

DEVELOPMENT OF GOOGLE FORM-BASED LEARNING OUTCOME EVALUATION INSTRUMENTS FOR ANIMAL DEVELOPMENT COURSES IN THE BIOLOGY EDUCATION STUDY PROGRAM FMIPA UNIMED

Nurmaya Teresia Simanullang¹ and Herbert Sipahutar^{1*}

¹ *Biology Education Study Program, Department of Biology, Faculty of Mathematics and Natural Sciences, Medan State University, Jalan Williem Iskandar Pasar V Medan Estate, North Sumatra 20221, Indonesia*

*Corresponding Author: herbert_sipahutar@yahoo.com

ARTICLE INFO:

ABSTRACT

Article History

Received December 20th, 2023

Revised December 25th, 2023

Accepted December 30th, 2023

Keywords:

Learning outcome evaluation instrument, google form, validity, reliability, readability test

This study aimed to develop a valid and reliable google form-based learning outcome evaluation instrument for the Animal Development course. The instrument was developed following the 4D development model (define, design, develop, disseminate). The research subjects were Biology Education Study Program (PSPB) class of 2021 class B students totaling 26 people. Data were obtained from validation questionnaires and student response questionnaires. The initial design of the learning outcome evaluation instrument consisted of 50 questions. Based on the validator's assessment, an instrument consisting of 25 valid and reliable questions was obtained. The results of validation by evaluation instrument experts obtained an average percentage of 98.6% with a very feasible category. Based on the validity analysis of the Aiken index, the value $V = 0.942$ was obtained with high validity criteria. Instrument reliability using Cronbach's Alpha obtained a value of 0.796 with high criteria (reliable), while the analysis of ICC reliability data obtained the results of inter-rater agreement of 0.796, and for each rater the consistency is 0.565. It can be concluded that the agreement between raters is very strong and each rater has a fairly good consistency. The results of the data analysis of student responses showed that each indicator was rated very well with an average percentage of 88%.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

How to Cite:

Manullang, N.T., Sipahutar, H. (2023). Development Of Google Form-Based Learning Outcome Evaluation Instruments For Animal Development Courses In The Biology Education Study Program Fmipa Unimed. *Jurnal Pelita Pendidikan*, 11(4)(2023), 121-130.

INTRODUCTION

Education is a series of activities that take place in various places with the aim of developing knowledge, understanding, assessment, growing, caring and behaving (Chazan, 2022). In practice, education requires several devices with the aim of obtaining learning outcomes and assessments for students. One of the benchmarks used to determine the level of success of students in the learning process is evaluation. Ralph Tyler said that evaluation is a process of collecting data to see to what extent, what and how the educational goals have been achieved (Nurfillaili et al., 2016).

Evaluation is a systematic and continuous process to determine the quality (value and meaning) of something, which is based on certain considerations and criteria used in making decisions in order to collect information about learning outcomes from students / students (Hadijah & Anggereni, 2016). Evaluation is one component in the learning system that must be carried out in an effort to provide quality learning (Merdekawati, 2017). Learning evaluation will provide information about student gains related to achieving learning objectives and how effective learning is taking place.

According to Ropii & Fahrurrozi (2017), the evaluation of learning outcomes aims to: 1) see the level of mastery of students related to the material that has been given, 2) see the skills, motivation, interests, talents, and attitudes of students towards the learning program provided, 3) see the level of progress and suitability of student learning outcomes to the established competency standards and basic competencies, 4) examine the advantages and disadvantages of students in following the process learning, 5) selecting and determining students who are in accordance with certain types of education, 5) determining class advancement, and 7) to place students based on their potential.

Generally, the evaluation instrument used to measure knowledge and mastery of certain content or material is in the form of tests. Evaluation of learning outcomes is usually carried out after the learning material has been discussed. Conventionally, the evaluation is carried out in the form of a written test using question sheets. Then, the students' answer sheets are checked to find out the results of the evaluation of the learning process. This will take a lot of time and energy. The rapid development of technology and information provides many benefits and conveniences in various fields, especially in the field of education (Utami, 2021).

