E-ISSN: 2502-3519 P-ISSN: 2798-3234
JGA, Vol. 10(1), March 2025 (29-41)
DOI: https://doi.org/10.14421/jga.2025.101.03

Golden Age
Jurnal Ilmiah Tumbuh Kembang Anak Usia Dini

OPEN ACCESS

A Self-Assessment Instrument for Early Childhood Teachers: Construct Validity and Reliability Analysis

Martha Christianti[1], Harun[2], Nur Hayati[3], Oktaviani Maharani[4], Nazaludin Nur Rahmat[5], Apry Aditya Saputra[6], Triandi Aprilio[7]
[1,2,3,4,5,6,7]Universitas Negeri Yogyakarta, Indonesia.

Keywords:
self-evaluation assessment, ECE teachers, construct validity, reliability, factor analysis.

Correspondence to
Martha Christianti, Department of Early Childhood Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia.
e-mail:
marthachristianti@uny.ac.id

Received 18 07 2024
Revised 10 01 2025
Accepted 24 02 2025
Published Online First 28 02 2025

Abstract
Early childhood education teachers in Indonesia often lack standardized tools to assess their pedagogical and professional competence. This study aims to develop and validate a reliable self-evaluation instrument by identifying key competency dimensions to support early childhood teachers in reflecting on their professional practices. The instrument consists of 39 items encompassing seven competency dimensions: evaluation and reflection skills, holistic educational approach, curriculum mastery, learning activity design, professionalism, technology integration, and communication. Data were collected using a 4-point Likert scale from 276 respondents for exploratory factor analysis (EFA) and 978 respondents for confirmatory factor analysis (CFA). EFA was used to identify the instrument's underlying factor structure, while CFA tested the model's fit with empirical data. The EFA results indicate that seven factors with eigenvalues greater than 1 were extracted, explaining 60.378% of the total variance across the 39 items. CFA results confirm that the conceptual model aligns with empirical data, with all standardized loading factors found to be significant. Construct reliability, measured using Composite Reliability (CR), shows that four of the seven dimensions have satisfactory reliability (CR > 0.70), while the remaining three dimensions demonstrate moderate reliability. This study contributes to early childhood education by presenting a validated and reliable self-assessment instrument to measure early childhood teachers' pedagogical and professional competence. The instrument supports teachers in reflecting on their teaching practices and promotes ongoing professional development. It also offers a foundational tool for policymakers and institutions to improve teacher quality. Future research can examine its applicability in various educational contexts and evaluate its effectiveness in guiding targeted development programs.

To cite: Christianti, M., Harun, H., Hayati, N., Maharani, O., Rahmat, N. N., Saputra, A. A., & Aprilio, T. (2025). A self-assessment instrument for early childhood teachers: Construct validity and reliability analysis. *Golden Age: Jurnal Ilmiah Tumbuh Kembang Anak Usia Dini, 10*(1), 29-41. *https://doi.org/10.14421/jga.2025.10-03*

Introduction

The school accountability movement, teacher effectiveness, and teacher assessment systems aimed at improving academic standards have long received significant attention. Unfortunately, in Indonesia, there is no credible system in place to measure the quality of teacher performance. Several experts argue that the current teacher evaluation procedures are superficial, merely formal, and inconsistent (Damanik, 2019; Kelly et al., 2020; Öztabak & Polatlar, 2020). In general, the existing evaluation systems fail to enhance teacher quality or improve student learning outcomes.

These shortcomings stem from several issues. First, the teacher evaluation system does not accurately measure teacher quality, as it fails to distinguish between effective and ineffective teachers. Second, it does not contribute to the development of highly skilled teaching staff (Gates & Gates, 2010; Kane et al., 2014). Some evaluations are unable to differentiate between teachers, the quality of their instruction, and their impact on student achievement. Studies also reveal that most current evaluation programs give high ratings to nearly all teachers, thereby overlooking teaching effectiveness (Cohen & Goldhaber, 2016; Darling-Hammond, 2015; Sholeh, 2017). Such evaluations hinder professional growth and provide an inaccurate portrayal of classroom performance.

Given the limitations of traditional evaluation systems, there is a growing need for a more reflective, teacher-centered approach, such as self-evaluation. Teachers rarely receive constructive feedback that fosters improvement, and school principals often do not use evaluations to enhance instructional practices that can boost student achievement (Grissom et al., 2015; Juma, 2024; Kraft & Gilmour, 2016). According to Danielson, several external barriers contribute to this problem, including unclear teacher competency standards, ineffective evaluation instruments, and insufficient time for proper implementation (Best, 2016; Moss, 2015; Supadi et al., 2021). Internal constraints include a lack of high-quality professional development for evaluators and school cultures that do not promote critical feedback or negative evaluations (Boud, 2015; Buyamin, 2023). Other barriers include limited oversight or incentives for administrators to conduct accurate assessments (Hornstein, 2017; Kusumawardhani, 2017).

Given these challenges, self-evaluation instruments can serve as a viable alternative. They encourage continuous reflection and professional growth while addressing existing gaps in traditional systems. Self-assessment allows teachers to take greater ownership of their development process. It also fosters a culture of accountability and improvement within educational institutions. Therefore, implementing self-evaluation instruments may lead to more meaningful and sustainable improvements in teaching quality.

