Lisensi Creative Commons

Murhum : Jurnal Pendidikan Anak Usia Dini
e-ISSN: 2723-6390, hal. 2133-2147
Vol. 7, No. 2, Desember 2026
DOI: 10.37985/murhum.v7i2.2356

Developing and Validating an Ethnopedagogical Gross Motor Instrument through Traditional Songs for Preschool Children

Riska Maratus Sholikah[1], Ruqoyyah Fitri[2], Miftakhul Jannah[3], Mohammad Syahidul Haq[4], dan Emmanuel Etim Uye[5]

[1,2,3,4] Pendidikan Anak Usia Dini, Universitas Negeri Surabaya
[5] Psikologi Sosial dan Lingkungan, University of Ibadan

ABSTRAK. *Penelitian ini mengembangkan dan memvalidasi instrumen penilaian motorik kasar berbasis etnopedagogi melalui lagu tradisional untuk anak prasekolah usia 4-5 tahun. Penelitian menggunakan desain pengembangan instrumen kuantitatif yang mengacu pada prinsip International Test Commission yang dilaksanakan pada maret sampai mei 2026. Tahapan penelitian meliputi analisis kebutuhan lapangan, definisi konstruk, penyusunan indikator, pengembangan butir, validasi ahli, pelatihan penilai, dan uji coba pada 100 anak dari empat satuan pendidikan anak usia dini. Data dikumpulkan melalui lembar validasi ahli, observasi kinerja anak, dan catatan lapangan. Validitas isi dianalisis menggunakan Aiken's V, validitas butir menggunakan korelasi item-total, konsistensi internal menggunakan Cronbach's Alpha, serta reliabilitas antarpenilai menggunakan ICC dan Cohen's Kappa. Hasil menunjukkan Aiken's V sebesar 0,87-0,93, korelasi item-total 0,57-0,67, Cronbach's Alpha 0,86, ICC 0,83-0,86, dan Kappa 0,78-0,81. Instrumen mengintegrasikan Cublak-cublak Suweng, Sluku-sluku Bathok, dan Gundul-gundul Pacul untuk menilai koordinasi, keseimbangan, dan kelincahan secara kontekstual. Temuan ini menunjukkan bahwa asesmen motorik kasar dapat tetap kuat secara psikometrik sekaligus ekologis dan responsif budaya.*

Kata Kunci : *Keterampilan Motorik Kasar; Etnopedagogi; Lagu Tradisional; Penilaian*

ABSTRACT. *This study developed and validated an ethnopedagogical gross motor assessment instrument using traditional songs for preschool children aged 4-5 years. A quantitative instrument development design was used, guided by International Test Commission principles* were conducted from March to May 2026*. The procedure covered field needs analysis, construct definition, indicator specification, item development, expert validation, rater training, and pilot testing with 100 children from four early childhood education institutions. Data were collected through expert validation sheets, performance-based observation, and field notes. Content validity was examined using Aiken's V, item validity using item-total correlations, internal consistency using Cronbach's Alpha, and inter-rater reliability using ICC and Cohen's Kappa. The findings showed strong psychometric quality: Aiken's V = 0.87-0.93, item-total correlations = 0.57-0.67, Cronbach's Alpha = 0.86, ICC = 0.83-0.86, and Kappa = 0.78-0.81. The instrument embeds Cublak-cublak Suweng, Sluku-sluku Bathok, and Gundul-gundul Pacul to assess coordination, balance, and agility in culturally familiar activities. The study contributes a culturally responsive assessment model that maintains psychometric rigor while improving ecological validity in early childhood motor assessment.*

Keyword : *Gross Motor Skills; Ethnopedagogy; Traditional Songs; Assessment*

Copyright (c) 2026 Riska Maratus Sholikah dkk.

✉ Corresponding author : Riska Maratus Sholikah
Email Address : 25011545002@mhs.unesa.ac.id
Received 2 Juni 2026, Accepted 19 November 2026, Published 19 November 2026

INTRODUCTION

Gross motor skills are a central component of early childhood development because they support physical activity, self-regulation, and later participation in structured learning [1], [2]. These skills are also closely associated with children's broader motor competence and developmental readiness [3]. For children aged 4–5 years, assessment is needed not only to document movement performance but also to help teachers identify developmental needs and plan appropriate stimulation. Therefore, a valid and reliable assessment instrument must represent the construct of gross motor competence while remaining suitable for the learning context in which children actually move and play[4], [5] .

Existing standardized instruments, including TGMD-3, provide important psychometric references for measuring fundamental motor skills [6]–[8]. However The validity and reliability of TGMD-3 have also been examined with Indonesian children [9]. Nevertheless, standardized assessment tasks may not always reflect how movement competence is naturally displayed in children's everyday play and culturally familiar learning activities. In Indonesian early childhood settings, this creates an ecological-validity concern: an instrument may demonstrate sound psychometric properties while still requiring careful consideration of children's lived cultural experiences [10], [11].

Preliminary field observations and informal teacher interviews at the participating early childhood education setting showed that teachers commonly observed children's gross motor performance during play, singing, and movement activities. However, they did not use a structured rubric with clear indicators and scoring criteria. Teachers also reported that traditional songs were frequently used to engage children, yet these activities were not systematically converted into assessment items. These initial findings indicate the need for an instrument that can translate culturally familiar activities into measurable gross motor indicators.