The results of observations on students of the Biology Education Study Program (PSPB) class B class of 2021 found data that the evaluation carried out still did not utilize technology, namely still using question sheets and answers. The next acquisition data is that students prefer if the evaluation of learning outcomes is carried out online using *google forms* because it is more effective and easy to use. Based on these data, researchers want to develop a google form-based learning outcome evaluation instrument.

The current development of ICT, one of which is smartphones, offers various applications that can be used to assess *online* learning. One that can be used as a web/online-based learning evaluation tool is *Google Form* which is part of the *Google Docs* component provided by Google as software that can be accessed for free and is quite easy to use (Meirawati et al., 2021). Along with the rapid development of ICT, it is expected to reduce excessive use of paper. With the change from paper-based evaluation (paper-based test) to *computer-based* (computer-based test). The application of a computer-based evaluation system (*computer-based* test) by utilizing *google forms* is quite efficient and makes it easier to assess learning / test results, and is more practical to use because it can be accessed using a *smartphone* anywhere as long as the internet network can be reached and can be carried everywhere (Alwiyah et al, 2020).

Google forms is a tool that can be used to help plan events, send surveys, create quizzes, and collect data or information quickly and efficiently. Forms can be linked to spreadsheets, if the responses needed are related to forms and responses. If you want to make it easier, users can see a summary of responses that can be accessed on the response menu page (Utami, 2021).

The *Google Form application* is very suitable for students, teachers, lecturers, office employees, and professionals who like to make quizzes, forms and online surveys. *Google forms* can be accessed openly or exclusively by using accessibility options, such as *read-only* or *editable*. *Google form* is an application that is not paid or free, so it can be accessed at all levels with the condition that you have a universal Google account. Because this application is web-based, everyone can quickly respond or answer to quizzes or questionnaires wherever they are using computer/laptop internet applications or SmartPhone (Coal, 2016).

Batubara (2016) states that *google forms* in the world of education function, among others: 1)

provide online practice / test assignments in the form of website pages, 2) collect other people's opinions through website pages, 3) *collect data on students / educators through website pages*, 4) create *online* registration forms for schools, 5) Distribute questionnaires to people *online*. The use of *google forms* as a learning evaluation tool saves the use of paper as a form of environmental care. In addition, the energy and time needed to distribute questionnaires and process data becomes more efficient and easier.

Evaluation of learning outcomes based on *google form*, namely evaluation by utilizing technology carried out online using an internet network connection, students can evaluate anywhere and anytime, and lecturers are able to set time limits for the evaluation work given. Compared to evaluation using paper and pen, evaluation using *google form* is more practical.

Technology-based evaluations using *google forms* can be timed and questions can be randomized. This will make students focus more on the questions and have no opportunity to cheat on their friends, so it will minimize cheating when the evaluation is carried out. After the evaluation, the result score will immediately appear and this will minimize the occurrence of answer correction errors so that the results can be seen quickly and accurately (Rachmawati & Kurniawati, 2020).

The general purpose of this study is to develop a *google form*-based learning outcome evaluation instrument in the hope that it can help and facilitate the process of assessing learning outcomes. The specific objectives of this study are to: 1) know the design of learning outcomes evaluation instruments based on *google form* of the Animal Development course, 2) know the validity of the evaluation instruments developed, 3) know the reliability of the evaluation instruments developed, 4) know student responses to the instruments developed.

METHOD

This research was carried out at the Biology Education Study Program, Faculty of Mathematics and Natural Sciences, State University of Medan (UNIMED) located on Jl. William Iskandar Psr IV, Medan, North Sumatra.

The type of research is a 4-D development model (define, design, develop, disseminate). In this research only up to the development stage (develop).

The subjects of this study were students of the Biology Education Study Program (PSPB) 21-B which amounted to 26 students.

Development with a 4D model was chosen because each stage in this model has detailed and clear steps for each stage to make a product. Research is carried out until the development stage only because of limited time from researchers and researchers will only produce products in the form of learning outcome evaluation instruments that have been tested for feasibility by experts. The 4D research procedure (define, design, develop) is as follows.