Teacher evaluation is therefore an important tool for independently measuring pedagogical and professional competence (Filipe et al., 2015; Kane et al., 2014; Maghfuroh et al., 2020). Early childhood education (ECE) teachers play a critical role in implementing ECE and must possess adequate competencies to support optimal child development. Evaluating their competencies is essential for maintaining and enhancing teaching quality and student learning outcomes (Bank, 2020; Kementerian Pendidikan Nasional RI, 2014; Retnawati et al., 2018; Suryadarma & Jones, 2015).

Self-evaluation enables teachers to reflect on their practices, identify strengths and areas for improvement, and plan targeted professional development (Fireside & Lachlan-Hache, 2015; Fitria et al., 2019; Juma, 2024; Nur Efendi & Muh Ibnu Sholeh, 2023). However, valid and reliable assessment instruments remain a challenge, particularly in the Indonesian early childhood education context (Maghfuroh et al., 2020; Manggaberani & Putro, 2023; Rista et al., 2020; Suranto et al., 2014). Many existing instruments do not fully capture the complexity of competencies required of ECE teachers, especially amid technological advances and dynamic curricular changes.

One effective method for developing valid and reliable instruments is factor analysis. This statistical approach includes two main techniques: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA and CFA are widely used to identify and confirm the factor structure of a measurement instrument (Manggaberani & Putro, 2023). EFA helps explore the underlying dimensions without prior assumptions, while CFA tests whether the proposed structure fits the data. These methods are essential to ensure that the instrument aligns with the intended theoretical framework.

This study aims to develop a self-evaluation instrument tailored to the needs of early childhood education teachers in Indonesia, focusing on pedagogical and professional competencies. The instrument's development is based on teacher competency standards outlined in Government Regulation No. 137 of 2022 concerning National Standards for Early Childhood Education, as well as Danielson's theory of teaching competence. The resulting instrument is intended to serve as a practical tool to enhance teachers' reflective abilities and professionalism, ultimately supporting high-quality early childhood education.

This study proposes a conceptual model comprising seven key dimensions: evaluation and reflection skills, holistic educational approach, curriculum mastery, learning activity design, professionalism, technology integration, and communication. Each dimension is measured through specific indicators developed and tested using factor analysis. The instrument is expected to offer a comprehensive understanding of ECE teacher competencies and serve as a foundation for improving the quality of instruction in early childhood education institutions.

Methods

Design

This exploratory study aims to describe the quality of standardized pedagogical and professional competency assessment instruments developed to evaluate the competence of early childhood education (ECE) teachers in Indonesia. The assessment must be developed according to standards, as the results are expected to provide meaningful conclusions and have significant implications for PAUD (Pendidikan Anak Usia Dini) teachers in understanding their competencies. This study utilizes data collected from questionnaires completed by PAUD teachers from five districts in Yogyakarta. The assessment is presented as a series of items rated on a four-point Likert scale (1–4) that best reflects the teachers' characteristics.

The study focuses on describing the quality of the competency assessment instrument—both pedagogical and professional—based on two aspects: validity evidence and estimated reliability of each component's construct. To address these objectives, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were employed to examine whether the test construction is empirically supported. Additionally, we estimated the reliability of the test to assess the accuracy of the measurement results.

Participants

This study involved responses from 1,254 early childhood education teachers who voluntarily completed the developed instrument. A total of 276 PAUD teachers were included in the EFA stage, while the remaining 978 teachers participated in the CFA stage. The participants represented various PAUD institutions, including TPA, KB, TK, and PAUD posts in the Special Region of Yogyakarta (DIY).

Instruments

The study used assessment instruments developed specifically for this research. The primary constructs measured were the pedagogical and professional competencies of ECE teachers. The instrument consisted of 39 items, which were analyzed using EFA to explore the underlying constructs. Each item was rated on a four-point Likert scale, with the scoring system shown in Table 1.

Table 1. Scoring System

Answer Option	Score	Criteria
Unsatisfactory	1	Does not meet expectations
Basic Stage	2	Meets basic requirements
Satisfactory	3	Meets expectations
Very Satisfactory	4	Exceeds expectations

Procedure

The assessment was disseminated through a standardized procedure. Once finalized, the instrument was sent in the form of a link to the Chairperson of HIMPAUDI Yogyakarta and the Chairperson of IGTKI Yogyakarta. Each chairperson distributed the link to teachers through coordination with school principals and community groups. The responses were automatically collected via the form platform, and the data were subsequently analyzed. The identities of participating teachers were kept confidential to ensure the privacy of each respondent.

Data Analysis

The data analysis was conducted in four stages. First, we analyzed the factors associated with each item to identify the underlying constructs. This included evaluating the goodness-of-fit criteria of the conceptual model and the factor loadings of each manifest variable. Model fit was assessed using the following statistical indicators:

- Comparative Fit Index (CFI) ≥ 0.90 (Hu & Bentler, 1999)
- Tucker-Lewis Index (TLI) ≥ 0.90 (Hu & Bentler, 1999)

- Root Mean Square Error of Approximation (RMSEA) between 0.06 and 0.08, with a 90% confidence interval (Schreiber et al., 2006)
- Standardized Root Mean Square Residual (SRMR) ≤ 0.08 (Hu & Bentler, 1999)

To evaluate the standardized factor loadings of each manifest variable, we used the criteria $\beta > |0.5|$ with $p < 0.05$.

Result

The results of the research on the construction of self-evaluation instruments for PAUD teachers began with a descriptive statistical explanation to determine the characteristics of the data. In addition, the next stage was to examine the factors of each instrument item to identify the fundamental aspects of the teacher self-evaluation instrument using Exploratory Factor Analysis (EFA). The subsequent step was to assess the alignment between the conceptual model and fit statistics with empirical data. The data are presented with standardized loading factors to determine the contribution of each indicator to its corresponding latent variable. In the final stage, we present the results of the reliability analysis of the construct and the average variance explained for each latent variable.

