



Development of Performance Instruments to Assess Student Performance in Congregational Prayer Practices for Islamic Elementary School Grade II

Ernawati Ernawati^{1✉}, Saiful Ridlo², Zaim Elmubarak³

^{1,2,3}Universitas Negeri Semarang

Article Info

History Articles

Received:

09 May 2024

Accepted:

07 June 2024

Published:

30 August 2024

Keywords:

*Analysis , instruments,
performance*

Abstract

This research is a research and development of performance assessment instruments for class II of Islamic Elementary School students in the practice of congregational prayer on fiqh subjects. The research was conducted at Al-Qolam Islamic Elementary School , Nurul Ilmi Islamic Elementary School, and Sambinae Islamic Elementary School in Bima City. Research method using the Mix method, where the research subject was Islamic Elementary School teacher that teach subject Fiqh at class II.

The results showed that there were 3 instruments for assessing student performance in Congregational Prayer Practices, namely the performance instrument for the Priest which consisted of 23 statement items, the Main Congregation which consisted of 24 statement items and the Masbuk Congregation which consisted of 25 statement items. All instruments developed have been tested for validity, namely the performance instrument of Priest has an average V'aiken 0.87, the performance instrument of the Main Congregation is 0.85, and the performance instrument of the Masbuk Congregation is 0.86. While the reliability of the instruments analyzed using the inter-rater reliability (ICC) showed a reliable instrument with coefficient of reliability 0.79 for the performance instrument of Priests, 0.78 for the performance instrument of Main Congregation, and 0.73 for the performance instrument of Masbuk Congregation.

✉Correspondence Address :

Faculty of Graduate Studies, Semarang State University, Indonesia
50229

E-mail : wati87765@students.unnes.ac.id

p-ISSN 2252-6420

e-ISSN 2503-1732

INTRODUCTION

Assessment is an important aspect in the world of education because assessment is a systematic process of gathering information about the development of students in participating in learning. According to (Edi Istiyono, Djemari Mardapi, 2014). To monitor the process, progress in order to support continuous improvement of learning, an assessment is required. Acher describes the Assessment as an important process that can provide the information about students' progress specifically since the learning process and its information can be used to assess that the student can be continued to the next level (Guangul et al., 2020). According to Oerman (2018) Assessment not only the process to provide information about students' progress but also the process that can help teacher to describe the next goals of learning process (Immonen et al., 2019). According to Regulation of the minister of education No.20 (2007) Assessment is the process of collecting and processing information on student development to determine the level of achievement of students in the learning process.

Assessment is teachers responsibility to measure the outcome of learning that not only to describe the learning outcomes but also to describe the quality of learning process. According to (Dobrowolska et al., 2016) Mentors have the responsibility to assess the learning outcomes achieved by students. According to (Mardapi, 2012) The quality of learning depends on the assessment.

A good assessment must fulfill the principles of openness, fairness and meaningfulness. Open means that all students can find out how the process and results of the assessment are carried out, fair refers to equality of treatment in assessment both from the quality of the instrument, supervision, scoring, to assessment, and meaningfulness means that the assessment carried out has

more value if it is not compared to not carrying out an assessment. According to (Anderson, 2002) Assessment must meet 3 principles, namely meaningful, transparency, and fairness.

In the context of education, measurement and assessment do not only end with one aspect of competence, but are complex, that is more than one aspect of competence. This means that assessment and measurement do not only measure and assess student understanding or cognitive aspect, but also reach the stage of measuring and assessing abilities students in coordinating their understanding into real actions (performance). According to (Sudirtha, Mayuni, & Budhyani, 2014) assessment must cover all aspects of student competence. According to (Herni Yuniarti Suhendi1, et al 2018) Assessment process must be conducted by three competences such Cognitive, Affective, and psychomotor.

