.49         0.15        1.04      0.69      1.05      0.51
var7         1.13         0.16        1.20      2.81      1.34      3.09
var8        -1.49         0.18        0.96     -0.39      1.09      0.53
var9         1.28         0.16        0.95     -0.67      0.90     -0.91
var10        0.85         0.15        0.80     -3.27      0.74     -3.33
var11        0.82         0.15        0.97     -0.46      1.00      0.03
var12        0.30         0.15        0.98     -0.29      0.95     -0.63
var13        0.37         0.15        0.91     -1.48      0.91     -1.23
var14        0.57         0.15        1.12      1.93      1.15      1.84
var15        0.73         0.15        1.00     -0.05      0.99     -0.11
var16       -1.94         0.21        0.93     -0.45      0.74     -1.04
var17        0.68         0.15        0.82     -2.98      0.78     -2.96
var18       -1.00         0.17        0.90     -1.36      0.84     -1.07
var19       -1.29         0.18        0.95     -0.51      1.01      0.11
var20        0.55         0.15        0.84     -2.62      0.79     -2.89
=========================================================================
```

The level of difficulty of the test items can influence the shape of the distribution of test questions. For very difficult tests, the distribution is positively skewed, while for easy tests the distribution is negatively skewed (Fatimah & Alfath, 2019) . The level of difficulty of a question item shows how many respondents are likely to be able to answer a question item correctly (Erfan et al., 2020) . If you look at the Rasch analysis table above, it can be concluded that:

Question	Index	Criteria
Question 1	-1.26	Easy Question
Question 2	-0.38	Medium Problem
Question 3	-0.54	Medium Problem
Question 4	-0.24	Difficult Question
Question 5	1.36	Easy Question
Question 6	0.49	Medium Question
Question 7	1.13	Easy Question
Question 8	-1.49	Easy Question
Question 9	1.28	Easy Question

Question 10	0.85	Easy Question
Question 11	0.82	Easy Question
Question 12	0.30	Difficult Question
Question 13	0.37	Medium Question
Question 14	0.57	Medium Question
Question 15	0.73	Easy Question
Question 16	-1.94	Easy Question
Question 17	0.68	Medium Question
Question 18	-1.00	Easy Question
Question 19	-1.29	Easy Question
Question 20	0.55	Medium Question

From the table above it can be concluded that of the 20 questions given regarding information literacy for prospective teacher students, there were 2 questions with a difficult level of difficulty (10%), 7 items with a medium level of difficulty (35%), and 11 items with a level of difficulty. easy (55%). For difficult questions, there may be three follow-up actions, namely that the question item is no longer used and is not removed again in the next test results; reviewed, researched and traced so that the factors causing the participant's failure to answer the question can be identified; Difficult question items should have the benefit of being able to use the test items in a loose selection test, in the sense that most of the testees will be declared to have passed the selection test (Dalimunthe et al., 2021). In this condition, providing easy questions will provide an opportunity for many testees to pass the selection test or exam being held (Magdalena et al., 2021). (Manfaat & Nurhairiyah, 2021)provides an explanation of the factors that cause questions to be categorized as difficult, namely the question item "may" have the wrong answer key, the question item has 2 or more correct answers, the material has not been taught or the study has not been completed, the statement or sentence is too complex. So it is necessary to carry out further research regarding difficult questions (Amin & Harahap, 2023).

The reliability coefficient value is a measure of stability in measuring participants' abilities. The higher the estimated value of the instrument's reliability coefficient calculation indicates the higher the confidence in the participant's test results as measured using the instrument. (Manfaat &

Nurhairiyah, 2021)states that if the test reliability value obtained is high then the test developed can give the same results if given to the same group even by different people, different times or occasions, and different places so that consistency the test is considered high and trustworthy. The degree of confidence in the reliability of an instrument is expressed in the form of a reliability coefficient. In multiple choice questions, the reliability coefficient estimation formula commonly used is the KR-20 formula. By using KR 20, reliability estimates are calculated based on the number of scores of students who answered correctly or the correct score method (Nusantari, 2016). In this study, the KR-20 value obtained was 0.740, so it is included in the minimum criteria (Whitley, 2002)or can be said to be sufficient.

In Rasch modeling, data is also known regarding the reliability of items and people which is shown in the image below

```
SCALE QUALITY STATISTICS
=================================================
Statistic                   Items      Persons
-------------------------------------------------
Observed Variance           0.9497     1.0566
Observed Std. Dev.          0.9745     1.0279
Mean Square Error           0.0259     0.2927
Root MSE                    0.1608     0.5410
Adjusted Variance           0.9239     0.7639
Adjusted Std. Dev.          0.9612     0.8740
Separation Index            5.9764     1.6156
Number of Strata            8.3019     2.4875
Reliability                 0.9728     0.7230
=================================================
```

Based on the table, it can be seen that the Person Reliability value of 0.72 is included in the sufficient category. The person reliability value between the KR-20 formula and Rasch modeling is not much different, indicating that the consistency of students' answers when answering questions regarding information literacy is consistent. Based on the Item Reliability value of 0.97 in the good category, it can be concluded that the quality of the items in the instrument has a good reliability aspect (Adawiyah et al., 2023).