Construct Validation

The initial step in determining the factors of the self-evaluation instrument was conducted by empirically testing the EFA. The results of the analysis indicated that the performance assessment of PAUD teachers, which refers to Government Regulation No. 137 of 2022 and Danielson's theory, is divided into 39 question items. These items encompass competencies such as organizing developmental aspects according to early childhood characteristics; analyzing play theories based on developmental aspects and stages, needs, potentials, talents, and interests of early childhood; designing early childhood development activities based on the curriculum; organizing educational development activities; utilizing technology, information, and communication to implement educational development activities; developing children's potential for self-actualization; communicating effectively, empathetically, and politely; organizing and compiling assessment reports on early childhood learning processes and outcomes; determining the scope and objectives of assessment for early childhood learning processes and outcomes; using the results of assessment, development, and evaluation for the benefit of early childhood development; undertaking reflective, corrective, and innovative actions to improve the quality of early childhood development processes and outcomes; developing materials, structures, and scientific concepts that support and align with early childhood developmental needs and stages; designing various development activities creatively in accordance with early childhood developmental stages; and continuously developing professionalism through reflective practices.

Each observed variable in the instrument was analyzed using EFA to verify the construct validity of the instrument according to its developmental framework. The data, collected from 288 respondents, were based on a Likert scale ranging from 1 to 4. Construct validity testing began with exploratory factor analysis, which grouped the 39 items into seven factors. The naming of these new factors was carried out through a focus group discussion (FGD) involving experts and researchers. The seven resulting factors were identified as educational holistic competence, curriculum, activity design, professionalism, technology integration, communication, and evaluation and reflection. The determination of the number of factors was based on eigenvalues greater than 1. These seven factors collectively explained 60.378% of the variance of the 39 analyzed items, with the variance explained by each factor as follows: 38.582%, 5.704%, 3.699%, 3.395%, 3.180%, 2.971%, and 2.847%. The variables within each factor were rotated using the Varimax method with Kaiser normalization. A loading factor coefficient with an absolute value of 0.4 or greater was considered to indicate a strong loading. Table 1 presents the results of the rotated component matrix, while Table 2 provides the naming and definitions of each factor.

Table 2. Rotated Component Matrix

Butir Item	Component						
	1	2	3	4	5	6	7
10a	,734	,215	,147	,070	,093	,238	,023
10b	,724	,108	,329	,137	,025	,046	,123
8a	,721	,179	,118	-,034	,001	,172	,122
15c	,697	,106	,237	,248	,218	,085	,041
8b	,676	,160	,047	,006	,093	,060	,329
15d	,664	,064	,264	,301	,236	,067	-,029
9b	,661	,196	,194	,118	,000	,185	,235
11a	,650	,200	,142	,150	,210	,149	-,015
17b	,620	,135	,269	,141	,449	,119	-,048
14b	,601	,081	,170	,140	,503	,095	,009
15b	,572	,045	,116	,451	,148	,130	,197
11b	,536	,127	-,014	,275	,200	,202	,142
12b	,531	,147	,142	,509	,035	,080	,322
6b	,529	,266	,089	,057	,071	,422	,292
17d	,504	-,016	,371	,197	,373	,135	,160
9a	,495	,326	,320	,027	-,019	,129	,193
16c	,487	,228	-,065	,342	,397	,154	,016
4b	,471	,144	,021	,208	,183	,189	,038
2b	,143	,696	,157	,204	,073	,087	,019
6a	,209	,688	-,005	,147	,211	,201	-,050
2a	,243	,687	,306	,018	,016	,018	,211
1a	,137	,628	,361	-,017	-,020	,068	,159
16a	,338	,342	,076	,178	,276	,294	,060
12a	,199	,339	,697	,189	,132	,052	,079
15a	,334	,315	,656	,203	,218	,019	,072
3a	,195	,166	,639	-,098	,045	,379	-,027
13b	,347	,254	,538	,291	,213	-,029	,112
13a	,501	,182	,027	,594	,034	,075	,182
4a	,189	,180	,160	,575	,122	,292	,203
3b	,134	,066	,469	,507	,023	,300	,020
16b	,156	,383	,280	,407	,377	-,044	-,121
17a	,109	-,013	,128	-,035	,705	,147	,138
14a	,229	,217	,064	,162	,632	,047	,249
5a	,180	,057	,146	,190	,102	,632	-,011
5b	,327	,083	-,019	,148	,069	,626	,028
17c	,029	,163	,360	-,074	,290	,459	,266
1b	,285	,392	,281	,089	,045	,402	,146
7b	,149	,028	,091	,121	,204	,014	,796
7a	,348	,212	,038	,216	,113	,146	,609

