

Development And Validity-Reliability Testing Of An Early Childhood Education Institution Performance Instrument Based On A Psychometric Approach

Melati Indri Hapsari*, Ani Rusilowati, Nina Oktarina

Universitas Negeri Semarang, Semarang, Indonesia

*Corresponding Author: melatikesling06@students.unnes.ac.id

Abstract. This study aimed to develop and examine the validity and reliability of a psychometric-based performance instrument for Early Childhood Education (ECE) institutions. The study employed a Research and Development (R&D) method using a quantitative psychometric approach. Instrument development was conducted through several stages, including construct identification, indicator formulation, expert validation, instrument trials, and statistical validity-reliability testing. The instrument was developed based on four dimensions of institutional performance: input, process, output, and outcome, consisting of 30 statement items using a Likert scale. The research sample involved 300 teachers and administrators of ECE institutions. Content validity was analyzed using Aiken's V index, construct validity was tested through Exploratory Factor Analysis (EFA), and reliability was measured using Cronbach's Alpha coefficient. The results showed that all instrument items met the validity criteria, with Aiken's V values ranging from 0.80 to 0.90 and factor loading values above 0.50. The Kaiser-Meyer-Olkin (KMO) value exceeded 0.50, and Bartlett's Test of Sphericity was significant, indicating the feasibility of factor analysis. The reliability test produced a Cronbach's Alpha value of 0.973, indicating very high internal consistency. Therefore, the instrument was declared valid, reliable, and feasible for evaluating ECE institutional performance.

Keywords: early childhood education; institutional performance instrument; psychometric approach; validity and reliability testing

INTRODUCTION

Early Childhood Education (ECE) serves as a fundamental foundation in the development of human resources because early childhood is a critical period characterized by rapid growth in children's cognitive, socio-emotional, language, moral, and motor aspects. The quality of ECE services is an essential factor in determining children's developmental success at subsequent educational stages. Therefore, ECE institutions are required to demonstrate effective, professional, and quality-oriented organizational performance. The performance of ECE institutions encompasses not only administrative achievements but also the quality of learning processes, educators' competencies, institutional management, community relations, and the attainment of national education standards (Directorate of ECE, 2021).

In addition, the development of quality ECE institutions is closely related to the achievement of the Sustainable Development Goals (SDGs), particularly Goal 4: Quality Education, which aims to ensure inclusive and equitable quality education and promote lifelong learning opportunities for all. One of the specific targets of

SDG 4 emphasizes access to quality early childhood development, care, and pre-primary education for all children. The performance of Early Childhood Education (ECE) institutions is one of the important indicators in determining the quality of educational services for young children. Institutional performance is not only measured from administrative aspects, but also includes the effectiveness of learning, the quality of educators, leadership, institutional management, learning environment, parental involvement, and the achievement of educational quality standards. High-performing ECE institutions play a strategic role in supporting children's readiness for primary education, reducing educational inequality, and improving the quality of human resources in the long term. Therefore, strengthening institutional performance in ECE is not only a national educational priority but also part of the global commitment toward sustainable development.

Furthermore, the SDGs also highlight the importance of effective institutions, accountability, and evidence-based policymaking, as reflected in Goal 16 concerning strong institutions and Goal 17 regarding partnerships for sustainable development. In this context, valid and

reliable performance measurement instruments are necessary to provide accurate data for institutional evaluation, quality assurance, and educational policy improvement. The availability of scientifically developed instruments can support stakeholders in identifying institutional strengths and weaknesses, designing intervention programs, and improving the quality of ECE services sustainably.

In practice, the measurement of ECE institutional performance still faces various challenges. Many evaluation instruments currently used lack a strong developmental foundation, resulting in validity and reliability that have not been empirically tested. Instruments with low validity may produce biased data and fail to accurately represent the actual conditions of ECE institutions. Furthermore, most instruments remain administrative in nature and do not comprehensively measure the substantive dimensions of institutional performance. Consequently, performance evaluation results often cannot serve as an accurate basis for decision-making in institutional quality improvement.