Performance assessment is common assessment models to assess students' performance and it was a type of authentic assessment. According to (Gallardo, 2020) Performance-Based Assessment basically based on students' Performance and was the part of authentic assessment. (Juniadi, Aisyah E. Palupi, & Euis Ismayati, 2013) Performance assessment is the process of measuring the performance of students in using their knowledge to work on certain skills. (AlKhateeb, 2018) Performance-Based Assessment have significance impact on student achievement and self-efficacy for students in Yordan. Based on (Abualrob & Al-Saadi, 2019) Performance-Based Assessment is Assessment model that more motivating students. According to (Umami, 2018) performance assessment is an appropriate for assessing students' performance in Islamic education subject.

Procedurally, measurements can be carried out using instruments or measurements and assessments, both in the form of test and non-test activities. According to (Rusilowati, A. Dewi & Fianti, 2021)

measurement not only use instruments in the form of tests such multiple choice question, but also can use non-test measuring instruments such as observation sheet. (Nugroho, Yudha, Sundari, & Praja, 2021) the observation sheet is a good tool for assessing students' practical performance. (Nadhiroh & Sigit, 2018) conducted research on the development of performance instruments in the form of observation sheets used by teachers in assessing students' practical performance on acid-base materials, salt hydrolysis, and buffer solutions.

The information used in the assessment process is certain numbers obtained through the measurement process, where measurement means a process of compare certain sizes or numbers that represent an indicator that describes the condition of the measuring object or students on certain variables. According to (SA Jabar & S Arikunto, 2004) measurement is the process of comparing something with a certain size so that the quantitative nature of a thing is formed.

The process of determining numbers in measurements can be carried out with the help of a measurement scale, namely the agreement used as a reference. Criteria certain so that the value of a variable measured can be expressed in the form of numbers. According to (P. D Sugiyono, 2017) The measurement scale is an agreement that is used as a guide to determine the long and short intervals in the measuring instrument so that the measuring instrument used can produce quantitative data. The measurement scale consists of several models, namely Likert, Thurstone, Semantic Differential, Staple, and Guttman scales (Verma, 2019). Likert and Semantic Differential are the most popular scale used Measuring Instrument in sociology psychology, politics and other fields (Taherdoost, 2019).

In principle, a good assessment and measurement process must use good instruments or tools, namely instruments that are tested for quality both in terms of validity

and reliability as well as in terms its practicality. (Azwar, 2011) based on the agreement of psychometric experts, the quality of the instrument is determined by its validity, reliability and practicality. According to (P. Dr Sugiyono, 2016) a good instrument is an instrument that has good validity and reliability. Based on Carmines & (Carmines, E. G., & Zeller, 1979) there are two basic characteristics that can be used as a reference in considering the quality of the instrument, namely Validity and Reliability.

The process of producing good instruments can be done with a research and development approach. According to (Adib, 2017) instrument development can be carried out using a research and development approach.

The availability of performance instruments for assess students' performance practice of congregational prayer in class II Islamic Elementary School has not been found in the scientific literature. Several classroom action studies conducted an assessment of students' practice scores before and after the action was carried out but the quality of the assessment instruments used was not guaranteed, such as research conducted by (Muhson, 2015) who conducted action research by applying a jigsaw cooperative model but the assessment instruments used (Instrument test and non-test) validity and reliability have not been guaranteed. A similar study was also conducted by (Khoerodin, 2015) who conducted action research by applying the modeling the way method but the assessment instruments used were tested for validity and reliability, both test instruments and practice.

Although the results of a literature review show that there is no standard instrument for performing congregational prayer practices, it is necessary to carry out an empirical study by conducting a needs analysis in Islamic Elementary School, to ensure that the instrument for performing prayer is urgently needed.

The results of empirical studies through interviews show that teachers have not used The standard instrument for assessing students' abilities in practice congregational prayer.

Apart from the lack of teacher ability in developing assessment instruments, the unavailability of standard assessment instruments that can be used by teachers is also one of the factors teachers do not use standard instruments in conducting assessments. Practice in prayer students quality in terms of their level of validity and reliability as well as practicality. Research conducted by Khoerudin (2015) applies the the way modeling by assessing practical prayer in congregation of students but the instruments used are not standard. Research conducted by Muchson (2015) applied a jigsaw cooperative model by conducting pre-test and post abilities practical prayers of students and students' understanding of congregational prayers but the instruments used are not standard. Therefore it is necessary to carry out developmental research to develop a standard performance instrument as an instrument to assess level of student performance in congregational prayer practice for Islamic elementary school students.