Table 3. Naming of ECE Teacher Performance Competency Factors

Competency	**Description**	**Item Code**
Evaluation and reflection competencies (A)	Teachers' ability to evaluate and analyse activities, assessments, and learning materials, as well as conduct regular and innovative reflection for continuous improvement in teaching practices and program development.	4b, 6b, 8a, 8b, 9a, 9b, 10a, 10b, 11a, 11b, 12b, 14b, 15b, 15c, 15d, 16c, 17b, 17d
Educational holistic competencies (B)	The ability of teachers to understand and apply child development theories, identify individual potential, and design and implement creative learning activities that are in accordance with children's needs and interests.	1a, 2a, 2b, 6a, 16a

doi: *https://doi.org/10.14421/jga.2025.101-03*

Competency	Description	Item Code
Curriculum competencies (C)	The ability of teachers to understand and apply curriculum structures and scientific concepts relevant to early childhood development, including the adjustment of activities according to the stages of child development.	3a, 12a, 13b, 15a
Activity design competencies (D)	The ability of teachers to design and organize educational activities that are educational, creative, and in accordance with the goals of the curriculum and stages of child development.	3b, 4a, 13a, 16b
Professional competence (E)	Teachers' commitment to continuously improve their professionalism through continuous self-development is an important part of a career as a teacher.	14a, 17a
Technology integration competencies (F)	The ability of teachers to utilize technology in the implementation and innovation of learning activities, as we as to organize activities that support children's development and participate in professional development.	5a, 5b, 1b, 17c
Communication competencies (G)	The ability of teachers to communicate effectively, empathetic, and politely with children.	7a, 7b

Descriptive Statistics

The pedagogical and professional competency construct comprises seven latent variables: evaluation and reflection ability (A), holistic education (B), curriculum (C), activity design (D), professional competence (E), technology integration (F), and communication (G). Each latent variable contains a varying number of indicators, referred to as manifest variables. Evaluation and reflection consist of 18 manifest variables, holistic education includes 5, while curriculum, activity design, and technology integration each consist of 4 manifest variables. Professional competence and communication each include 2 manifest variables. In total, the model comprises 39 manifest variables. Descriptive statistics including the mean, standard deviation, minimum, and maximum values for each latent and manifest variable are presented in Table 4.

Table 4. Indicators of the Pedagogical and Professional Components and Descriptive Statistics

Indicator	Abbreviation	*M*	*SD*	*Min.*	*Max.*
Evaluation and Reflection	A	60.38	7.70	18	72
Holistic Educative	B	16.56	2.47	5	20
Curriculum	C	13.05	2.00	4	16
Activity Design	D	14.24	1.67	4	16
Professional	E	7.28	0.86	2	8
Technology Integration	F	12.63	1.77	4	16
Communication	G	7.48	0.78	2	8

Model Fit Test

To further validate the construct of the self-evaluation instrument for PAUD teachers, a Confirmatory Factor Analysis (CFA) was conducted. This stage aimed to assess the compatibility between the proposed conceptual model and the empirical data obtained from a sample of 978 participants during the second stage of the study. The proposed conceptual model is illustrated in Figure 1. In evaluating the model fit, several goodness-of-fit indices were used, with the results presented in Table 5.

Table 5. Goodness of Fit Indices for CFA of ECE Self-Assessment Instrument

Fit Indices	Good Fit	Proposed Model
RMSEA	$0 \leq \text{RMSEA} \leq 0.05$	0.028
CFI	$0.90 \leq \text{CFI} \leq 1.00$	0.908
TLI	$0.90 \leq \text{TLI} \leq 1.00$	0.900
SRMR	$0.00 \leq \text{SRMR} \leq 0.05$	0.045

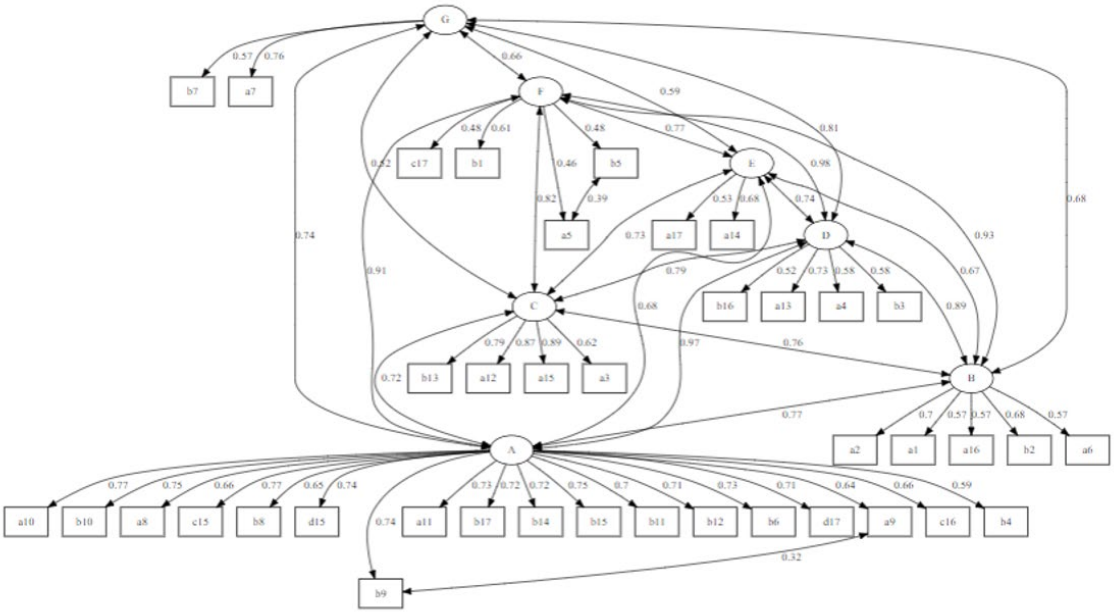

Figure 1. First-order factor model of self-evaluation assessment of ECE teachers with standardized parameter estimates

In Figure 1, it can be seen that each of the latent variables is interconnected. Within latent variable A, indicators a10 and b8 show the highest loading factors, while b4 demonstrates the lowest loading. In latent variable B, indicator a2 has the highest loading, whereas a1, a16, and a6 show the lowest loadings about latent variable B. For latent variable C, indicator a15 presents the highest standardized loading, and a3 the lowest. In latent variable D, a13 has the highest loading, while b16 has the lowest. Within latent variable E, indicator a14 has the highest loading, whereas a16 has the lowest. For latent variable F, indicator b1 demonstrates the highest loading, while a5 shows the lowest. In latent variable G, a7 has the highest loading, and b7 the lowest. Table 6 shows that all standardized loadings in the first-order factor model are statistically significant, although they vary in magnitude.