Previous studies have mainly examined two separate areas. The first group of studies focuses on music and rhythm based interventions to improve children's motor development [12]–[14]. Related studies have investigated movement, dance, and music-integrated activities that support children's physical and motor abilities [15]–[17]. Other intervention studies have examined physical activity and traditional play as strategies for stimulating motor development [18]–[20]. The second group focuses on instrument development and psychometric validation, including standardized gross motor assessments and measurement models[6], [7]. However, local cultural practices are not always positioned as the core assessment context in these studies [21], [22]. The present study addresses the gap between these two areas by developing an assessment instrument that is both psychometrically tested and culturally embedded.

The conceptual framework of this study links ethnopedagogy, traditional songs, observable movement tasks, and psychometric validation [10]. Ethnopedagogy provides the cultural-pedagogical foundation, while culturally rooted songs and play practices strengthen the local relevance of the assessment context [23], [24]. Tembang dolanan provides rhythmic and social movement contexts that enable children to demonstrate their motor competence through familiar activities [25], [26]. These contexts are

translated into indicators of coordination, balance, and agility, and the resulting items are evaluated through content validity, item validity, internal consistency, and inter-rater reliability [22], [27], [28].

Conceptual Framework



Figure 1. Conceptual framework of the ethnopedagogical gross motor assessment instrument

Accordingly, this study aims to develop and validate a culture-based gross motor assessment instrument for children aged 4–5 years by integrating traditional songs within an ethnopedagogical framework [12]–[14]. The novelty of this study lies in positioning traditional songs not merely as learning media or intervention activities but as the core measurement context for assessing gross motor competence [15]–[17]. Unlike prior instrument adaptation and standardized assessment studies, this research positions cultural practices as the central measurement context rather than as an additional contextual feature [6], [7], [21]. The proposed approach therefore offers a culture-embedded assessment model that seeks to balance psychometric rigor and ecological validity.

METHOD

This study employed a quantitative approach using an instrument development design, guided by the International Test Commission (ITC) guidelines [27], [28]. Unlike traditional research and development models such as Borg and Gall, which focus on procedural stages of product development [22], the present study emphasizes psychometric principles to ensure the quality and validity of the developed instrument. These guidelines highlight key principles of measurement quality, including construct clarity, validity, reliability, and contextual appropriateness. In line with contemporary instrument development practices, this study also considers both process-oriented and product-oriented measurement approaches to capture children's motor performance more comprehensively, ensuring that observed behaviors reflect both movement quality and performance outcomes [29], [30].

The development process consisted of several stages: (1) preliminary needs analysis, (2) construct definition, (3) indicator specification, (4) item development, (5) expert validation and item revision, (6) pilot testing and psychometric analysis, and (7) instrument finalization. The construct of the instrument was based on Fundamental Motor Skills (FMS), which include locomotor, non-locomotor, and manipulative skills [4], [31]. These constructs served as the foundation for developing observable indicators aligned with the developmental characteristics of children aged 4–5 years. Furthermore, the construct specification was informed by global evidence indicating that FMS should be assessed as a multidimensional construct involving coordination, balance, and agility as core components of early motor competence [3], [20].

The research activities were conducted from March to May 2026. Preliminary needs analysis, construct definition, indicator specification, and initial item development were conducted in March 2026. Expert validation, item revision, and rater training were conducted in April 2026, while pilot testing, psychometric analysis, and instrument finalization were conducted in April–May 2026. The complete development procedure is presented in Figure 2.

Research Design and instrument Development Stages

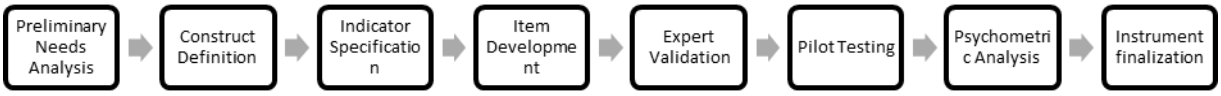


Figure 2. Research design and stages of instrument development

The instrument was designed as an observation-based assessment tool, embedded within movement and music activities using traditional songs (tembang dolanan), including Cublak-cublak Suweng, Sluku-sluku Bathok, and Gundul-gundul Pacul. This integration aimed to enhance the ecological validity and contextual relevance of the assessment [25], [32]. Embedding assessment within culturally familiar activities is consistent with recent findings that authentic, play-based, and rhythm-integrated environments can improve the accuracy of motor skill assessment by reducing artificial task constraints [15], [33]. To ensure content quality, the instrument underwent expert validation, followed by a pilot study to evaluate item clarity, feasibility, and initial psychometric properties. This procedure aligns with previous studies emphasizing the importance of systematic validation in instrument development [7], [34]. Additionally, this study follows recent recommendations in psychometric research emphasizing the importance of preliminary pilot testing before large-scale validation to ensure item functioning and feasibility in real classroom contexts [35].

The study was conducted from March to May 2026 at four private early childhood education institutions located in Sukolilo District, Surabaya City, East Java, Indonesia: TK A, TK B, TK C, and TK D. These institutions constitute formal kindergarten settings serving children in an urban area of Surabaya. The assessment activities were conducted in the children's regular classrooms or school play areas during movement and music based learning activities to maintain natural and ecologically valid assessment conditions. According to the official Kemendikdasmen education database accessed on April 23, 2026, the four institutions had a combined enrollment of 300 children. These enrollment figures were used to describe the institutional context and sampling frame and should not be interpreted as the demographic distribution of the 100 children included in the study [36].