METHODS

This research is a development research using the 4 D model developed by Thiagarajan which consists of the stages of instrument needs analysis, instrument design, instrument development, and instrument deployment (Hendriana, Rofi'i, & Hartono, 2021). This research only took three development steps, namely needs analysis, design, and instrument development.

This research aims to develop a good performance instrument as an instrument to assess the level of student performance in congregational prayer practice for Islamic elementary school students.

The research was conducted at Al-Qolam Islamic Elementary School, Nurul Ilmi Islamic Elementary School, and Sambinae Islamic Elementary School in Bima City.

Method of data collection used in this research is the interview method and document study, while the instruments used are a short list of questions and observation sheets, as well as expert validation sheets (P. Dr Sugiyono, 2018) The interview instrument at the needs analysis stage uses digital technology assistance in the form of google form and WhatsApp.

Qualitative data analysis in this study used the 3-step method from (Ridder, Miles, Michael Huberman, & Saldaña, 2014), namely data reduction, data presentation, and concluding research data, while quantitative data was analyzed using SPSS Microsoft Excel (Setiawan, Pusporini, & Dardjito, 2020).

Content validity analysis was carried out using the Aiken V (Suciati, Munadi, Sugiman, & Ratna Febriyanti, 2020).

$$V = \sum s / n (c-1)$$

$$S = r - lo$$

$$r = \text{Score given by Expert}$$

$$lo = \text{Lowest score}$$

$$c = \text{Highest score (Triwibowo, Rusilowati, Anggani, & Bharati, 2020)}$$

Test the reliability of performance instruments students in prayer are carried out with the help of the SPSS data processing application (Darren George, 2018). The analytical model used is the reliability test between raters using the ICC or Intraclass Correlation Coefficient (Tong, Tang, Irby, Lara-Alecio, & Guerrero, 2020).

RESULTS AND DISCUSSION

Research on the development of performance instruments to assess student performance in congregational prayer practice activities resulted in 3 performance instruments, each of which can be used to assess the performance of Priests praying in

congregation, Main Congregation, and *Masbuk* Congregation.

The performance instrument for the Priest praying congregation consists of 23 question items consisting of 4 variables measurement syaf, reading, movement, and cooperation. The Main Congregation performance instrument consists of 24 statement items and consists of the same 4 variables as the priest's performance instrument, while the *Masbuk* Congregation performance instrument consists of 25 statement items and 5 variables, namely syaf settings, readings, movements, competency as an Priest *Masbuk*, and cooperation.

The three instruments that have been developed have been tested for quality through the assessment of five experts consisting of academics and practitioners. According to (Wakano, Isnaeni, & Ahmadi, 2022) using expert judgement for assessing the quality of instrument that developed. Lyn

et al suggested that minimum number of expert for expert judgement is three expert (Almanasreh, Moles, & Chen, 2019).

The results of the expert validation of the instrument for the performance of Priests praying in congregation which were developed were then analyzed using the Aiken or Aiken V using the help of Microsoft Excel software. According to (Maulita, Sukarmin, & Marzuki, 2019) using validity Aiken for measure the content validity of instrument that developed and its analyze by using Microsoft excel.

The results of the analysis showed that all statement consisted of 23 statement items for priest performance instruments, 24 statement items for core performance instruments, and 25 statement items for advanced performance instruments that were developed. Valid as the data shown in Table 1, Table 2, and Table 3.