CONCLUSION

From the analysis results obtained, a valid and reliable instrument for measuring the information literacy of prospective teacher students has been produced. Initially the questions consisted of 22 items, then validation was carried out by experts. There were 20 CVR values above 0.99, so 2 questions were eliminated because they were invalid. The results of the content validity test involving 5 experts had a CVI value of 0.93, this result is said to be very good and valid for use at the next stage. Then, in the next stage, the Rasch model is used to ensure that the data obtained

is accurate, objective and consistent because measurements using the Rasch model can describe interactions between respondents and statement items. Based on the results of this research, the student information literacy instrument created is valid and reliable with a person reliability value of 0.72, proving that the consistency indicator of respondents' answers is at a fairly good level and the item reliability value is 0.96, which shows that the item quality indicators in the instrument is a good level. So that the information literacy instrument can be used to see the level of information literacy possessed by prospective teacher students, later it can also be used for self-development to improve their information literacy skills.

REFERENCES

ACRL, B. (2019). Information Literacy Framework for Higher Education. *Library Waves*, 36.

Adawiyah, R., Kholizah, S., & Harahap, R. D. (2023). *Literature Study : Utilization of Audio-Visual Media in Science Subject Matter for Junior High Sch ool Students*. *May*, 68–72.

Aksu, G., Güzeller, C. O., & Eser, M. T. (2019). Jmetrik: Classical test theory and item response theory data analysis software. *Journal of Measurement and Evaluation in Education and Psychology*, *10*(2), 165–178. https://doi.org/10.21031/epod.483396

Alam, U. F. (2013). Kemampuan Literasi Informasi Mahasiswa Dan Peranan Perpustakaan Dalam Proses Belajar Mengajar Di Perguruan Tinggi. *Pustakaloka*, *5*(1), 92–105.

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, *15*(2), 214–221. https://doi.org/10.1016/j.sapharm.2018.03.066

American Library Association. (2000). *Information Literacy For Higher Education*.

Amin, M. T., & Harahap, R. D. (2023). Analysis of students learning interest in biology subject matter of viruses in grade X senior high school. *BIO-INOVED : Jurnal Biologi-Inovasi Pendidikan*, *5*(2), 179. https://doi.org/10.20527/bino.v5i2.15549

Arikunto, S. (2010). *Prosedur Penelitian Suatu Pendekatan Praktik* (Edisi Revi). Rineka Cipta.

Bashooir, K., & Supahar. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *22*(2), 219–230. https://doi.org/10.21831/pep.v22i2.20270

Beile, P. M. (2007). The ILAS-ED: A standards-based instrument for assessing pre-service teachers' information literacy levels. *Society for Information Technology & Teacher Education International Conference*, *2007*(1), 1–27.

Dalimunthe, R. R., Harahap, R. D., & Harahap, D. A. (2021). Analisis Minat Belajar Siswa Sekolah Dasar Terhadap Mata Pelajaran IPA Pada Masa Pandemi Covid-19. *Jurnal Basicedu*, *5*(3), 1341–1348. https://doi.org/10.31004/basicedu.v5i3.888

Erfan, M., Maulyda, M. A., Hidayati, V. R., Astria, F. P., & Ratu, T. (2020). Tes Klasik Dan Model Rasch. *Indonesian Journal of Educational Research and Review*, *3*(1), 11–19.

Eynon, A. (2013). Welsh Information Literacy Project. *Library and Information Research*, *37*(114), 17–22.

Fatimah, L. U., & Alfath, K. (2019). Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor. *Jurnal Komunikasi Dan Pendidikan Islam*, *8*(2), 37–64.

Fitriyanto, J. N., Widjanarko, D., & Khumaedi, M. (2019). Validity and Reliability Test of Assessment Instrument of the Suitability of Electric Power Steering Media. *Journal of*

Vocational Career Education (Jvce), *4*(1), 61–69.

Hanbidge, A. S., Tin, T., & Sanderson, N. (2018). Information literacy skills on the go. *Journal of Information Literacy*, *12*(1), 118. https://doi.org/10.11645/jil.v12i1.2322

Ibnu, M., Indriyani, B., Inayatullah, H., & Guntara, Y. (2019). Aplikasi Rasch Model: Pengembangan Instrumen Tes untuk Mengukur Miskonsepsi Mahasiswa. *Prosiding Seminar Nasional Pendidikan FKIP*, *2*(1), 205–210.