Table 1. Research locations, institutional enrollment, and sample allocation

Institution	NPSN	Official enrollment	Research sample
TK A	20559761	43	25
TK B	69946714	52	25
TK C	20559773	124	25
TK D	69948191	81	25
Total		300	100

The participants consisted of two groups: expert validators and children involved in the pilot study. The expert validators included three individuals: two lecturers in early childhood education with a minimum of a master's degree and a kindergarten teacher with at least five years of experience in movement and music based learning. Their role was to evaluate the relevance, clarity, and representativeness of the instrument items in relation to the construct being measured. The pilot study involved 100 children aged 4–5 years. The participating institutions were selected purposively based on their location in Sukolilo District, institutional permission, availability of children who met the eligibility criteria, and suitability for conducting movement- and music-based assessment activities. A quota-based purposive sampling procedure was then used to recruit 25 eligible children from each institution. The use of a sample of 100 children was intended to improve the stability of the initial psychometric estimates, particularly internal consistency and inter-rater reliability coefficients [37]. The final sample consisted of 40 boys and 60 girls. Based on chronological age at the time of assessment, 50 children were four years old and 50 children were five years old.

Table 2. Participant demographic characteristics and sampling criteria

Component	Description
Research location	Four private kindergartens in Sukolilo District, Surabaya City, East Java, Indonesia
Research period	March–May 2026
Sample size	100 children aged 4–5 years
Institutional distribution	25 children from each participating institution
Gender distribution	Boys: 40; Girls: 60
Age distribution	Four years old: 50; Five years old: 50
Sampling technique	Quota-based purposive sampling
Number of raters	Two trained raters
Assessment duration	Approximately 20–25 minutes per child

The inclusion criteria were: (1) children aged 4–5 years at the time of data collection; (2) enrollment in one of the four participating institutions; (3) regular attendance during the research period; (4) physical ability to participate in routine gross motor activities; (5) written parental or legal guardian consent; and (6) willingness to participate in the assessment activities. The exclusion criteria were: (1) medical or physical restrictions that prevented participation in gross motor activities; (2) absence during both the original and rescheduled assessment sessions; and (3) incomplete performance or assessment data. Data were collected using three techniques: (1) expert validation sheets to assess content validity, (2) observation sheets to measure children's gross motor performance during activities, and (3) field notes to document implementation processes and children's responses.

The observation instrument used a 4-point rating scale consisting of 1 = poor, 2 = fair, 3 = good, and 4 = very good, representing levels of children's motor skill achievement. Observations were conducted in natural learning settings to ensure authentic assessment conditions [32]. This approach is consistent with recommendations that early childhood motor assessment should be conducted in ecologically valid environments to better reflect children's natural movement behaviors [38]. The complete assessment required approximately 20–25 minutes per child, including activity instructions, movement demonstrations, performance of the three

song-based movement tasks, and recording of scores. Children who were absent from the initial assessment were given one rescheduled assessment opportunity during the data collection period. Only children with complete scores from both raters were included in the reliability analysis. Before assessment, permission was obtained from the heads of the participating institutions, and written informed consent was obtained from the children's parents or legal guardians. The children were given age appropriate explanations of the activities and were allowed to discontinue their participation at any time.

Data analysis was conducted using descriptive quantitative techniques to evaluate the psychometric quality of the instrument, including content validity and reliability. Data coding, descriptive statistics, corrected item-total correlation, Cronbach's Alpha, and inter-rater reliability analyses were conducted using JASP version 0.98.8 [39]. The level of statistical significance was set at ($p<0.05$), and reliability coefficients were reported with 95% confidence intervals where available. Content validity was analyzed using Aiken's V, based on expert judgments. This index measures the degree of agreement among experts regarding the relevance of each item to the construct. Aiken's V was calculated using the following formula:

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

where $s = r - l_o$, (r) is the score assigned by an expert, l_o is the lowest possible score, n is the number of expert validators, and (c) is the number of rating categories. Items were considered valid if they achieved a coefficient of $V \geq 0.80$ [34], [40]. Items with coefficients below 0.80 were revised based on the validators' qualitative comments and were reconsidered before pilot testing.

Item functioning was examined using corrected item-total correlation. An item was retained when it produced a corrected item-total correlation of at least 0.30 and remained conceptually consistent with the dimension being measured. Items below this criterion were examined for unclear wording, overlapping indicators, scoring ambiguity, or limited score variation before a decision was made to revise or remove them [41].Reliability was assessed using Cronbach's Alpha to determine the internal consistency of the instrument. Cronbach's Alpha was calculated using the following formula:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right)$$

Where (k) is the number of instrument items, S_i^2 is the variance of each item, and S_t^2 is the variance of the total instrument score. A coefficient of $\alpha \geq 0.70$ was considered acceptable, indicating adequate consistency among items measuring the same construct [7]. The use of Aiken's V and reliability testing follows established procedures in educational instrument validation [41].

In addition to internal consistency, inter-rater reliability was examined using the Intraclass Correlation Coefficient (ICC) and Cohen's Kappa to assess the level of agreement between raters. ICC was used to evaluate the consistency of continuous

scores across raters, while Cohen's Kappa was applied to examine agreement in categorical scoring levels [42], [43]. A two-way random effects, absolute agreement, single measure model ICC(2,1) was applied because every child was independently assessed by the same two raters and the analysis focused on agreement between individual ratings. ICC was calculated using the following formula:

$$ICC(2,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k(MS_C - MS_E)}{n}}$$

where MSR is the mean square for participants or rows, MSC is the mean square for raters or columns, MSE is the residual error mean square, k is the number of raters, and n is the number of participants. The use of ICC is widely recommended in reliability studies involving observational data, as it provides a robust estimate of agreement beyond simple correlation [42]. Because the observational ratings consisted of four ordered categories, weighted Cohen's Kappa was used to evaluate agreement between the two raters at the categorical level. Cohen's Kappa was calculated using the following formula:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the observed proportion of agreement and P_e is the proportion of agreement expected by chance. Cohen's Kappa offers a complementary measure by accounting for chance agreement in categorical ratings, making it particularly suitable for performance-based assessment instruments [37].