Table 1. Content Validity of Performance Instruments For Piers of Congregational Prayer

Item	Expert Score (r)				1	S1	S2	S3	S4	Σs	n	c	c-	n (c-	$\Sigma s/ n (c-1)$	Description
	I	II	III	IV	o								1	1)		
1	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
2	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
3	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid
4	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
5	4	3	3	3	1	3	2	2	2	9	4	4	3	12	0.75	Valid
6	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
7	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid
8	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
9	4	3	3	3	1	3	2	2	2	9	4	4	3	12	0.75	Valid
10	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
11	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
12	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid
13	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
14	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
15	4	3	3	3	1	3	2	2	2	9	4	4	3	12	0.75	Valid
16	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
17	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid
18	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
19	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid
20	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
21	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
22	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid

23	3	4	4	4	1	2	3	3	3	11	4	4	3	12	0.92	Valid
Aiken V Average															0.87	Valid

Data expert judgement and the validity analysis of Aiken V in table 1 above shows that all statement items in the performance assessment instrument to assess the performance of the priest in practice are Valid, where the Aiken validity value of each statement item is more than 0.30 and the average Aiken V all statement items 0.87 or more than 0.30. According to (Sahrul, Khumaedi, & Masrukan, 2022) Using 0.30 as minimum standard of Aiken validity criteria on their study. And (Rusilowati, A. Dewi & Fianti, 2021) declare 0.74 as Valid criteria.

Thus the performance instrument for the Priest praying congregation is valid with a very high level of validity because the average validity value is more than 0.80 (Tomoliyus , 2020). In addition to the performance instrument for the Priest praying in congregation, the performance instrument for the core of the congregational prayer consisting of 24 statement items was also assessed by experts and analyzed for the validity of Aiken with valid results as shown in Table 2.

Table 2. Content Validity of Performance Instruments For Main Congregation of Congregational Prayer

Item	Expert Score				S1	S2	S3	S4	$\sum$ s	n	c	c-1	n (c-1)	$\sum$ s/ n (c-1)	Description	
	I	II	III	IV												
1	4	2	4	4	1	3	1	3	3	10	4	4	3	12	0.83	Valid
2	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid
3	3	3	4	3	1	2	2	3	2	9	4	4	3	12	0.75	Valid
4	4	3	3	4	1	3	2	2	3	10	4	4	3	12	0.83	Valid
5	4	3	4	4	1	3	2	3	3	11	4	4	3	12	0.92	Valid
6	4	3	3	4	1	3	2	2	3	10	4	4	3	12	0.83	Valid
7	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid
8	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid
9	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid
10	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid
11	4	4	3	4	1	3	3	2	3	11	4	4	3	12	0.92	Valid
12	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid
13	4	3	4	4	1	3	2	3	3	11	4	4	3	12	0.92	Valid
14	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid
15	4	3	4	4	1	3	2	3	3	11	4	4	3	12	0.92	Valid
16	3	4	4	3	1	2	3	3	2	10	4	4	3	12	0.83	Valid
17	4	3	3	4	1	3	2	2	3	10	4	4	3	12	0.83	Valid
18	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid
19	4	4	3	4	1	3	3	2	3	11	4	4	3	12	0.92	Valid
20	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid
21	4	3	3	4	1	3	2	2	3	10	4	4	3	12	0.83	Valid
22	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid
23	4	3	3	4	1	3	2	2	3	10	4	4	3	12	0.83	Valid
24	4	1	4	4	1	3	0	3	3	9	4	4	3	12	0.75	Valid

Aiken V Average	0.85	Valid
-----------------	------	-------

Data from the analysis of the validity instrument in the Table 2 shows that all of the statement items in the instrument of main Congregation of congregational prayer are valid because their validity index exceeds 0.30. which is between 0.75 and 0.92 with an

average validity value of 0.85. The *Masbuk* Congregation performance instrument which consists of 25 statement items tested is valid because the Aiken validity index exceeds 0.30 as the data is shown in Table 3. Below.

Table 3. Content Validity of Performance Instruments For *Masbuk* Congregation of Congregational Prayer

Item	Expert (r)				Score	lo	S1	S2	S3	S4	Σs	n	c	c-1	n (c-1)	Σs/ n (c-1)	Description
	I	II	III	IV													
1	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
2	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid	
3	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
4	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid	
5	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
6	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid	
7	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
8	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid	
9	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
10	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid	
11	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid	
12	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid	
13	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
14	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
15	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid	
16	4	4	3	3	1	3	3	2	2	10	4	4	3	12	0.83	Valid	
17	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
18	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
19	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid	
20	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
21	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
22	3	3	4	4	1	2	2	3	3	10	4	4	3	12	0.83	Valid	
23	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
24	3	3	3	4	1	2	2	2	3	9	4	4	3	12	0.75	Valid	
25	4	4	4	3	1	3	3	3	2	11	4	4	3	12	0.92	Valid	
Aiken V Average																0.86	Valid

Source: Expert Validation Data and Aiken V Analysis with Microsoft Excel

After the instrument is tested for validity. the next step is to perform a reliability analysis using ICC analysts or Intraclass Correlation Coefficient through the SPSS data processing application.