1. Defining Stage (Define)

The define stage is aimed at determining and defining the needs in the learning process and collecting various information related to the product to be developed. At this stage, needs analysis activities are carried out to determine student needs for *google form*-based learning outcome evaluation instruments, material analysis to determine parts of the material learned in animal development learning by referring to RPS (Semester Learning Plan). After knowing the material and scope of what study material is studied in animal development, it will then be adjusted to the making of learning indicators, questions, and analysis of learning achievement indicators to see the material for each meeting so that it can be adjusted to the cognitive level of the test to be developed.

2. Design Stage

The design stage aims to obtain the design of the learning outcome evaluation instrument developed. At this stage, a grid of learning outcome evaluation instruments is made based on data obtained from the defining stage (define). Question development is tailored to student needs, learning materials, and learning outcome indicators. The number of questions developed consists of 50 questions. The preparation of these 50 questions is adjusted to the instrument grid that has been made based on the CPMK from the RPS animal development course. Questions are made as many as 2 points from each indicator so that 50 questions are obtained. Of the total 50 questions developed, the questions that will be used will be 25 questions that represent each indicator based on the instrument grid made. The question will be validated and corrected based on suggestions from validators until the question is said to be valid. These 25 questions will be selected based on the validation results of 3 animal development lecturers with the highest scores. The results at this stage of design are called preliminary drafts (draft I).

3. Development Stage

The learning outcome evaluation instrument that has been developed consists of 50 questions first validated by expert lecturers on evaluation questions to see whether the questions developed are suitable or not for use.

The next stage is instrument validation by 3 course lecturers. Based on the assessment of the three lecturers, 25 questions were selected to be used. The selection is based on the highest total score after summing the scores of the three validators. The question items are then tested for validity based on assessments between raters (validators).

Reliability tests are carried out after data on the validity of learning outcome evaluation instruments are obtained. The results of the reliability test will see how the agreement between raters on 25 questions that have been validated. Reliability tests were analyzed using SPSS.

Test readability to students to see student responses to the google form-based learning outcome evaluation instrument that has been developed and to see whether the evaluation instrument can be read clearly and easily understood. Students are given questionnaires through google form to respond to the instruments developed.

The technique used to analyze the data in this study consists of three parts, namely instrument validity analysis, instrument reliability analysis and student response test analysis.

a. Instrument Validity Test

The data were analyzed using the Aiken index (Aiken, 1980), where each question item was assessed with the following calculation:

$$V = \frac{\sum s}{n(c - 1)}$$

Information:

S = r - lo

V : Rater deal index

r : The value given rater

Lo : Lowest score score

n : The number of raters

c : Highest score score

The results of the analysis using the Aiken index are summed up in the form of categorizing validity classifications (Table 1).

Table 1. Instrument Validation Assessment Criteria

Average rating	Validation Criteria
0.80 < rxy < 1.00	High (very decent)
0.60 < rxy < 0.80	Medium (decent)

0.40 < rxy < 0.60	Low (decent enough)
0.20 < rxy < 0.40	Very low (not worth it)

(Guilford, 1956)

The assessment sheet filled in by the evaluation expert is analyzed to determine the quality of the instrument using the formula:

$$P = \frac{f}{N} \times 100$$

Information:

P = Percentage number

f = Raw score obtained

N = Max score

The results of the analysis are categorized under the level of validity (Table 2).

Table 2. Validity Criteria

Achievement Criteria (%)	Validity Criteria
81 - 100	Very Worth It
61 - 80	Proper
41 - 60	Pretty Decent
21 - 40	Not Worth It
0 - 20	Very Unworthy

(Sudijono, 2015)

b. Instrument Reliability Test

The reliability test of each question item on the developed instrument was carried out based on the score given by the validator (rater) analyzed using Cronbach's Alpha SPSS and ICC (Interclass Correlation Coefficients).