The decision criteria for evaluating the validity and reliability evidence were organized systematically as presented in Table 3.

Table 3. Decision criteria for validity and reliability analysis

Analysis	Coefficient or criterion	Interpretation and decision
Aiken's V	V ≥ 0.80	Strong content validity; item retained
Aiken's V	V < 0.80	Item reviewed and revised
Corrected item-total correlation	≥ 0.30	Acceptable item discrimination; item retained
Corrected item-total correlation	< 0.30	Item reviewed, revised, or removed
Cronbach's Alpha	α ≥ 0.70	Acceptable internal consistency
ICC	<0.50	Poor reliability
ICC	0.50–0.74	Moderate reliability
ICC	0.75–0.90	Good reliability
ICC	>0.90	Excellent reliability
Cohen's Kappa	0.01–0.20	Slight agreement
Cohen's Kappa	0.21–0.40	Fair agreement
Cohen's Kappa	0.41–0.60	Moderate agreement
Cohen's Kappa	0.61–0.80	Substantial agreement
Cohen's Kappa	0.81–1.00	Almost perfect agreement

Accordingly, Aiken's V ≥ 0.80 indicated strong content validity; corrected item-total correlation ≥ 0.30 indicated acceptable item functioning; Cronbach's Alpha ≥ 0.70 indicated acceptable internal consistency; ICC ≥ 0.75 indicated good inter-rater reliability; and Cohen's Kappa values of 0.61–0.80 and above 0.80 indicated substantial and almost perfect agreement, respectively[7], [37], [42].

RESULTS AND DISCUSSION

The instrument development process produced a blueprint consisting of three gross motor aspects, fifteen observable items, and a 4-point scoring rubric. Figure 3 summarizes the transition from cultural activity selection to validated assessment output.

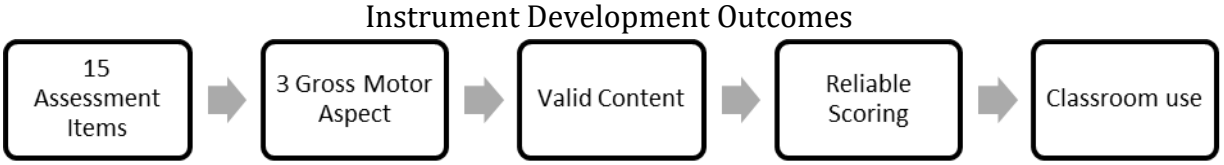


Figure 3. Summary of instrument development outcomes

Table 4. Instrument Blueprint

No	Aspect	Indicator	Activity (Tembang Dolanan)	Item	Assessment Type
1	Coordination	Synchronizing body movement with rhythm	Cublak-cublak Suweng	1-5	Performance based observation
2	Balance	Maintaining body stability (static & dynamic)	Sluku-sluku Bathok	6-10	Performance based observation
3	Agility	Changing movement quickly and accurately	Gundul-gundul Pacul	11-15	Performance based observation

Content validity was evaluated using Aiken's V index based on expert judgments. The results indicate that all items achieved coefficients ranging from 0.87 to 0.93, exceeding the minimum acceptable threshold ($V \geq 0.80$).

Table 5. Content Validity Results (Aiken's V)

No	Item Description	V
1	Clapping while walking following rhythm	0.90
2	Clapping body parts according to beat	0.91
3	Imitating simple movement patterns	0.89
4	Controlled object manipulation	0.87
5	Selective body movement based on rhythm	0.92
6	Standing on one foot	0.90
7	Walking on a straight line	0.91
8	Shifting body without losing balance	0.88
9	Landing with both feet	0.89
10	Squat-stand movement control	0.90
11	Running without collision	0.88
12	Changing direction with tempo	0.91
13	Moving through obstacles	0.87
14	Galloping/side movement	0.89
15	Responding to stop/go signal	0.93

A closer examination of Table 2 reveals that items related to rhythmic coordination (e.g., Item 15: responding to stop/go signal, V = 0.93) consistently achieved higher validity scores compared to items involving object manipulation (e.g., Item 4: controlled object manipulation, V = 0.87). This pattern indicates that experts perceived rhythm-based movements as more clearly observable and contextually relevant within the tembang dolanan framework.

This finding demonstrates a high level of agreement among experts regarding the relevance and clarity of the instrument items. Importantly, this variation should not be interpreted as a weakness of the instrument, but rather as evidence that construct representation is influenced by the ecological characteristics of the assessment context.

In this case, tembang dolanan inherently emphasize rhythmic and coordinated movement, while offering more limited affordances for object manipulation [34], [40]. From a theoretical standpoint, this finding supports the argument that content validity in culturally embedded instruments is context-dependent, challenging the assumption of uniform construct representation across different assessment environments [30].

Item Validity Analysis, following content validation, item-level analysis was conducted to examine the contribution of each item to the overall construct measurement. Item validity was further examined using item-total correlation analysis.