Indrasvari, M., Harahap, R. D., & Harahap, D. A. (2021). Analysis of the Impact of Smartphone Use on Adolescent Social Interactions During COVID-19. *Jurnal Penelitian Pendidikan IPA*, *7*(2), 167–172. https://doi.org/10.29303/jppipa.v7i2.622

Julianti, C., Harahap, R. D., & Safitri, I. (2022). The use of multimedia in biology learning: MAS Subulussalam Sumberjo student responses. *BIO-INOVED : Jurnal Biologi-Inovasi Pendidikan*, *4*(3), 306. https://doi.org/10.20527/bino.v4i3.13812

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575.

Livingston, S. . (2018). Test Reliability—Basic Concepts. In *Research Memorandum No. RM-18-01*. Princeton, NJ: Educational Testing Service.

Magdalena, I., Nurul Annisa, M., Ragin, G., & Ishaq, A. R. (2021). Analisis Penggunaan Teknik Pre-Test Dan Post-Test Pada Mata Pelajaran Matematika Dalam Keberhasilan Evaluasi Pembelajaran Di Sdn Bojong 04. *Jurnal Pendidikan Dan Ilmu Sosial*, *3*(2), 150–165.

Manfaat, B., & Nurhairiyah, S. (2021). Pengembangan Instrumen Tes untuk Mengukur Kemampuan Penalaran Statistik. *Jurnal Pembelajaran Matematika*, *1*(1), 1–19.

Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Nuha Medika.

Mohajan, H. K. (2017). Two Criteria for Good Measurements in Research: Validity and Reliability. *Annals of Spiru Haret University*, *17*(3), 58–82.

Mok, M., & Wright, B. (2004). *Overview of Rasch Model Families. In Introduction to Rasch Measurment: Theory, Models and Application (hal : 1-24)*. Jam Press.

Muntashir. (2016). Standar Kompetensi Literasi Informasi Mahasiswa Ilmu Perpustakaan Pada Perguruan Tinggi Agama Islam. *JIPI (Jurnal Ilmu Perpustakaan Dan Informasi)*, *1*(1), 102–124.

Nurfadillah, T., Elvia, R., & Elvinawati. (2023). Pengembangan Instrumen Tes Kimia Berbasis Literasi Sains untuk Mengukur Literasi Sains Siswa. *Alotrop*, *7*(1), 44–56. https://doi.org/10.33369/alo.v7i1.28253

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

Prihandoko, L. (2021). Self-Assessment on Information Literacy Competency of Students in Indonesia-Papua New Guinea Border Area During The COVID-19 Epidemic. *ICOPE*. https://doi.org/10.4108/eai.16-10-2020.2305235

Puger, I. G. N. (2021). PENGUJIAN VALIDITAS ISI TES HASIL BELAJAR YANG DINILAI OLEH SUBJECT MATTER EXPERT (SME). *Daiwi Widya*, *8*(3).

Puspitasari, W. D., & Febrinita, F. (2021). Pengujian Validasi Isi (Content Validity) Angket Persepsi Mahasiswa terhadap Pembelajaran Daring Matakuliah Matematika Komputasi. *Journal Focus Action of Research Mathematic (Factor M)*, *4*(1), 77–90. https://doi.org/10.30762/factor_m.v4i1.3254

Ranaweera, P. (2020). Importance of Information Literacy skills for an Information Literate society CORE View metadata, citation and similar papers at core.ac.uk provided by E-LIS. *Senior Lecturer, National Institute of Library & Information Sciences, University of Colombo*, 1–13.

Retnawati, H. (2016). *Validitas reliabilitas dan karakteristik butir: (panduan untuk peneliti, mahasiswa, dan psikometrian)*. Parama Publishing.

Segal, D. ., & Coolidge, F. . (2018). *Reliability*. SAGE Publications. https://doi.org/http://dx.doi.org/10.4135/9781506307633.n683

Sudijono, A. (2011). *Pengantar Evaluasi PEndidikan*. Rajagrafindo persada.

Taherdoos, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management (IJARM)*, *5*(3), 28–36.

Wati, M., & Mahtari, S. (2017). Pengembangan Instrumen Kognitif Fisika Siswa SMP. *Jurnal Penelitian Dan Pengkajian Ilmu Pendidikan: E-Saintika*, *1*(1), 45. https://doi.org/10.36312/e-saintika.v1i1.5

White, A. M. J. (2019). *Handbook of Research on Critical Thinking and Teacher Education Pedagogy*. Handbook of Research on Critical Thinking and Teacher Education Pedagogy: Global. https://doi.org/10.4018/978-1-5225-7829-1.ch020

Whitley, B. E. (2002). *Principals of Research and Behavioural Science*. McGraw-Hill.

Zaenab, S., Khasanah, N., & Salimi, M. (2017). Literasi Informasi Untuk Mahasiswa Kependidikan. *Universitas Sebelas Maret*, *2003*, 260–265.

Zeeshan, M., Siddique, N., & Idrees, H. (2020). Measuring Information literacy skills of graduate students of Lahore University of Management Sciences (LUMS). *Library Philosophy and Practice*, *2020*.