The development of high-quality measurement instruments requires a systematic scientific approach. One relevant approach is the psychometric approach. The psychometric approach emphasizes the importance of validity, reliability, item discrimination, internal consistency, and the construction of indicators capable of accurately measuring variables. According to Robert F. DeVellis, proper instrument development should involve stages including construct conceptualization, indicator formulation, expert validation, instrument trials, item analysis, and statistical testing of validity and reliability. Thus, the psychometric approach can produce instruments that are objective, accurate, and trustworthy in measuring the performance of ECE institutions.

The importance of using a psychometric approach is also supported by various previous studies. Research conducted by Lee J. Cronbach demonstrated that instrument reliability is a primary requirement for obtaining consistent measurement results. In addition, Samuel Messick emphasized that validity is not only related to the content of an instrument but also to the interpretation of the measurement results produced. In the educational context, valid and reliable instruments are essential for supporting continuous educational quality evaluation.

Several previous studies have also revealed

that measuring the performance of educational institutions requires multidimensional indicators. A study conducted by the National Association for the Education of Young Children emphasized that the quality of ECE is influenced by leadership, teacher quality, learning environment, parental involvement, and institutional management. Previous studies have shown that the quality of educational evaluation instruments greatly influences the quality of research data and educational decision-making. Research conducted by Samuel Messick emphasized that validity is not only related to the content of the instrument, but also to the interpretation and use of measurement results. In addition, the study by Yusup (2018) demonstrated that validity and reliability testing are important stages in quantitative research to ensure that the instrument is appropriate for data collection. In the field of Early Childhood Education (ECE), several previous studies have primarily focused on the evaluation of learning programs, institutional accreditation, or the quality of early childhood education services. Research conducted by Mashburn et al. (2008) found that classroom quality and the learning environment in ECE institutions have a significant relationship with children's academic and social development. Similarly, Slot (2018) demonstrated that structural quality and learning processes are key factors in determining the quality of ECE institutions.

However, research on the development of psychometric-based performance instruments for ECE institutions remains relatively limited, particularly in Indonesia. Research on the development of psychometric-based performance instruments for ECE institutions in Indonesia remains relatively limited. Most existing studies focus more on program evaluation or institutional accreditation rather than on developing measurement instruments that comprehensively meet psychometric standards.

Furthermore, the development of educational policies in Indonesia demands a more accountable and data-driven institutional evaluation system. The implementation of educational quality assurance, ECE accreditation, and the Merdeka Belajar policy requires measurement instruments capable of providing valid information regarding institutional conditions and performance achievements. Without standardized and empirically tested instruments, the evaluation process may lose objectivity and hinder efforts to improve the quality of early childhood education.

Based on the explanations above, the

development and validity-reliability testing of psychometric-based performance instruments for ECE institutions are important to conduct. This study is expected to produce instruments that are valid, reliable, and feasible to be used as comprehensive tools for evaluating the performance of ECE institutions. The results of this study are also expected to support decision-making, quality assurance, the achievement of Sustainable Development Goals (SDGs), and the development of early childhood education policies in Indonesia.

This study aims to develop and examine the validity and reliability of a psychometric-based performance instrument for Early Childhood Education (ECE) institutions. Specifically, the objectives of this study are:

1. To develop a performance measurement instrument for Early Childhood Education (ECE) institutions that encompasses aspects of input, process, output, and outcome.
2. To test the content validity of the instrument through expert judgment to ensure the suitability of the items with the constructs being measured.
3. To examine the construct validity of the instrument using statistical approaches such as Exploratory Factor Analysis (EFA) or Confirmatory Factor Analysis (CFA).
4. To test the reliability of the instrument through internal consistency analysis, such as Cronbach's Alpha coefficient, in order to determine the reliability level of the instrument.