The results of the analysis using Cronbach's Alpha SPSS and ICC concluded the level of reliability between the raters in the form of categorization / classification (Table 3).

Table 3. Instrument Reliability Criteria

Level of reliability	Category
> 0.75	Excellent
0,60 – 0,75	Good
0,40 – 0,60	Enough
< 0.40	Bad

(Fleiss, 1975)

c. Student Response Test

Eligibility criteria for the instrument based on student responses use the Likert scale which is calculated using the formula:

$$\%respon = \frac{\text{skor yang diperoleh}}{\text{skor maksimum}} \times 100$$

The students responses were concluded based on Table 4.

Table 4. Instrument Validity Criteria based on Response Test

Achievement Criteria (%)	Validity Level
81 - 100	Excellent
61 - 80	Good
41 - 60	Good enough
21 - 40	Bad
0 - 20	Very Not Good

(Arikunto, 2016)

RESULTS AND DISCUSSION

Research Results

1. Description of Learning Outcomes Evaluation Instrument

The learning outcome evaluation instrument developed is arranged based on the grid of questions that have been made. The grid is derived from the Semester Learning Plan (RPS) developed by lecturers, covering learning topics, material scope, indicators, cognitive levels and question numbers. The final product of the development is an instrument consisting of 25 question points adjusted to the CPMK indicator. In each indicator, 2 questions were developed, so that draft I of the learning outcome evaluation instrument consisted of 50 questions.

The instrument is first validated by 3 lecturers who teach courses and experts on evaluation questions. After validation and the instrument is improved according to the suggestions and input provided by the validator, a draft II of the learning outcome evaluation instrument consisting of 50 corrected questions is obtained. From the validation results, an assessment sheet was obtained from the three validators for each question item. Then, from the assessment sheets of the three validators, 25 questions with the highest scores were selected from the accumulated assessment data of the three validators obtained. Furthermore, the 25 questions become the final questions that will be used (draft III) and tested to their validity and reliability.

From draft I consisting of 50 questions, it will be validated by 3 animal development lecturers and improved according to suggestions until each question item is valid, so that draft II consists of 50 questions that have been revised several times and have been corrected. After that, a draft III consisting of 25 questions was obtained which was selected based on the highest score from the assessment given by validators representing each indicator. Draft III of the learning outcomes evaluation instrument consisting of 25 questions was validated using Aiken's index in Microsoft Excel. Aiken's validation aims to see the feasibility of each question item

based on the assessment agreement between raters (validators). Then continued with reliability tests of Cronbach's Alpha and Interclass Correlation Coefficients (ICC) value analysis at SPSS. This reliability test is carried out to see the level of agreement between raters (validators) in assessing each question item on the learning outcome evaluation instrument. After the validity test and reliability test, the 25 questions were tested for readability by giving questionnaires to students.

2. Define Stage

The questionnaire was distributed through google form to 26 students of PSPB class 21-B to find out the basic needs in developing products. From the first question given, which is the question of evaluating animal development given by the lecturer in what form. 100% of students answered that the question work used question sheets and answers given by the lecturer.

The next acquisition data is that students prefer if the work on the evaluation questions for animal development courses is carried out online. As many as 80.8% of students chose answers to work on online evaluation questions.

The material used as a reference to develop animal development evaluation instruments is seen from the RPS (Semester Learning Plan). Indicators of each question are developed evaluation instruments that have been adjusted to the topic and scope of the study material. From obtaining these data, researchers developed a google form-based learning outcome evaluation instrument for animal development courses.

3. Design Stage

The products produced at this stage are a grid of questions and essay-shaped question items. The test instrument is designed based on the scope of the study material and the learning outcomes of the animal development course.

The grid of learning outcome evaluation instruments consists of learning topics, material scope, learning achievement indicators, cognitive levels and question numbers.

The questions developed are 50 items, but the final questions that will be used are as many as 25 questions after comparing the validation results of 3 animal development lecturers. The questions were first validated by expert lecturers, the evaluation questions followed by the validation of instrument feasibility by 3 lecturers who taught animal development courses.