In Figure 1, all standardized loading factors fall within the strong category. This indicates that the seven latent variables in the model significantly contribute to measuring the performance competencies of early childhood education (ECE) teachers through self-evaluation. All loading factors are statistically significant (see Table 6).

Table 6. Loading Factor, Residual, and R-square of the First-order Factor Model

Variable Laten	Variable Manifest	Estimate	Std. Err	z-Value	P(>|z|)	St. Loading	Resid.	R square
A	a10	1.000				0.495	0.168	0.594
	b10	0.940	0.038	25.060	0.000	0.465	0.172	0.557
	a8	0.898	0.041	21.651	0.000	0.445	0.257	0.435
	c15	0.957	0.037	26.137	0.000	0.474	0.152	0.597
	b8	0.819	0.038	21.367	0.000	0.406	0.223	0.425
	d15	0.911	0.037	24.944	0.000	0.451	0.165	0.553
	b9	0.974	0.039	24.912	0.000	0.482	0.189	0.552
	a11	0.951	0.039	24.526	0.000	0.471	0.191	0.537
	b17	0.875	0.037	23.911	0.000	0.433	0.177	0.515
	b14	0.927	0.039	23.996	0.000	0.459	0.196	0.518
	b15	0.875	0.035	25.224	0.000	0.433	0.146	0.563
	b11	0.808	0.035	23.140	0.000	0.400	0.168	0.487
	b12	0.824	0.035	23.798	0.000	0.408	0.159	0.511
	b6	0.864	0.035	24.439	0.000	0.428	0.160	0.534

doi: *https://doi.org/10.14421/jga.2025.101-03*

Variable Laten	Variable Manifest	Estimate	Std. Err	z-Value	P(>|z|)	St. Loading	Resid.	R square
	d17	0.902	0.038	23.458	0.000	0.447	0.201	0.499
	a9	0.726	0.035	21.017	0.000	0.359	0.183	0.413
	c16	0.731	0.034	21.746	0.000	0.362	0.168	0.438
	b4	0.631	0.033	19.249	0.000	0.313	0.179	0.353
B	b2	1.000			0.000	0.441	0.232	0.457
	a6	0.640	0.041	15.647	0.000	0.282	0.168	0.322
	a2	0.871	0.046	18.738	0.000	0.384	0.156	0.486
	a1	0.628	0.040	15.699	0.000	0.277	0.160	0.324
	a16	0.610	0.039	15.761	0.000	0.269	0.149	0.327
C	a12	1.000			0.000	0.547	0.100	0.749
	a15	1.058	0.029	35.976	0.000	0.579	0.089	0.790
	a3	0.570	0.027	21.032	0.000	0.312	0.159	0.380
	b13	0.890	0.030	30.082	0.000	0.487	0.141	0.628
D	a13	1.000			0.000	0.405	0.145	0.531
	a4	0.735	0.041	18.102	0.000	0.298	0.177	0.333
	b3	0.782	0.043	18.328	0.000	0.317	0.194	0.341
	b16	0.872	0.054	16.291	0.000	0.353	0.334	0.272
E	a17	1.000			0.000	0.273	0.192	0.279
	a14	1.542	0.126	12.217	0.000	0.420	0.209	0.458
F	a5	1.000			0.000	0.252	0.234	0.214
	b5	1.382	0.095	14.540	0.000	0.349	0.408	0.230
	c17	1.255	0.111	11.336	0.000	0.317	0.335	0.231
	b1	1.764	0.136	12.926	0.000	0.445	0.339	0.369
G	b7	1.000			0.000	0.223	0.102	0.328
	a7	1.761	0.124	14.201	0.000	0.392	0.112	0.578

Note: variant residue a10, b2, a12, a13, a17, a5, b7 fixed at zero so that it can be identified

The "Estimate" column in Table 6 presents the unstandardized loading factors for each manifest and latent variable. For interpretation purposes in this study, we rely solely on the standardized loading factors shown in the "St. Loading" column. These values range from 0.223 to 0.579, representing the degree of contribution of each manifest variable. The "P(>|z|)" column confirms the significance of each loading factor in the first-order factor model, with all values indicating significance. The "Residual" column shows the estimated error values, with the highest error associated with b5 and the lowest with a15.

The "R-square" column indicates the total variance explained by each manifest variable. Among them, a15 explains the highest variance (79%), while a5 explains the lowest (21.5%). For latent variable A, the indicators explain between 35.3% and 59.7% of the total variance. Latent variable B explains between 32.2% and 48.6% of its indicators' variance. The remaining latent variables—C, D, E, F, and G—account for between 21.4% and 79% of the variance across their respective indicators. These findings indicate that the seven latent variables—evaluation and reflection competence, holistic education, curriculum, activity design, professional competence, technology integration, and communication—are effective in explaining the pedagogical and professional abilities of ECE educators.