METHODS

This study employed a Research and Development (R&D) method using a psychometric approach to develop a performance instrument for Early Childhood Education (ECE) institutions. This method was selected because it aims to produce an instrument that is valid, reliable, and feasible to be used as an evaluation tool for ECE institutional performance. The psychometric approach was applied to ensure that the instrument possesses good measurement quality through empirical validity and reliability testing. This study employed an instrument development design based on a quantitative approach. The instrument development model in this study referred to the stages of instrument development proposed by Robert F. DeVellis and Nunnally & Bernstein, which include: (1) identifying the constructs and dimensions of ECE

institutional performance; (2) developing indicators and instrument items; (3) conducting content validation through expert judgment; (4) carrying out instrument trials; (5) analyzing validity and reliability; and (6) revising and finalizing the instrument.

The research population consisted of all teachers and administrators of Early Childhood Education (ECE) institutions, particularly kindergartens, with relatively similar characteristics in Pati Regency. The research sample was determined using purposive sampling or proportional random sampling techniques based on the characteristics of the respondents. The sample size for the instrument testing consisted of 300 participants. Data collection was conducted through questionnaires as the main research instrument. The research instrument used a Likert scale with five alternative responses: strongly disagree, disagree, neutral, agree, and strongly agree.

Validity testing was conducted in several stages: 1) Content Validity. Content validity was assessed through expert judgment. The level of agreement among experts was analyzed using Aiken's V index to determine the relevance of the items to the measured constructs; 2) Construct Validity. Construct validity was tested using Exploratory Factor Analysis (EFA) to identify the factor structure of the instrument. The adequacy of factor analysis was determined based on the following criteria 1) Kaiser-Meyer-Olkin (KMO) > 0.50 ; 2) Bartlett's Test of Sphericity significant (< 0.05); 3) Factor loading ≥ 0.50 . Instrument reliability was analyzed using Cronbach's Alpha coefficient to determine the internal consistency of the instrument. The instrument is considered reliable if it has a Cronbach's Alpha value of ≥ 0.70 , as suggested by Nunnally & Bernstein (1994). The research data were analyzed using descriptive and inferential statistics with the assistance of SPSS software.

RESULTS AND DISCUSSION

Description of the Development of the ECE Institutional Performance Instrument

The performance of Early Childhood Education (ECE) institutions refers to the level of institutional success in achieving educational goals through resource management, the implementation of learning processes, and the achievement of educational outcomes and impacts effectively and efficiently. According to Stephen P. Robbins and Timothy A. Judge, organizational

performance is the work outcome achieved by individuals or groups within an organization according to their roles and responsibilities in order to accomplish organizational goals (Robbins & Judge, 2017). In the context of early childhood education, UNESCO (2015) explains that the quality and performance of ECE institutions are determined by the integration of resources (inputs), learning processes, and children's developmental outcomes. This is further supported by OECD (2018), which emphasizes that early childhood education performance includes service quality, learning interactions, and children's developmental outcomes. Therefore, the performance of ECE institutions in this study is defined as a multidimensional construct that reflects the level of institutional effectiveness and quality in managing inputs, implementing learning processes, producing educational outputs, and creating outcomes that positively impact learners and society.

The development of the Early Childhood Education (ECE) institutional performance instrument was conducted to produce a measurement tool capable of evaluating institutional performance objectively, validly, and reliably. The development of this instrument was based on the need for an institutional evaluation system that is not only oriented toward administrative aspects but also capable of measuring substantive dimensions related to the quality of early childhood education services. The performance of ECE institutions is understood as the level of institutional success in managing resources, implementing learning processes, achieving educational outcomes, and providing

positive impacts on children's development and community satisfaction.

Conceptually, the instrument development refers to educational evaluation theory and the psychometric approach, which emphasize the importance of instrument validity and reliability. According to Robert F. DeVellis, instrument development should begin with clear construct identification so that each item can accurately represent the measured variable. In addition, Nunnally and Bernstein emphasized that a good instrument must have high internal consistency so that the measurement results can be trusted.