The results of data analysis using the SPSS application show that three performance instruments developed are reliable as the data presented in Table 4. Table 5 and Table 6 below.

Table 4. Reliability of Performance Instruments for Priest of Congregational Prayer

Intraclass Correlation Coefficient							
	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.144 ^a	.020	.744	4.861	3	66	.004
Average Measures	.794 ^c	.317	.985	4.861	3	66	.004

Table 5. Reliability of Performance Instruments for Main Congregation of Congregational Prayer

Intraclass Correlation Coefficient							
	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df 2	Sig
Single Measures	.105 ^a	.006	.686	3.828	3	69	.013
Average Measures	.739 ^c	.135	.981	3.828	3	69	.013

Table 6. Reliability of Performance Instruments for *Masbuk* Congregation of Congregational Prayer

Correlation Coefficient							
	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.124a	.015	.715	4.555	3	72	.006
Average Measures	.780c	.275	.984	4.555	3	72	.006

Source : Results of Reliability Analysis ICC Through SPSS

Data in Table 4. Table 5. and Table 6 shows that according to the results of the SPSS analysis explains the performance instrument for practicing prayer that is developed to be reliable. where the reliability coefficient is shown through the Intraclass Correlation in the Average Measure line.

Table 4 shows the results of the reliability of the performance for priests praying in congregation with an Average Measure value of 0.794. Table 5 shows the reliability results of the performance for the core congregation of congregational prayers with an Average Measure value of 0.739 and Table 6 displays the results of the reliability of the performance for the congregational Prayer with an Average Measure value of 0.780. Analysis reliability of three performance instruments for the practice of prayer . it can be concluded that all performance instruments developed are

reliable with high and very high reliability criteria because the reliability coefficient of each instrument more than 0.50 and 0.75. According to (Khoirul Bashoor & Supahar, 2018) coefficient Reliability more Or equal to 0.50 categorize as High level of Reliability and coefficient Reliability more Or equal to 0.75 categorize Very high Reliability. (Perinetti, 2018) categorize coefficient Reliability between 0.50 to 0.75 as fair Reliability, between 0.75 to 0.90 is good reliability and coefficient Reliability between 0.90 to 1 categorize as high reliability.

CONCLUSION

Based on the research and development results. there are three instruments for assessing student performance in congregational prayer namely an instrument for assessing the performance of

the Priest consisting of 23 statement items that consisting of 4 variables measurement such *syaf*, reading, movement, and cooperation. an instrument for assessing the performance of main congregation that consisting of 24 statement items and have the same variables as the priest's performance instrument . An instrument for assessing the performance of the *Masbuk* Congregation which consists of 25 statement items and 5 variables, namely *syaf* settings, readings, movements, competency as an *Masbuk* congregation, and cooperation.

The performance instrument for assessing student performance in practices congregational prayer are valid and reliable base on expert judgement. Even the instruments that developed are valid and reliable, it is necessary to carry out further research with the process of testing the instrument for factor analysis and internal consistency of the instrument.