Table 6. Item Validity Results

No	Item	r-value	r-table (0.05)	Remark
1	Item 1	0.61	0.44	Valid
2	Item 2	0.63	0.44	Valid
3	Item 3	0.60	0.44	Valid
4	Item 4	0.58	0.44	Valid
5	Item 5	0.65	0.44	Valid
6	Item 6	0.62	0.44	Valid
7	Item 7	0.64	0.44	Valid
8	Item 8	0.59	0.44	Valid
9	Item 9	0.60	0.44	Valid
10	Item 10	0.63	0.44	Valid
11	Item 11	0.57	0.44	Valid
12	Item 12	0.66	0.44	Valid
13	Item 13	0.58	0.44	Valid
14	Item 14	0.61	0.44	Valid
15	Item 15	0.67	0.44	Valid

The results show that all items have correlation coefficients exceeding the critical value (r = 0.44), indicating that each item contributes significantly to measuring the underlying construct. As shown in Table 3, item-total correlation values range from 0.57 to 0.67, with Item 15 showing the highest contribution (r = 0.67) and Item 11 the lowest (r = 0.57). This relatively narrow range suggests a balanced contribution across items, without any item dominating or weakening the construct measurement. The relatively balanced correlation values (0.57–0.67) indicate that the instrument captures gross motor skills as a multidimensional construct, consistent with the fundamental motor skills framework [4], [31].

Reliability of the Instrument, the reliability analysis yielded a Cronbach's Alpha coefficient of 0.86, indicating good internal consistency.

Table 7. Reliability Results

Variable	Number of Items	Sample Size	Cronbach's Alpha	Category
Gross Motor	15	100	0.86	Reliable

It is important to note that Table 4 consistently reports an alpha value of 0.86 for a sample of 100 children, indicating that the reliability estimate is based on a sufficiently stable dataset. Inter-Rater Reliability (ICC & Cohen's Kappa), given that the instrument relies on observational scoring, inter-rater reliability analysis is essential to ensure that the assessment results are not dependent on individual rater subjectivity. The results indicate high to very high agreement among raters across all aspects, with ICC values ranging from 0.83 to 0.86 and Cohen's Kappa values ranging from 0.78 to 0.81.

Table 8. Inter-Rater Reliability Results

No	Aspect	ICC	Cohen's Kappa	Interpretation

1	Coordination	0.84	0.79	High agreement
2	Balance	0.86	0.81	Very high agreement
3	Agility	0.83	0.78	High agreement

A closer examination of Table 5 shows that the balance aspect achieved the highest agreement (ICC = 0.86; Kappa = 0.81), indicating that static and dynamic stability movements are relatively easier to observe and score consistently compared to more complex movement patterns. In contrast, agility (ICC = 0.83; Kappa = 0.78) shows slightly lower agreement, which may reflect the dynamic and rapid nature of movement changes that are more difficult to evaluate in real-time. This pattern suggests that different motor components present varying levels of observational complexity. The alignment between ICC and Cohen's Kappa further strengthens the reliability evidence, as both indices consistently indicate strong agreement, suggesting that the instrument produces stable measurements both in continuous scoring and categorical classification. This result is particularly significant because it demonstrates that observational assessment conducted in naturalistic and culturally embedded settings can still achieve high reliability, addressing common concerns regarding subjectivity in performance-based evaluation [42].

Rubric Scoring, based on the validity and reliability results, the finalized instrument consists of fifteen items from three indicators assessed using a 4 point scale. The use of a 4-point scale without a neutral midpoint was intentionally designed to reduce central tendency bias and encourage more discriminative judgments by raters, particularly in early childhood performance assessment [29].

Table 9. Rubric Scoring

No	Score	Description
1	4 (Very Good)	Performs movement accurately, consistently, and in full synchronization with rhythm
2	3 (Good)	Performs movement correctly with minor inconsistencies
3	2 (Fair)	Performs movement but lacks control or rhythm accuracy
4	1 (Poor)	Unable to perform movement appropriately

A closer examination of the rubric shows a progressive differentiation across four performance levels, ranging from complete mastery (score 4) to inability to perform (score 1). This structure reflects a concise and practical instrument that can be directly implemented in early childhood learning contexts. From a measurement perspective, this rubric aligns with performance-based assessment principles, where scoring is based on the quality of observable actions rather than abstract judgment, thereby improving both reliability and ecological validity [30].

Operational Definitions, based on the content validity of the instrument, the operational definition of each aspect is obtained in the table below.

Table 10. Operational Definitions of Each Aspect

No	Aspect	Descriptio Operational Definition
1	Coordination	The ability to synchronize body movements with musical rhythm and perform controlled motor actions
2	Balance	The ability to maintain body stability in both static and dynamic positions
3	Agility	The ability to change body position quickly and accurately in response to stimuli

To enhance measurement clarity, each operational definition is linked to observable movement indicators within the assessment activities, ensuring that raters

evaluate performance based on specific behavioral evidence rather than subjective interpretation. By clearly defining each construct at the operational level, the instrument strengthens both validity and reliability, as it minimizes ambiguity and ensures consistent interpretation across different raters and contexts [35].

Contextualized Implementation through Tembang Dolanan. The integration of the instrument within movement and music activities based on traditional tembang dolanan including Cublak-cublak Suweng, Sluku-sluku Bathok, and Gundul-gundul Pacul. Enables assessment to occur within authentic and meaningful contexts. These songs provide structured rhythmic patterns and culturally familiar movement sequences that facilitate children's engagement. Specifically, Cublak-cublak Suweng supports coordination and rhythm, Sluku-sluku Bathok emphasizes balance, and Gundul-gundul Pacul promotes agility. The developed 15-item instrument demonstrated strong preliminary psychometric performance: all content-validity coefficients exceeded the predetermined criterion (Aiken's V = 0.87–0.93), item-total correlations ranged from 0.57 to 0.67, internal consistency reached $\alpha = 0.86$, and inter-rater coefficients indicated high agreement (ICC = 0.83–0.86; Cohen's Kappa = 0.78–0.81). The reliability evidence also aligns with studies of established gross motor measures such as TGMD-3, although the present instrument differs by embedding the assessment tasks in traditional-song activities instead of relying solely on standardized, decontextualized tasks [7], [8], [43].