The initial stage of instrument development was carried out through literature review and needs analysis related to the performance of ECE institutions. The theoretical review was derived from national education standards, ECE policies, and previous studies regarding the quality of early childhood education. Based on the findings of the review, several main dimensions were identified as the basis for constructing the instrument, namely: (1) input aspects, (2) process aspects, (3) output aspects, and (4) outcome aspects. The performance of ECE institutions was measured using the Input–Process–Output–Outcome (IPOO) approach, which was elaborated into dimensions, indicators, and instrument items based on a 1–5 Likert scale. Based on these four dimensions, the researchers developed an instrument blueprint consisting of indicators and statement items using a Likert scale. Each item was designed to comprehensively represent the construct of ECE institutional performance.

Table 1. Instrument Blueprint

Variable	Dimension	Indicator	Item	Scale
ECE Institutional Performance	Input	Educator qualifications	1	Likert 1–5
		Teachers' pedagogical competence	2	Likert 1–5
		Teacher–student ratio	3	Likert 1–5
		Play facilities	4	Likert 1–5
		Safe and comfortable learning environment	5	Likert 1–5
		Institutional curriculum	6	Likert 1–5
		Learning tools and materials	7	Likert 1–5
		Institutional funding	8	Likert 1–5
	Process	Learning planning	9	Likert 1–5
		Active learning implementation	10	Likert 1–5
		Varied learning methods	11	Likert 1–5
		Use of instructional media	12	Likert 1–5
		Utilization of technology	13	Likert 1–5
		Teacher–child interaction	14	Likert 1–5
		Learning evaluation	15	Likert 1–5
		Principal's supervision	16	Likert 1–5

Output	Institutional administration	17	Likert 1–5
	Program evaluation	18	Likert 1–5
	Children's cognitive development	19	Likert 1–5
	Socio-emotional development	20	Likert 1–5
	Language development	21	Likert 1–5
	Motor development	22	Likert 1–5
	Readiness for elementary school	23	Likert 1–5
	Children's attendance	24	Likert 1–5
Outcome	Parents' satisfaction	25	Likert 1–5
	Graduates' adaptation	26	Likert 1–5
	Institutional reputation	27	Likert 1–5
	Community trust	28	Likert 1–5
	Growth in student enrollment	29	Likert 1–5
	Program sustainability	30	Likert 1–5

After the item construction process, content validity was conducted through expert judgment involving ECE experts and educational management expert. This stage aimed to ensure that the instrument items were consistent with the theoretical constructs being measured.

Furthermore, the instrument was piloted with teachers and administrators of ECE institutions to obtain empirical data. The pilot test data were analyzed using a psychometric approach through construct validity and reliability testing. Construct validity was analyzed using Exploratory Factor Analysis (EFA) and/or Confirmatory Factor Analysis (CFA), while reliability was analyzed using Cronbach's Alpha coefficient.

Results of Instrument Content Validity

Content validity testing was conducted through expert judgment involving five experts in the fields of early childhood education and

educational management. The assessment used a 1–5 Likert scale to evaluate the relevance of each item to the measured construct. The content validity analysis method employed was Aiken's V.

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

Explanation:

- s = the score assigned by the experts minus the lowest score
- n = the number of experts
- c = the number of rating categories

Criteria:

- $V > 0,90 \rightarrow$ highly valid
- $0,8 \leq V < 0,90 \rightarrow$ valid
- $V < 0,80 \rightarrow$ not valid
- p is in the range 0.04 to 0.07 $\rightarrow$ content validity is significant

Table 2. Results of Aiken's V Calculation

Item	Σs	V	Kategori
1.	18	0.9	Valid
2.	19	0.95	Highly valid
3.	16	0.8	Valid
4.	15	0.75	Not valid
5.	17	0.85	Valid
6.	19	0.95	Highly valid
7.	17	0.85	Valid
8.	17	0.85	Valid
9.	17	0.85	Valid
10.	19	0.95	Highly valid
11.	20	1.00	Highly valid
12.	17	0.85	Valid
13.	17	0.85	Valid