Construct Reliability

Table 7 presents the construct reliability (CR) values for the pedagogical and professional competency constructs. As illustrated in the conceptual model in Figure 1, there are seven constructs measured: the overarching construct of pedagogical and professional ability, and its subconstructs—evaluation and reflection (A), holistic education (B), curriculum (C), activity design (D), professional (E), technology integration (F), and communication (G). Among these subconstructs, four exhibit satisfactory reliability with CR values greater than 0.70. The remaining three subconstructs have CR values below 0.70, indicating lower reliability.

Table 7. Construct Reliability

Construction	CR
Pedagogical and Professional Skills	0.97
Evaluation and Reflection (A)	0.94
Holistic Educative (B)	0.75
Curriculum (C)	0.82
Activity Design (D)	0.71
Professional (E)	0.57
Technology Integration (F)	0.65
Communication (G)	0.64

Discussion

The results of our study show that the pedagogical and professional competency assessment instrument for early childhood education (ECE) teachers is both valid and reliable. This study successfully identifies seven essential factors that comprehensively measure the pedagogical and professional competencies of ECE teachers: evaluation and reflection skills, holistic education, curriculum, activity design, professionalism, technology integration, and communication. These dimensions reflect not only the theoretical foundations of teacher competence but also their practical applications in the classroom. By capturing these multifaceted components, the instrument provides a meaningful tool for assessing teacher effectiveness in early childhood settings.

Moreover, the inclusion of 39 indicators across these seven factors demonstrates the instrument's depth and scope. Each indicator contributes significantly, offering a rich variety of content relevant to the core competencies of ECE teaching. The confirmed construct validity shows that the instrument effectively measures the intended theoretical constructs, as supported by previous literature (Hartono & Muchtar, 2018; Otaya et al., 2020; Shaffer et al., 2016; Umar & Nisa, 2020). Furthermore, the test's construction is based on a sound conceptual framework, aligning with current standards and policy directions in early childhood education. This foundation enhances the instrument's relevance and utility for both research and practical evaluation purposes.

Evaluation and reflection competencies constitute a major portion of the instrument, comprising 18 manifest variables. This highlights the critical role of reflection and evaluation in the ongoing professional development of teachers. As stated by Schön, reflection is central to sustainable professional growth (Al Riyami, 2015; Goh, 2019). The EFA results identified seven distinct factors that significantly represent the constructs of pedagogical and professional competence. These seven factors explained 60.378% of the variance across the 39 items—an acceptable proportion for newly developed psychometric instruments (Filgueiras et al., 2014). The factor structure was further validated through focus group discussions (FGDs) with experts, which provided both theoretical grounding and appropriate naming for each factor, thereby strengthening construct validity.

The standardized loading factors in the confirmatory factor analysis ranged from 0.223 to 0.579, suggesting a range in the strength of the indicators' contributions to their respective latent constructs. While variability exists, these values remain within acceptable thresholds for psychometric measurement instruments, indicating that the instrument components perform adequately in capturing the constructs they represent (Kline, 2023). Among the latent factors, evaluation and reflection competencies (Factor A) emerged as the most influential, further reinforcing the centrality of reflective practice in teacher professional development. This finding aligns with a substantial body of literature that underscores reflection as a cornerstone of effective teaching and lifelong learning (Schön, 1983).

In addition, the composite reliability (CR) analysis reveals that four out of seven factors achieved high reliability (CR > 0.70), indicating strong internal consistency among their indicators. The evaluation and reflection factor, in particular, demonstrated the highest reliability, suggesting that the items in this domain are well-constructed and cohesively represent the underlying competency. On the other hand, three factors—professionalism,

technology integration, and communication—showed relatively lower reliability values (CR < 0.70), signaling the need for further refinement. While these values are still acceptable at an early stage of instrument development (Neumann et al., 2019), efforts to improve reliability through item revision or expansion would enhance the overall robustness of the assessment tool.

This research aligns well with previous theoretical frameworks and empirical findings that emphasize the importance of a comprehensive approach to teacher evaluation. Danielson's (2008) model, for example, highlights evaluation, activity design, and technological integration as key domains of teacher effectiveness—elements that are reflected in the structure of this instrument. Furthermore, the results support the view that teacher assessment should include dimensions such as communication, reflection, and professional development. These competencies are integral to a holistic understanding of teaching, particularly in early childhood settings where socio-emotional interaction, curriculum adaptation, and responsive practice are central to effective pedagogy.

The confirmatory factor analysis further supports the structural soundness of the instrument, with all standardized loading factors proving statistically significant. Although some indicators showed relatively lower contributions, their inclusion remains theoretically justified and consistent with real-world teaching practices. These results affirm the relevance and applicability of the instrument in assessing teacher competencies. Moreover, the structure supports an integrated model of teacher competence that aligns with early childhood education philosophies, which promote whole-child development and interdisciplinary approaches to teaching (Manggaberani & Putro, 2023; Rista et al., 2020). This instrument, therefore, offers not only a valid and reliable means of evaluation but also a conceptual model for professional growth and quality improvement in early childhood education.

Research Limitations and Future Directions

Despite its contributions, this study has several limitations. First, the sample used for both the EFA and CFA stages was limited to specific regions, potentially affecting the generalizability of the results to all ECE teachers in Indonesia. Future research should aim to expand the sample to include broader geographic areas to enhance the external validity of the instrument. Second, certain factors, such as communication and professionalism, require further refinement to improve reliability. Future studies could explore additional indicators or alternative scoring methods to enhance the precision of the measurements.