At this stage, draft I of the learning outcome evaluation instrument was produced which amounted to 50 questions.

4. Development Stage

a. Valid test resultstop

At this stage, draft I of the learning outcome evaluation instrument which amounted to 50 questions was first validated by expert lecturers on evaluation questions, followed by instrument validation by 3 lecturers who taught animal development courses.

The results of the feasibility calculation of the evaluation questions show that the learning outcome evaluation instrument is very feasible to use. With an average percentage obtained is 96.8% with very feasible criteria. After validation by experts, the evaluation questions continued with instrument validation by lecturers who teach animal development courses.

The validation results from 3 animal development lecturers obtained 50 questions that have been corrected based on suggestions and input given by validators (draft II).

The assessment of the validator is then processed by summing the total score of each indicator given by the validator for each question item. From the summation data, 25 questions will be selected based on the highest number of scores given by validators. All 25 questions will be the final questions obtained after validation (draft III).

The results of the expert judgment for 25 questions were analyzed quantitatively using Aiken's formula in Microsoft Excel. Validator assessments use a scale of 1 to 4. The results of validity data analysis can be seen in Table 5.

Table 5. Aiken Index Validity Data Analysis Results

Question Point	Validators			S1	S2	S3	Σs	V	Information
	I	II	III						
Items 1-25	93	97	97	68	72	72	212	0,942222	Tall

Table 5 shows that the average value of Aiken's coefficient for each question item is 0.942. Based on the classification of validity proposed by Guilford, it is concluded that the validity of learning outcome instruments with a total of 25 questions has high validity.

b. Reliability test results

The results of the inter-rater reliability test (validator) were analyzed using Cronbach's Alpha analysis in SPSS, the results are presented in Table 6.

Table 6. Reliability of Cronbach's Alpha Item Questions in the Instrument

Reliability Statistics	
Cronbach's Alpha 0,796	N of Items 3

Table 6 shows that Cronbach's Alpha value is 0.796. This shows an average agreement between raters of 0.796. Based on the reliability coefficient table, it is obtained that if the value of Cronbach's Alpha is 0.796, it can be concluded that this instrument is reliable with high reliability criteria.

The results of the interrater agreement test or interrater reliability were analyzed using interclass correlation coefficients (ICC) with SPSS. The results of the analysis are shown in Table 7.

Table 7. Results of Interclass Correlation Coefficients (ICC) Reliability Data Analysis

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.565a	.336	.757	4.896	24	48	0,000
Average Measures	.796c	.603	.903	4.896	24	48	0,000

Table 7 shows the results of the ICC, the average result of agreement between raters is 0.796, while for each rater the consistency is 0.565. Based on the reliability classification put forward by Fleiss (1975) in accordance with the ICC

results obtained, it was concluded that the agreement between raters was very strong and each rater had a fairly good consistency.

c. Readability test results

The evaluation instrument that has been validated and has been tested for reliability consists of 25 questions. Furthermore, the questions will be tested for readability to students by giving questionnaires of responses to students.

Data on the calculation of student response questionnaires to google form-based learning outcome evaluation instruments can be seen in Table 8.

Table 8. Data Analysis of Student Response Questionnaire to Google Form-Based Learning Outcome Evaluation Instrument

No	Assessment Aspect	Criteria (%)	Information
Highlights			
1	In my opinion, evaluation of animal development learning outcomes based on google form is more effective and efficient	87	Excellent
2	In my opinion, the visual appearance (photos, images, etc.) in this google form-based evaluation is very interesting.	87	Excellent
3	I prefer to do evaluation questions using google forms.	93	Excellent
Content Quality			
4	The questions presented are in accordance with the material I studied.	84	Excellent
5	The question material presented is easy to understand.	86	Excellent
6	This evaluation of learning outcomes can test how far I understand the Animal Development course.	88	Excellent
Language			
7	The sentences used in the questions are clear and easy to understand.	91	Excellent
8	The language used in evaluating the learning outcomes of the Animal Development course is simple and easy to understand	86	Excellent
9	The foreign language (Latin) in the question material is in accordance with the norms used in biology	85	Excellent
Ease			
10	I think this Google Form application is easily accessible anytime and anywhere.	92	Excellent
11	This Google Form application makes it easier for me to evaluate the learning outcomes of the Animal Development course.	91	Excellent
Average		88	Excellent

Table 8 shows that each indicator is rated very well with an average percentage gain of 88%. From these data, it can be concluded that the readability test against the google form-based evaluation instrument for animal development courses is very good.