14.	18	0.9	Valid
15.	19	0.95	Highly valid
16.	18	0.9	Valid
17.	16	0.8	Valid
18.	18	0.9	Valid
19.	18	0.9	Valid
20.	19	0.95	Highly valid
21.	18	0.9	Valid
22.	20	100	Highly valid
23.	19	0.95	Highly valid
24.	18	0.9	Valid
25.	17	0.85	Valid
26.	18	0.9	Valid
27.	17	0.85	Valid
28.	17	0.85	Valid
29.	18	0.9	Valid
30.	17	0.85	Valid

The results of the analysis using Aiken's V index show that 29 items have V values between 0.80 and 1.00, which indicates that they fall into the valid and highly valid categories. There is 1 invalid item, namely item number 4, with a V value of 0.75, which is below 0.80. This item was therefore removed. Thus, the instrument consisting of 29 items is considered to have met content validity requirements and is deemed suitable for further testing in the next stage.

Results of Construct Validity Testing (Construct Validity)

The results of the construct validity test using Exploratory Factor Analysis (EFA) indicate that all indicators of the research variables meet the validity criteria and are appropriate for use in the study. Based on the test results, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value is above 0.50, indicating that the sample size is adequate for factor analysis. In addition, the results of Bartlett's Test of Sphericity show a significance value of less than 0.05, meaning that the data meet the requirements for further analysis using EFA.

The analysis results indicate that most indicators have factor loading values above 0.50, suggesting that they are valid in representing the research constructs. However, there is one indicator with a communality value below 0.50, namely item number 28. This finding indicates that the indicator is not adequately explained by the extracted factors. Therefore, indicators with communality values below 0.50 are generally considered for elimination to improve the overall construct model and enhance its

representativeness. Nevertheless, based on the results of the Rotated Component Matrix, item 28 has a factor loading value greater than 0.50. Consequently, item 28 does not need to be removed or revised, as it still demonstrates sufficient validity in measuring the intended construct.

Results of Instrument Reliability Testing

The results of the instrument reliability test using the Cronbach's Alpha coefficient show that the Cronbach's Alpha value is 0.973. This value indicates that the instrument has a very high level of internal consistency. Based on the reliability criteria, an instrument is considered reliable if it has a Cronbach's Alpha value of ≥ 0.70 . Therefore, the Cronbach's Alpha value of 0.973 indicates that all items in the instrument have very good interrelatedness and consistency in measuring the construct of PAUD institution performance. These results indicate that the developed instrument is appropriate for use as a data collection tool in the research because it is capable of providing consistent and reliable measurement results. The high reliability value also indicates that each instrument item has been systematically arranged and is able to optimally represent the measured variables.

Instrument Revision and Finalization

The revision and finalization stage of the instrument was conducted after the completion of the content validity, construct validity, and reliability testing processes. The purpose of the instrument revision was to improve the quality of the items so that they would be more aligned with

the measured theoretical constructs, easier for respondents to understand, and capable of providing more optimal measurement performance. The revision process was carried out based on the results of statistical analysis and input from experts (expert judgment).

Based on the results of content validity testing using Aiken's V index, 29 questionnaire items were found to be valid, while 1 item was deemed invalid and was therefore removed from the instrument. Nevertheless, several items that obtained relatively lower validity scores compared to the other items were revised to improve their clarity, specificity, and to minimize potential ambiguity. These revisions focused on language usage, the clarity of indicators, and the alignment of the items with the dimensions of the Early Childhood Education (ECE) institutional performance construct.

Furthermore, the results of construct validity testing through Exploratory Factor Analysis (EFA) showed that most items had factor loading values above 0.50 and were therefore considered valid in representing the research constructs. There was one item, namely item number 28, that had a communality value below 0.50. However, based on the results of the Rotated Component Matrix, the item still had a factor loading value above 0.50, indicating that it was still capable of representing the extracted factor. Therefore, item number 28 was retained and not eliminated from the instrument.