REFERENCES

- Abualrob, M. M. A., & Al-Saadi, S. H. (2019). Performance-Based Assessment: Approach and Obstacles by Higher-Elementary Science Teachers in Palestine. *Journal of Education and Learning*, 8(2),198.
- Adib, H. S. (2017). Teknik Pengembangan Instrumen Penelitian Ilmiah di Perguruan Tinggi Keagamaan Islam. *Sains Dan Teknoogi*, 139–157.
- AlKhateeb, M. A. (2018). The effect of using performance-based assessment strategies to tenth-grade students' achievement and self-efficacy in Jordan. *Cypriot Journal of Educational Sciences*, 13(4), 489-500.
- 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.
- Anderson, L. W. (2002). *Classroom Assessment Enhancing the Quality of Teacher Decision Making* (1st ed.).
- Azwar, S. (2011). *Reliabilitas Dan Validitas*. Yogyakarta: Pustaka Pelajar.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. New Delhi: Sage publications.
- Darren George, P. M. (2018). *IBM SPSS Statistics 25 Step by Step (A Simple Guide and Reference)* (15th ed.).
- Dobrowolska, B., McGonagle, I., Kane, R., Jackson, C. S., Kegl, B., Bergin, M., ... Palese, A. (2016). Patterns of clinical mentorship in undergraduate nurse education: A comparative case analysis of eleven EU and non-EU countries. *Nurse Education Today*, 36, 44–52.
- Edi Istiyono, Djemari Mardapi, S. (2014). Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (Pythots)Peserta Didik SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 2–12.
- Gallardo, K. (2020). Competency-Based Assessment and the Use of Performance-Based Evaluation Rubrics in Higher Education: Challenges Towards the Next Decade. *Problems of Education in the 21st Century*, 78(1), 61–79.
- Guangul, F. M., Suhail, A. H., Khalit, M. I., & Khidhir, B. A. (2020). Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College. *Educational Assessment, Evaluation and Accountability*, 32(4), 519–535.
- Hendriana, D., Ruffi, R., & Hartono, H. (2021). The Development of Web-Based Learning Models as A Learning Medium for Students of Audio Video Electronics Competencies. *Edcomtech Jurnal Kajian Teknologi Pendidikan*, 6(1), 100–109.
- Herni Yuniarti Suhendi1*, Muhammad Ali Ramdhani2, F. S. I. (2018). Verification Concept of Assesment for Physics Education Student Learning Outcome. *International Journal of Engineering & Technology (IJET)*, 7, 321–325.
- Immonen, K., Oikarainen, A., Tomietto, M., Kääriäinen, M., Tuomikoski, A. M., Kaučič, B. M., ... Mikkonen, K. (2019). Assessment of nursing students' competence in clinical practice: A systematic review of reviews. *International Journal of Nursing Studies*, 100.
- Juniadi, Aisyah E. Palupi, & Euis Ismayati. (2013). Pengembangan Perangkat Penilaian Kinerja Praktik perawatan Mesin