Discussion

The research result instrument, initially consisting of 50 questions, the feasibility was first assessed by expert lecturers on evaluation questions and data was obtained that the questions were very feasible to use with an average score of 98.6%. After validation by experts on evaluation, the instrument was then assessed for feasibility by 3 animal development lecturers. After several revisions until a suitable instrument was obtained, from the assessment of the three validators, the question item that had the highest score of 25 questions was selected. The 25

questions represent each learning outcome. Based on the assessment of the three validators on the 25 question items, the validity of each question item was tested using the Aiken index.

An instrument is said to be valid if the results of data analysis conform to predetermined criteria. Arikunto (2016), said that an instrument is said to be valid if the results of the validity of the instrument are in accordance with the criteria, meaning that the test results with predetermined criteria have parallel.

Based on the validity data of Aiken that has been done, more details can be seen in table 5. Question items 1-25 based on validation conducted by 3 validator lecturers concluded that the questions had high validity with an average gain of $V = 0.942$. In line with research conducted by Nabil et al. (2022), the results of the Aiken index analysis on 20 items of AKM (Minimum competition assessment) numeracy literacy instrument questions obtained data that 20

questions were valid with a value of $V > 0.75$. The Aiken V index is a formula for calculating the content-validity coefficient based on the results of expert assessments of n people on an item and in terms of the extent to which the item represents the measured construct (Aiken, 1985). This formula was used in this study to see the validity of the instrument based on the assessment results of 3 validators.

Research conducted by Putri et al. (2020), the data from validation using Aiken's index is valid with $V = 0.931$. Data was obtained from the 15 questions developed, all questions were said to be valid (usable) based on the results of Aiken's index analysis using rater assessment (validator). In this study the value of V'Aikens is seen based on the standard table compiled by Aiken (1985) using 4 rating categories and 3 raters on 25 questions, then the minimum standard of Aiken's V for this study is 0.92. So it can be concluded that all items of the learning outcome evaluation instrument for the animal development course are valid in content or this learning outcome evaluation instrument has high validity. The closer the Aiken index value is to 1, the better an item is because it is more relevant to the indicator (Retnawati, 2016).

Instrument reliability is to see the extent to which a measurement can display consistent measurement results if repeated measurements are made using the same measuring instrument. The reliability of a measuring instrument is seen from how the measuring instrument gives consistent and stable results when repeated measurements are made with different times and the same object (Arikunto, 2013). An instrument is said to have a high level of confidence if it provides consistent results.

Reliability tests are performed to see the degree of consistency of the instrument and to determine the extent to which the measuring instrument is reliable and reliable. Based on the assessment given by validators (raters), reliability tests between raters were carried out to see the level of agreement between validators in assessing each question item in instruments developed using Cronbach's Alpha SPSS and ICC (Interclass Correlation Coefficients).

The result of Cronbach's Alpha reliability test on the question item based on rater assessment is 0.796. If the result of Cronbach's Alpha value is 0.796, it can be concluded that this instrument is reliable with high reliability criteria. Cronbach's Alpha is a measure of reliability with a value range of 0-1.00 (Tomoliyus & Sunardianta, 2020). The minimum value of Cronbach's Alpha

reliability level is 0.70 (Fleiss et al., 2004) if Cronbach's Alpha value is > 0.7 then the google form-based learning outcome evaluation instrument for animal development courses is very reliable.