In addition, the reliability test results showed a Cronbach's Alpha value of 0.973, indicating a very high level of internal consistency. The results of the Cronbach's Alpha if Item Deleted analysis also showed that there were no items that significantly improved reliability if removed. Thus, all items in the instrument were considered consistent and appropriate to retain.

Based on the overall testing results, the psychometric-based PAUD institution performance instrument met the criteria of validity and reliability. The final instrument consisted of 29 statement items covering the dimensions of input, process, output, and outcome. The instrument was declared suitable for use as an evaluation tool for PAUD institution performance in research activities, educational quality assurance, and data-driven decision-making in early childhood education.

Discussion of Research Findings

The research findings indicate that the Early Childhood Education (PAUD) institution

performance instrument developed using a psychometric approach has met the criteria of validity and reliability, making it suitable for use as an institutional evaluation tool. The instrument development process was carried out through systematic stages, beginning with construct identification, indicator formulation, expert validation, instrument trials, and continuing through statistical validity and reliability testing. This approach is consistent with the view of Robert F. DeVellis, who states that instrument development must follow scientific procedures in order to produce accurate and trustworthy measurement tools.

Conceptually, this study successfully identified the construct of PAUD institution performance into four main dimensions: input, process, output, and outcome. This multidimensional approach demonstrates that the performance of educational institutions is not solely measured from administrative aspects, but also includes learning quality, children's developmental achievements, and the institutional impact on society. These findings support the view of UNESCO that the quality of early childhood education is determined by the integration of resources, learning processes, and student developmental outcomes. In addition, the results of this study are consistent with the research of Pamela Morris and colleagues, who emphasize the importance of comprehensive PAUD quality evaluation to improve the quality of early childhood education services.

The results of the content validity assessment using Aiken's V index showed that 29 instrument items were categorized as valid to highly valid, while 1 item was found to be invalid. The Aiken's V values exceeding the minimum acceptable threshold indicate that the experts judged the instrument items to be consistent with the theoretical constructs being measured. These findings suggest that the development of the indicators was carried out systematically based on relevant theories, educational standards, and empirical needs in the field. According to Lewis R. Aiken, content validity is a crucial stage in instrument development because it determines the extent to which the items comprehensively represent the construct under investigation.

Furthermore, the results of construct validity testing using Exploratory Factor Analysis (EFA) showed that the instrument met the requirements for factor analysis feasibility. The Kaiser-Meyer-Olkin (KMO) value above 0.50 indicates that the research sample size was adequate for factor analysis, while the significant Bartlett's Test of

Sphericity results indicate sufficiently strong correlations among items. These results suggest that the instrument construct structure has strong interrelationships, enabling the formation of clear measurement dimensions. According to Henry F. Kaiser, an adequate KMO value indicates that the data are appropriate for factor analysis.

The study also found that most items had factor loading values above 0.50 and were therefore considered valid in representing the research constructs. High factor loading values indicate that the items contribute well to the extracted factors. These findings support psychometric theory, which states that items with high factor loadings have a better ability to explain the measured latent variables. Although one item had a communality value below 0.50, the item was retained because the Rotated Component Matrix results showed a factor loading above 0.50. The decision to retain the item was based on theoretical and statistical considerations to ensure that the instrument construct continued to comprehensively represent the outcome aspect. In addition to validity, the research results also showed that the instrument had very high reliability, with a Cronbach's Alpha value of 0.973. This value indicates that the instrument has excellent internal consistency. According to Lee J. Cronbach, a high reliability coefficient indicates that each item in the instrument consistently measures the same construct. These findings are consistent with the studies of Jum C. Nunnally and Ira H. Bernstein, who stated that an instrument is considered reliable if it has a Cronbach's Alpha value above 0.70.

The high reliability value of the instrument indicates that the developed items were systematically organized and possessed good homogeneity. It also demonstrates that the instrument is capable of providing stable and consistent measurement results when used under relatively similar conditions. Therefore, this instrument can be used as an institutional evaluation tool that supports objective and data-based quality assurance processes in early childhood education.