In this study, Cublak-cublak Suweng, Sluku-sluku Bathok, and Gundul-gundul Pacul functioned as repeatable contexts for eliciting and scoring coordination, balance, and agility. Recent motor-assessment research emphasizes the need for developmentally appropriate measures [44]; the specific contribution of the present study is to demonstrate an initial psychometric basis for measuring motor performance through culturally familiar activities. However, the results establish feasibility and preliminary validity in the studied setting and should not be interpreted as evidence that the instrument is superior to standardized gross motor tests. The variation in content-validity coefficients provides additional evidence about how the assessment context shaped item observability. Item 15, responding to a stop/go signal, obtained the highest Aiken's V value (0.93), whereas Item 4, controlled object manipulation, obtained the lowest value (0.87). In contrast, object manipulation depends on additional elements such as object type, force, trajectory, hand-eye coordination, and spatial control, while the selected songs provide fewer natural opportunities to display these behaviors [34], [40].

The aspect-level agreement results further clarify the observational characteristics of the instrument. Balance produced the highest agreement (ICC = 0.86; Kappa = 0.81), while agility produced slightly lower values (ICC = 0.83; Kappa = 0.78). Static and controlled dynamic balance behaviors offer relatively stable visual criteria, whereas rapid changes of direction require raters to judge timing, accuracy, and spatial response simultaneously, which increases observational complexity [38], [42]. Nevertheless, all coefficients met the criteria for good ICC reliability and substantial-to-almost-perfect Kappa agreement [37], [43]. For early childhood educators, the findings support a practical classroom use of the instrument as a structured observation tool.

Teachers can administer the 15 items during familiar movement-and-song activities, score each child using the four performance levels, and compare patterns across coordination, balance, and agility. The resulting profile can help teachers identify movements requiring additional stimulation, plan targeted play activities, and document progress over repeated observations without separating assessment from the normal learning routine.

CONCLUSION

This study achieved its objective by developing and obtaining preliminary validity and reliability evidence for a 15-item, culture-based gross motor assessment instrument for children aged 4–5 years using three traditional tembang dolanan. All items met the content- and item-validity criteria (Aiken's V = 0.87–0.93; r = 0.57–0.67), internal consistency was good (Cronbach's α = 0.86), and the two trained raters demonstrated high agreement (ICC = 0.83–0.86; Cohen's Kappa = 0.78–0.81). The main contribution of the study is the use of traditional songs as the measurement context, rather than merely as intervention media, for observing coordination, balance, and agility in natural early childhood learning activities. In practice, the four-point rubric can support teachers in documenting children's performance and planning targeted gross motor stimulation. Further multisite research should expand manipulative-skill items and examine factor structure, test-retest reliability, and cross-cultural applicability before the instrument is used beyond the population and cultural context examined in this study. The findings support a more precise theoretical proposition: cultural activities can operate as part of the measurement structure when they provide observable, repeatable, and developmentally appropriate motor tasks.

AWARD

The author would like to thank the expert validators who provided input and assessments in the instrument development process, as well as the teachers and children in one of the kindergartens who participated in this research.

REFERENCE

[1] L. E. Bolger *et al.*, "Global levels of fundamental motor skills in children: A systematic review," *J. Sports Sci.*, vol. 39, no. 7, hal. 717–753, Apr 2021, doi: 10.1080/02640414.2020.1841405.

[2] C. M. Capio, C. S. Y. Choi, dan R. S. W. Masters, "Association of working memory with gross motor skills in early childhood," *Int. J. Sport Exerc. Psychol.*, vol. 21, no. 6, hal. 992–1005, Nov 2023, doi: 10.1080/1612197X.2022.2098355.

[3] C. Malambo, A. Nová, C. Clark, dan M. Musálek, "Associations between Fundamental Movement Skills, Physical Fitness, Motor Competency, Physical Activity, and Executive Functions in Pre-School Age Children: A Systematic Review," *Children*, vol. 9, no. 7, hal. 1059, Jul 2022, doi: 10.3390/children9071059.

[4] S. L. C. Veldman, J. S. Gubbels, A. S. Singh, J. M. Koedijker, M. J. M. Chinapaw, dan T.

M. Altenburg, "Correlates of Fundamental Motor Skills in the Early Years (0–4 Years): A Systematic Review," *J. Mot. Learn. Dev.*, vol. 12, no. 1, hal. 1–47, Apr 2024, doi: 10.1123/jmld.2023-0003.

[5] H. Q. Song, P. W. C. Lau, dan J. J. Wang, "Investigation of the motor skills assessments of typically developing preschool children in China," *BMC Pediatr.*, vol. 22, no. 1, hal. 84, Des 2022, doi: 10.1186/s12887-021-03098-w.

[6] A. Griffiths, R. Toovey, P. E. Morgan, dan A. J. Spittle, "Psychometric properties of gross motor assessment tools for children: a systematic review," *BMJ Open*, vol. 8, no. 10, hal. e021734, Okt 2018, doi: 10.1136/bmjopen-2018-021734.