In addition, the present study relied primarily on self-assessment. Future research could incorporate multi-source assessments—such as peer reviews or supervisor evaluations—to provide a more comprehensive understanding of teacher competency (Alhassan & Ali, 2019). Conducting longitudinal studies would also offer valuable insights into how teachers' competencies develop over time, particularly in response to professional development programs. Addressing these areas would support ongoing efforts to improve teacher evaluation systems and ensure they effectively contribute to teacher development and student learning outcomes in early childhood education.

Conclusion

This study aimed to develop and validate a self-assessment instrument designed to measure the pedagogical and professional competencies of early childhood education (ECE) teachers in Indonesia. The 39-item instrument encompasses seven core dimensions: evaluation and reflection, holistic educational approaches, curriculum mastery, learning activity design, professionalism, technology integration, and communication. Exploratory factor analysis (EFA) revealed a strong factor structure, explaining 60.378% of the total variance, while confirmatory factor analysis (CFA) supported the model's fit with empirical data. Construct reliability analysis indicated high reliability (CR > 0.70) in four dimensions and moderate, yet acceptable, reliability in the remaining three.

The validated instrument offers a comprehensive framework for evaluating ECE teacher competencies and can serve as a practical tool for both performance evaluation and

professional reflection. Although some indicators require further refinement, the instrument holds potential as a reference for policy-making and the development of professional programs. These findings highlight the critical role of reflection-based evaluation in enhancing the quality of early childhood education. Future research should expand validation efforts across broader regions and diverse contexts and incorporate multi-source evaluation approaches to further strengthen the instrument.

Declarations

Author contribution statement

All authors contributed to this research. The first author developed the instruments and served as the corresponding author for the resulting publications. The second and third authors acted as experts in the development of the instruments and the formulation of new terms for emerging factors. The fourth author managed the financial aspects and reported the research findings. The fifth to seventh authors were responsible for designing the instruments in digital format.

Funding Statement

This research was supported by the Directorate of Research and Community Service, Yogyakarta State University, through a research group within the Department of Early Childhood Education, under contract number 41/UN34.11/KONTRAK-PEP/KU/2024.

References

Al Riyami, T. (2015). Reflection: Is it a promising or spurious tool for teachers' professional development? *International Journal of Bilingual & Multilingual Teachers of English*, *03*(01), 47–58. https://doi.org/10.12785/ijbmte/030105

Alhassan, A., & Ali, H. I. H. (2019). *Towards a multi-source performance appraisal model in omani higer education institutions*. *10*(4), 299–312.

Bank, W. (2020). The promise of education in Indonesia. In *The World Bank*. The World Bank. https://doi.org/10.1596/34807

Best, D. A. (2016). *Developing urban elementary teacher leaders using danielson's framework: An action reserach study*.

Boud, D. (2015). Feedback: Ensuring that it leads to enhanced learning. *The Clinical Teacher's Toolbox*, *12*, 3–7.

Buyamin, B. (2023). Implementation of teacher performance assessment program. *Porceeding of International Conference on Education, Society and Humanity*, *01*(01), 338–347. https://www.ejournal.unuja.ac.id/index.php/icesh/article/view/5646

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, *45*(6), 378–387. https://doi.org/10.3102/0013189X16659442

Damanik, R. (2019). Hubungan kompetensi guru dengan kinerja guru. *Jurnal Serunai Administrasi Pendidikan*, *8*(2). https://doi.org/10.37755/jsap.v8i2.170

Danielson, C. (2008). *Enhancing professional practice: A framework for teaching* (Second Edi). Association for Supervision and Curriculum Development (ASCD).

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44*(2), 132–137. https://doi.org/10.3102/0013189X15575346

Filgueiras, A., Pires, P., & Landeira-Fernandez, J. (2014). Screening measures used in child daycare centers: A 15-years systematic review. *Psychology*, *05*(19), 2109–2119. https://doi.org/10.4236/psych.2014.519213

Filipe, M. N. M., Ferreira, F. A. F., & Santos, S. P. (2015). A multiple criteria information system for pedagogical evaluation and professional development of teachers. *Journal of the Operational Research Society*, *66*(11), 1769–1782. https://doi.org/10.1057/jors.2014.129

Fireside, D., & Lachlan-Hache, L. (2015). *Uncommon measures : Teacher self-evaluation to*