The findings of this study have important implications for the development of the PAUD evaluation system in Indonesia. A valid and reliable instrument can help PAUD institutions identify institutional strengths and weaknesses more comprehensively. In addition, this instrument can be used by the government, education offices, and accreditation institutions as a basis for decision-making in improving the

quality of early childhood education. These findings support national education policies that emphasize the importance of data-driven decision making in educational management.

This study also supports the achievement of the United Nations Sustainable Development Goals (SDGs), particularly Goal 4 concerning Quality Education. The use of standardized evaluation instruments can help improve the quality of PAUD services sustainably, thereby supporting the creation of inclusive, equitable, and quality education for all early childhood learners.

CONCLUSION

The results of the Aiken's V analysis indicate that 29 instrument items fall within the valid to highly valid categories, while 1 item is categorized as invalid. Consequently, the instrument is considered suitable for further testing and evaluation. Furthermore, construct validity testing using the Exploratory Factor Analysis (EFA) and/or Confirmatory Factor Analysis (CFA) approach showed that the factor structure of the instrument was consistent with the formulated theoretical constructs. The reliability testing results using Cronbach's Alpha coefficient also demonstrated that the instrument had a high level of internal consistency. Thus, the developed ECE institutional performance instrument was declared valid, reliable, and feasible to be used as an evaluation tool in quality assurance and decision-making in early childhood education. This instrument is expected to assist ECE institutions, government agencies, and educational stakeholders in conducting institutional evaluations objectively, systematically, and based on data.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Early Childhood Education (ECE) experts who provided valuable input regarding the concepts, dimensions, and performance indicators of ECE institutions to ensure alignment with the characteristics of early childhood education and national education standards.

REFERENCE

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142.

- Azwar, S. (2021). *Reliabilitas dan Validitas* (Edisi ke-4). Yogyakarta: Pustaka Pelajar.
- Direktorat PAUD. (2021). *Pedoman Penjaminan Mutu Pendidikan Anak Usia Dini*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.
- Ghozali, Imam. (2016). *Aplikasi Analisis Multivariete dengan Program IBM SPSS 23*. Semarang: Badan Penerbit Universitas Diponegoro.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
- Henry F. Kaiser. (1974). An Index of Factorial Simplicity. *Psychometrika*, 39(1), 31–36.
- Lee J. Cronbach (1951). *Coefficient alpha and the internal structure of tests*. *Psychometrika*, 16(3), 297–334.
- Mardapi, D. (2017). *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Parama Publishing.
- Mashburn, A. J., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749.
- National Association for the Education of Young Children. (2020). *Developmentally Appropriate Practice Position Statement*. Washington, DC.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). McGraw-Hill.
- OECD. (2018). *Engaging Young Children: Lessons from Research about Quality in Early Childhood Education and Care*. Paris: OECD Publishing.
- OECD. (2018). *Starting strong V: Transitions from early childhood education and care to primary education*. Paris: OECD Publishing
- Robbins, S. P., & Judge, T. A. (2017). *Organizational behavior* (17th ed.). Pearson
- Robert F. DeVellis (2017). *Scale Development: Theory and Applications* (4th ed.). Sage Publications.
- Samuel Messick. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Slot, P. L. (2018). Structural characteristics and process quality in early childhood education and care: A literature review. *OECD Education Working Papers*, No. 176.
- United Nations. (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: United Nations.
- UNESCO. (2015). *Early childhood care and education (ECCE)*. Paris: UNESCO.
- UNESCO. (2015). *Education for All 2000–2015: Achievements and Challenges*. Paris: UNESCO Publishing.
- Widoyoko, E. P. (2018). *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.
- Yusup, F. (2018). Uji validitas dan reliabilitas instrumen penelitian kuantitatif. *Jurnal Tarbiyah: Jurnal Ilmiah Kependidikan*, 7(1), 17–23.