[7] S. Salami, P. F. R. Bandeira, C. M. A. Gomes, dan P. S. Dehkordi, "The Test of Gross Motor Development—Third Edition: A Bifactor Model, Dimensionality, and Measurement Invariance," *J. Mot. Learn. Dev.*, vol. 10, no. 1, hal. 116–131, Apr 2022, doi: 10.1123/jmld.2020-0069.

[8] E. K. Webster dan D. A. Ulrich, "Evaluation of the Psychometric Properties of the Test of Gross Motor Development—Third Edition," *J. Mot. Learn. Dev.*, vol. 5, no. 1, hal. 45–58, Jun 2017, doi: 10.1123/jmld.2016-0003.

[9] W. I. Rizkyanto, I. Gani, A. Iswanto, D. Yudhistira, dan M. I. Shahril, "Validity and reliability of the Gross Motor Development III Test for Indonesian children," *Fizjoterapia Pol.*, vol. 24, no. 2, hal. 171–177, Jun 2024, doi: 10.56984/8ZG56083HS.

[10] S. A. Sakti, S. Endraswara, dan A. Rohman, "Revitalizing local wisdom within character education through ethnopedagogy apporach: A case study on a preschool in Yogyakarta," *Heliyon*, vol. 10, no. 10, hal. e31370, Mei 2024, doi: 10.1016/j.heliyon.2024.e31370.

[11] T. Lawson, S. Yousaf, A. Dizon, I. Burks, dan C. Longo, "Educational Outcomes, Culturally Responsive Practices, and Identity: A Review of Literature," *School Psych. Rev.*, vol. 55, no. 5, hal. 724–740, Sep 2026, doi: 10.1080/2372966X.2026.2637188.

[12] Y. Ji dan X. Li, "A play-based integrated curriculum with synchronous music efficiently enhance Children's fundamental movement skills (FMS)," *Learn. Instr.*, vol. 97, hal. 102086, Jun 2025, doi: 10.1016/j.learninstruc.2025.102086.

[13] H. E. Kragness, L. Anderson, E. Chow, M. Schmuckler, dan L. K. Cirelli, "Musical groove shapes children's free dancing," *Dev. Sci.*, vol. 26, no. 1, hal. e13249, Jan 2023, doi: 10.1111/desc.13249.

[14] H. Zhao *et al.*, "Effects of 8 weeks of rhythmic physical activity on gross motor movements in 4-5-year-olds: A randomized controlled trial," *J. Exerc. Sci. Fit.*, vol. 22, no. 4, hal. 456–462, Okt 2024, doi: 10.1016/j.jesf.2024.10.001.

[15] M. del Mar Bernabé-Villodre *et al.*, "Effectiveness of a movement and music programme on activity, sedentary, and sleep behaviours and motor and musical skills in young children: protocol for the MoviMusi cluster randomised controlled trial," *J. Act. Sedentary Sleep Behav.*, vol. 3, no. 1, hal. 28, Nov 2024, doi: 10.1186/s44167-024-00065-5.

[16] P. Radiah Wakulu dan N. Priyanti, "The Development of Music-Accompanied Gymnastics to Improve Motor Skills in Early Childhood," *J. Islam. Educ. Students*, vol. 5, no. 1, hal. 98–112, Apr 2025, doi: 10.31958/jies.v5i1.13737.

[17] S. R. O'Connell, J. E. Nave-Blodgett, G. E. Wilson, E. E. Hannon, dan J. S. Snyder, "Elements of musical and dance sophistication predict musical groove perception," *Front. Psychol.*, vol. 13, hal. 998321, Nov 2022, doi: 10.3389/fpsyg.2022.998321.

[18] S. Calero-Morales, G. C. Vinueza-Burgos, C. L. Yance-Carvajal, dan W. J. Paguay-Balladares, "Gross Motor Development in Preschoolers through Conductivist and Constructivist Physical Recreational Activities: Comparative Research," *Sports*, vol. 11, no. 3, hal. 61, Mar 2023, doi: 10.3390/sports11030061.

[19] A. P. Dobell, M. A. Faghy, A. Pringle, dan C. M. P. Roscoe, "Improving Fundamental Movement Skills during Early Childhood: An Intervention Mapping Approach," *Children*, vol. 10, no. 6, hal. 1004, Jun 2023, doi: 10.3390/children10061004.

[20] D. Zhang, K. G. Soh, Y. M. Chan, dan Z. Zaremohzzabieh, "Effect of intervention programs to promote fundamental motor skills among typically developing children: A systematic review and meta-analysis," *Child. Youth Serv. Rev.*, vol. 156, hal. 107320, Jan 2024, doi: 10.1016/j.childyouth.2023.107320.

[21] Y. Chen *et al.*, "Developing a Scale for Measuring the Fundamental Movement Skills of Preschool Children in China," *Int. J. Environ. Res. Public Health*, vol. 19, no. 21, hal. 14257, Nov 2022, doi: 10.3390/ijerph192114257.

[22] I. W. Gunartha, T. Sulaiman, S. P. Suardiman, dan B. Kartowagiran, "Developing instruments for measuring the level of early childhood development," *REID (Research Eval. Educ.*, vol. 6, no. 1, hal. 1–9, Jun 2020, doi: 10.21831/reid.v6i1.21996.

[23] S. S. Sinaga, "Children's Trilingual Songs (Indonesian, English, and Javanese) as Media for Strengthening the Diversity Character of Elementary School Students," *Harmon. J. Arts Res. Educ.*, vol. 22, no. 2, hal. 392–404, Des 2022, doi: 10.15294/harmonia.v22i2.41585.