- Penggerak Utama Kapal pada Peserta didik kelas XI TKPI SMK Negeri 3 Tarakan. *Jurnal Pendidikan Vokasi: Teori Dan Praktek*, Vol.1 No.1(1), 80–90.
- Khoerodin. (2015). *Upaya Meningkatkan Hasil Belajar Mata Pelajaran Fiqih Materi Shalat Berjama'ah Melalui Penerapan Metode Modeling The Way Pada Siswa Kelas Ii Mi Muhammadiyah*. Retrieved from tedunan.
- Khoirul Bashooir & Supahar. (2018). Validitas dan reliabilitas instrumen asesmen kinerja literasi sains pelajaran Fisika berbasis STEM. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(2), 219–230.
- Mardapi, D. (2012). *Pengukuran penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Maulita, S. R., Sukarmin, S., & Marzuki, A. (2019). The Content Validity: Two-Tier Multiple Choices Instrument to Measure Higher-Order Thinking Skills. *Journal of Physics: Conference Series*, 1155(1).
- Muhson, M. (2015). *Penerapan Model Pembelajaran Kooperatif Tipe Jigsaw Untuk Meningkatkan Hasil Belajar Fiqh Siswa MI Ringinsari Sanankulon Blitar* (Universitas Islam Negeri Sayyid Ali Rahmatullah Tulungagung).
- Nadhiroh, A., & Sigit, D. (2018). Pengembangan Instrumen Penilaian Sikap dan Keterampilan Psikomotorik pada Materi Asam Basa, Titrasi Asam Basa, Hidrolisis Garam, dan Larutan Penyangga. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, 3(7), 887-890.
- Nugroho, W. A., Yudha, R. P., Sundari, S., & Praja, H. N. (2021). Analisis Instrumen Asesmen Unjuk Kerja pada Pembelajaran PJOK di Sekolah Dasar Kota Cirebon. *Gelombang Olahraga: Jurnal Pendidikan Jasmani Dan Olahraga (JPJO)*, 4(2), 126-141.
- Perinetti, G. (2018). StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. *South European Journal of Orthodontics and Dentofacial Research*, 5(1), 3–5.
- Ridder, H. G., Miles, M. B., Michael Huberman, A., & Saldaña, J. (2014). Qualitative data analysis. A methods sourcebook. *Zeitschrift Fur Personalforschung*, 28(4), 485–487.
- Rusilowati, A. Dewi, C., & Fianti. (2021). The analysis of exploratory factors on the development of data, technology, and human literacy assessment instrument. *Journal of Physics: Conference Series*, 1918(5).
- SA Jabar & S Arikunto. (2004). *Evaluasi Program Pendidikan*. Jakarta: Bumi Aksara.
- Sahrul, S., Khumaedi, M., & Masrukan, M. (2022). Development of Instruments to Measure Self-Confidence and Creative Thinking in Mathematics Learning for Vocational High School Students. *Journal of Research and Educational Research Evaluation*, 11(1), 81–92.
- Setiawan, A., Pusporini, W., & Dardjito, H. (2020). Observation instrument for student social attitude in primary schools: Validity and reliability. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 24(1), 76-87.
- Suciati, Munadi, S., Sugiman, & Ratna Febriyanti, W. D. (2020). Design and validation of mathematical literacy instruments for assessment for learning in Indonesia. *European Journal of Educational Research*, 9(2), 865-875.
- Sudirtha, I. G., Mayuni, P. A., & Budhyani, I. D. A. M. (2014). Pengembangan Instrumen Penilaian Mata Kuliah Praktik Tata Busana. *Jurnal Pendidikan Indonesia*, 3(1), 326–336.
- Sugiyono, P. D. (2017). *Statistika Untuk Penelitian*. Bandung: Alfabeta.
- Sugiyono, P. Dr. (2016). *Penelitian kuantitatif, kualitatif, dan R&D*. Yogyakarta: Alfabeta.
- Sugiyono, P. Dr. (2018). *Metode Penelitian Kombinasi (Mixed Methods)*. Bandung: Alfabeta.
- Taherdoost, H. (2019). What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale. *International Journal of Academic Research in Management (IJARM)*, 8(1), 2296–1747.
- Tomoliyus, T., & Sunardianta, R. (2020). Validitas Aiken's instrumen tes untuk mengukur reaktif agility olahraga khusus tenis meja. *Jurnal Keolahragaan*, 8(2), 148-157.
- Tong, F., Tang, S., Irby, B. J., Lara-Alecio, R., & Guerrero, C. (2020). The determination of appropriate coefficient indices for inter-rater reliability: Using classroom observation instruments as fidelity measures in large-scale randomized research. *International Journal of Educational Research*, 99(June 2019), 101514.

- Triwibowo, F. D., Rusilowati, A., Anggani, D., & Bharati, L. (2020). Content Validity and Reliability of The Inter-Rater Instrument for Android- Based Speaking Performance Assessment. *Journal of Educational Research and Evaluation*, 9(1), 52–57.
- Umami, M. (2018). Penilaian Autentik Pembelajaran Pendidikan Agama Islam dan Budi Pekerti dalam Kurikulum 2013. *Jurnal Kependidikan*, 6(2), 222-232.
- Verma, J. P. (2019). Measurement and Scaling Techniques. In: *Statistics and Research Methods in Psychology with Excel. In Statistics and Research Methods in Psychology with Excel* .
- Wakano, R., Isnaeni, W., & Ahmadi, F. (2022). Developing Instruments Assessment of Students' Critical Thinking and Communication Skills in Biology Learning Using Hybrid Learning Models in 3T Areas. *Journal of Research and ...*, 11(1).