Meanwhile, to test the agreement between raters and the consistency of each rater, ICC (Intraclass Correlation Coefficients) analysis is used. ICC is used because the number of raters is more than two. In other words, more than two raters assess the google form-based learning outcome evaluation instrument for animal development courses through rating instruments. The result of agreement between raters is 0.796, while for each rater the consistency is 0.565. Based on the reliability classification put forward by Fleiss (1975) in accordance with the ICC results obtained, it was concluded that the agreement between raters was very strong and each rater had a fairly good consistency.

In line with research conducted by Tomoliyus & Sunardianta (2020), the analysis result of Cronbach's Alpha reactive agility test instrument is 0.875. According to Tavakol & Dennick (2011), the minimum value of Cronbach's Alpha reliability level is 0.70, so if Cronbach's Alpha > 0.7 then the aspect of the reactive agility test instrument is very reliable. As for the results of ICC analysis shows that the average agreement between raters is 0.875 and for one rater the consistency of each is 0.500.

The learning outcome evaluation instrument whose initial design consisted of 50 questions was then validated and has been improved several times based on suggestions from validators, 25 questions were selected based on the highest score given by validators. 25 questions that have been tested for validity and reliability are then compiled in the form of a google form to be distributed to students and tested for readability by providing response questionnaires.

The results of the analysis of readability test data showed that each indicator was rated very good with an average acquisition percentage of 88%. From these data, it can be concluded that the readability test against the google form-based evaluation instrument for animal development courses is very good.

Research conducted by Multin et al. (2018), the results of the instrument readability test have a percentage of 77.5% with a fairly good category. The readability test results are quite good because they are included in the high category. However, improvements are still needed for very high readability results. This shows that the display of the performance instrument is easy to practice and

can be used by anyone. In line with research conducted by Rachmawati & Kurniawati (2020), the average readability assessment of learning outcome evaluation instruments was obtained 80% in the good category. Based on the repon questionnaire, it can be said that online-based evaluation is effective and practical.

Based on the acquisition of validity data, reliability data and instrument readability data, for future research this instrument can be tested for effectiveness by disseminating it to students and assessing learning outcomes.

CONCLUSION

Based on the results of the research and data analysis that has been carried out, it can be concluded that: The initial design of the learning outcome evaluation instrument consisted of 50 questions. Based on the validator's assessment, an instrument consisting of 25 valid and reliable questions was obtained. The results of validation by evaluation instrument experts obtained an average percentage of 98.6% with a very feasible category. Based on the validity analysis of the Aiken index, the value $V = 0.942$ was obtained with high validity criteria. Instrument reliability using Cronbach's Alpha obtained a value of 0.796 with high criteria (reliable), while the analysis of ICC reliability data obtained the results of inter-rater agreement of 0.796, and for each rater the consistency is 0.565. It can be concluded that the agreement between raters is very strong and each rater has a fairly good consistency. The results of the data analysis of student responses showed that each indicator was rated very well with an average percentage of 88%.