encourage professional growth (Issue November).
Fitria, H., Kristiawan, M., & Rahmat, N. (2019). Upaya meningkatkan kompetensi guru melalui pelatihan penelitian tindakan kelas. *Abdimas Unwahas*, *4*(1), 14–25. https://doi.org/10.62509/ji.v1i1.20
Gates, B., & Gates, M. (2010). Learning about teaching. In *Bill & Melinda Gates Foundation*. https://doi.org/10.4324/9780203464014_chapter_2
Goh, A. Y. S. (2019). Rethinking reflective practice in professional lifelong learning using learning metaphors. *Studies in Continuing Education*, *41*(1), 1–16. https://doi.org/10.1080/0158037X.2018.1474867
Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3–28. https://doi.org/10.3102/0162373714523831
Hartono, S., & Muchtar, D. Y. (2018). Uji validitas konstruk pada instrumen dengan metode confirmatory factor analysis (CFA). *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, *6*(1), 79–86. https://doi.org/10.15408/jp3i.v6i1.8155
Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, *4*(1), 1–8. https://doi.org/10.1080/2331186X.2017.1304016
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Juma, A. A. (2024). Self-reflection in teaching: A comprehensive guide to empowering educators and enhancing student learning. *International Journal of Science and Research Archive*, *12*(1), 2835–2844. https://doi.org/10.30574/ijsra.2024.12.1.1113
Kane, T. J., Kerr, K. A., & Pianta, R. C. (2014). *Designing teacher evaluation systems*. Jossey-Bass A Wiley Brand.
Kelly, S., Bringe, R., Aucejo, E., & Fruehwirth, J. (2020). Using global observation protocols to inform research on teaching effectiveness and school improvement: Strengths and emerging limitations. *Education Policy Analysis Archives*, *28*, 1–34. https://doi.org/10.14507/EPAA.28.5012
Kementrian Pendidikan Nasional RI. (2014). Standar Nasional Pendidikan Anak Usia Dini No 137 Tahun 2014. In *Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia*. https://portaldik.id/assets/upload/peraturan/PERMEN KEMENDIKBUD Nomor 137 Tahun 2014 STANDAR NASIONAL PENDIDIKAN ANAK USIA DINI.pdf
Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly*, *52*(5), 711–753. https://doi.org/10.1177/0013161X16653445
Kusumawardhani, P. N. (2017). Does teacher certification program lead to better quality teachers? Evidence from Indonesia. *Education Economics*, *25*(6), 590–618. https://doi.org/10.1080/09645292.2017.1329405
Maghfuroh, L., Salimo, H., Qadrijati, I., & Subiyanto, A. A. (2020). Early childhood education teachers in Indonesia : Validity and reliability of preschool child development surveillance instruments. *Migration Letters*, *20*(8), 1079–1089.
Manggaberani, A. A., & Putro, N. H. (2023). Development of character assessment instrument on english learning for junior high school students. *Research and Development in Education (RaDEn)*, *4*(1), 374–389. https://doi.org/10.47006/ijierm.v5i3.282
Moss, J. T. (2015). *The danielson model of teacher evaluation: Exploring teacher perceptions concerning its value in shaping and improving instructional practice*.
Neumann, M. M., Anthony, J. L., Erazo, N. A., Neumann, D. L., Neumann, M. M., & Anthony, J. L. (2019). Assessment and technology : Mapping future directions in the early childhood classroom. *Frontiers in Education*, *4*(October), 1–13. https://doi.org/10.3389/feduc.2019.00116

Nur Efendi, & Muh Ibnu Sholeh. (2023). Manajemen pendidikan dalam meningkatkan mutu pembelajaran. *Academicus: Journal of Teaching and Learning*, *2*(2), 68–85. https://doi.org/10.59373/academicus.v2i2.25

Otaya, L. G., Kartowagiran, B., Retnawati, H., & Mustakim, S. S. (2020). Estimating the ability of pre-service and in-service Teacher Profession Education (TPE) participants using Item Response Theory. *Research and Evaluation in Education*, *6*(2), 160–173. https://doi.org/10.21831/reid.v6i2.36043

Öztabak, M. Ü., & Polatlar, D. Y. (2020). An overview of the administrative issues in preschools from preschool teachers' perspectives. *Journal of Education and Learning*, *9*(6), 183. https://doi.org/10.5539/jel.v9n6p183

Retnawati, H., Djidu, H., Kartianom, Apino, E., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thingking skills and its learning strategy. *Problem of Education in The 21st Century*, *76*(2), 215–230. http://repository.seku.ac.ke/handle/123456789/63%0Ahttps://hungary.pure.elsevier.com/en/publications/on-synergies-and-conflicts-between-the-sustainable-development-go

Rista, B. V, Agung, A. G., & Ujianti, P. R. (2020). Validity and reliability: An ability assessment instrument to identify the number concept. *Indonesian Journal of Educational Research and Review*, *3*(3), 144–152. https://ejournal.undiksha.ac.id/index.php/IJERR/article/view/30614%0Ahttps://ejournal.undiksha.ac.id/index.php/IJERR/article/download/30614/pdf

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, *19*(1), 80–110. https://doi.org/10.1177/1094428115598239

Sholeh, M. (2017). Keefektifan peran kepala sekolah dalam meningkatkan kinerja guru. *Jurnal Dinamika Manajemen Pendidikan*, *1*(1), 41. https://doi.org/10.26740/jdmp.v1n1.p41-54

Smith, C., & Gillespie, M. (2023). Research on professional Development and Teacher Change: Implications for Adult Basic Education. In *Review of adult learning and literacy: Vol.7: Connecting research, policy, and practice* (Vol. 7, pp. 205–244). https://doi.org/10.4324/9781003417996-7

Supadi, S., Soraya, E., Muhammad, H., & Halim, N. (2021). Making sense the teacher evaluation with the value of local culture (The voice of school principals in Indonesia). *International Journal of Educational Management*, *35*(6), 1124–1135. https://doi.org/10.1108/IJEM-08-2020-0408

Suranto, S., Muhyadi, M., & Mardapi, D. (2014). Pengembangan instrumen evaluasi uji kompetensi keahlian (ukk) administrasi perkantoran di smk. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 98–114. https://doi.org/10.21831/pep.v18i1.2127

Suryadarma, D., & Jones, G. W. (2015). Education in indonesia; Rising to the challenge. In *OECD*. https://doi.org/10.2307/3023860

Umar, J., & Nisa, Y. F. (2020). Uji validitas konstruk dengan CFA dan pelaporannya. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, *9*(2), 1–11. https://doi.org/10.15408/jp3i.v9i2.16964