REFERENCES

- Aiken, L.R. (1980). Content Validity and Reliability of Single Items or questionnaires. *Educational and Psychological Measurement*, 40(4):955-959. <https://doi.org/10.1177/001316448004000419>
- Aiken, L. R. (1985). Three coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1):131-142. <https://doi.org/10.1177/0013164485451012>
- Alwiyah, U., Syamsudduha, & Ali, A. (2020). Pengembangan Instrumen Penilaian kognitif Berbasis Google Form pada Mata Pelajaran Biologi. *Jurnal Al-Ahya*, 2(3): 129-148.
- Arikunto, S. (2013). *Prosedur Penelitian Suatu Pendekatan Praktek*. Jakarta: Rineka Cipta.
- Arikunto, S. (2016). *Manajemen Penelitian*. Jakarta: Rineka Cipta.
- Batubara, H. H. & Ariani, D. N. (2016). Workshoop Penggunaan *Google Form* sebagai Media Evaluasi Pembelajaran pada Dosen-dosen Fakultas Studi Islam. *Jurnal Al-Ikhlash*, 2(1): 39-44.
- Batubara, H. H. (2016). Penggunaan *Google Form* sebagai Alat Penilaian Kinerja Dosen di Prodi PGMI UNISKA Muhammad Arsyad Al Banjari. *Al-Bidayah: Jurnal Pendidikan Dasar Islam*, 8(1): 39-50.
- Chazan, B. (2022). *What Is "Education"?*. In: *Principles and Pedagogies in Jewish Education*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-83925-3_3
- Fleiss, J. L. (1975). Measuring Agreement Between Two Judges on the Presence or Absence of a Trait. *Biometrics*, 31(3):651. <https://doi.org/10.2307/2529549>
- Fleiss, J. L., Levin, B., & Paik, M. C. (2004). The Measurement of Interrater Agreement. In *Statistical Methods for Rates and Proportions* (pp.598-626). Jhon Wiley & Sons, Inc. <https://doi.org/10.1002/0471445428.ch18>
- Guilford, J. P. (1956). *Fundamental Statistics in Psychology and Education* (3rd ed). McGraw-Hill.
- Hadijah & Anggereni, S. (2016). Pengembangan instrumen tes hasil belajar kognitif mata pelajaran fisika pada pokok bahasan momentum dan impuls SMA kelas XI. *Jurnal Pendidikan Fisika*, 4 (1):30-34.
- Meirawati, E., Azis, A., & Rohmadi, M. (2021). Pemanfaatan *Google Form* sebagai Alat Evaluasi Pembelajaran Mata Pelajaran Pendidikan Agama Islam di SMKN 2 Palangka Raya. *Muallimun: Jurnal Kajian Pendidikan dan Keguruan*, 1(1): 15-26.
- Merdekawati, K. (2017). Pengembangan Instrumen Penilaian Untuk Kerja Praktikum Kimia Fisika. *Jurnal Penelitian Pendidikan*, 20(1): 24-32.
- Multin, H. A., Munawar, W., & Noor, R. A. M. (2018). Penyusunan dan Analisis Tes Kinerja (Performance Test) pada Kompetensi pRaktik Memasang Sistem Penerangan dan Wiring Kelistrikan di SMK. *Journal of Mechanical Engineering Education*, 5(2): 164-172.

- Nabil, N. R. A., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis Indeks Aiken untuk Mengetahui Validitas Isi Instrumen Assesmen Kompetensi Minimum Berbasis Konteks Sains Kimia. *Jurnal Penelitian pendidikan*, 25(2): 184-191.
- Nurfillaili, U., T. Yusuf, M., & Anggereni, S. (2016). Pengembangan Instrumen Tes Hasil Belajar Kognitif Mata Pelajaran Fisika Pada Pokok Bahasan Usaha Dan Energi SMA Negeri Khusus Jeneponto Kelas XI Semester I. *Jurnal Pendidikan Fisika*, 4 (2): 83-87.
- Rachmawati & Kurniawati, A. (2020). Pengembangan Instrumen Penilaian Tes Berbasis *Mobile Online* pada Prodi Pendidikan Matematika. *Prima: Jurnal Pendidikan Matematika*, 4(1): 46-63.
- Retnawati, H. (2016). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). *Research and Evaluation in Education*, 2(2): 155-164. doi:<http://dx.doi.org/10.21831/reid.v2i2.11029>
- Ropii, M. & Fahrurrozi, M. (2017). *Evaluasi Hasil Belajar*. Pancor Selong Lombok Timur, NTB: Universitas Hamzanwadi Press.
- Sudijono, A. (2015). *Pengantar Statistika Pendidikan*. Jakarta:Rajagrafindo.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Tomoliyus, T. & Sunardianta, R. (2020). Validitas dan reliabilitas Instrumen Tes Reaktif Agility Tenis Meja. *Jurnal Keolahragaan*, 8(2): 148-157.
- Utami, L. W. S. (2021). Penggunaan *Google Form* dalam Evaluasi Hasil Belajar Peserta Didik di Masa Pandemi Covid-19. *Jurnal Inovasi Keguruan dan Ilmu Pendidikan*, 1(3): 150-156.
- Z. Putri, R., Sari, R., Jumadi, & Ariswan. (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.