[24] N. Purba, D. C. Ovami, M. K. A. Kamaruddin, W. Hayati, dan T. R. Novita, "Revitalization of Socio-Cultural Based Dolanan in the Formation of Children's Character in Javanese Families in Binjai, Indonesia," *El-Usrah J. Huk. Kel.*, vol. 7, no. 1, hal. 165, Jun 2024, doi: 10.22373/ujhk.v7i1.23054.

[25] W. Wadiyo, S. Haryono, dan J. Wiyoso, "The Phenomenon of Dolanan Song in the Life of Modern Javanese ChildrenRunning head: Dolanan Song for Modern Javanese Children," in *Proceedings of International Conference on Science, Education, and Technology*, 2021, vol. 7, hal. 640–644. [Daring]. Tersedia pada: https://proceedings.unnes.ac.id/ISET/article/view/2025

[26] S. Supeni, O. Handini, dan L. Al Hakim, "The effort to Develop Character Education Child-Friendly School Based on Java Regional Art Culture at Elementary School in Surakarta City," in *Universitas Lampung International Conference on Social Sciences (ULICoSS 2021)*, 2022, hal. 697–703. doi: 10.2991/assehr.k.220102.096.

[27] I. T. C. (ITC), "The International Test Commission Guidelines on the Security of Tests, Examinations, and Other Assessments," *Int. J. Test.*, vol. 16, no. 3, hal. 181–204, Jul 2016, doi: 10.1080/15305058.2015.1111221.

[28] J. Gregoire, "ITC Guidelines for Translating and Adapting Tests (Second Edition)," *Int. J. Test.*, vol. 18, no. 2, hal. 101–134, Apr 2018, doi: 10.1080/15305058.2017.1398166.

[29] K. K. Palmer, D. F. Stodden, D. A. Ulrich, dan L. E. Robinson, "Using Process- and Product-oriented Measures to Evaluate Changes in Motor Skills across an Intervention," *Meas. Phys. Educ. Exerc. Sci.*, vol. 25, no. 3, hal. 273–282, Jul 2021, doi: 10.1080/1091367X.2021.1876069.

[30] K. K. Palmer, A. Pennell, B. Terlizzi, M. A. Nunu, D. F. Stodden, dan L. E. Robinson, "Performance Metrics From Product-Oriented Measures of Fundamental Motor Skills—A Comparison and Developmental Perspective," *J. Mot. Learn. Dev.*, vol. 11, no. 3, hal. 401–423, Des 2023, doi: 10.1123/jmld.2022-0074.

[31] P. Sutapa, K. W. Pratama, M. M. Rosly, S. K. S. Ali, dan M. Karakauki, "Improving Motor Skills in Early Childhood through Goal-Oriented Play Activity," *Children*, vol. 8, no. 11, hal. 994, Nov 2021, doi: 10.3390/children8110994.

[32] U. Faizah, D. Zuchdi, dan Y. Alsamiri, "An authentic assessment model to assess kindergarten students' character," *REID (Research Eval. Educ.*, vol. 5, no. 2, hal. 103–119, Des 2019, doi: 10.21831/reid.v5i2.24588.

[33] M. Laure dan K. Habe, "Stimulating the Development of Rhythmic Abilities in Preschool Children in Montessori Kindergartens with Music-Movement Activities: A Quasi-Experimental Study," *Early Child. Educ. J.*, vol. 52, no. 3, hal. 563–574, Mar 2024, doi: 10.1007/s10643-023-01459-x.

[34] H. Retnawati, "Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index)," *REID (Research Eval. Educ.*, vol. 2, no. 2, hal. 155–164, Des 2016, doi: 10.21831/reid.v2i2.11029.

[35] S. Chun *et al.*, "Comprehensive Assessment and Early Prediction of Gross Motor Performance in Toddlers With Graph Convolutional Networks–Based Deep Learning: Development and Validation Study," *JMIR Form. Res.*, vol. 8, hal. e51996, Feb 2024, doi: 10.2196/51996.

[36] M. Nursalim, *Belajar Mudah & Praktis Analisis Data dengan SPSS dan JASP*. 2022. [Daring]. Tersedia pada: https://repository.radenintan.ac.id/25436/

[37] C. B. Monti, F. Ambrogi, dan F. Sardanelli, "Sample size calculation for data reliability and diagnostic performance: a go-to review," *Eur. Radiol. Exp.*, vol. 8, no. 1, hal. 79, Jul 2024, doi: 10.1186/s41747-024-00474-w.

[38] B. Klingberg, N. Schranz, L. M. Barnett, V. Booth, dan K. Ferrar, "The feasibility of fundamental movement skill assessments for pre-school aged children," *J. Sports Sci.*, vol. 37, no. 4, hal. 378–386, Feb 2019, doi: 10.1080/02640414.2018.1504603.

[39] S. S. Hukkelberg dan K. Taraldsen, "The Strengths and Difficulties Questionnaire Among Youth with Serious Problem Behaviors: Validity, Reliability, and Change during Multisystemic Therapy," *J. Psychopathol. Behav. Assess.*, vol. 48, no. 4, hal. 70, Des 2026, doi: 10.1007/s10862-026-10331-2.

[40] N. S. Utami, J. Nurkamto, M. F. Hidayatullah, dan H. Yuliarto, "Measuring the Content Validity of Physical Literacy Assessment Instruments for Manipulative Games through Aiken Index Analysis," *KnE Soc. Sci.*, vol. 10, no. 9, hal. 163–174, Apr 2025, doi: 10.18502/kss.v10i9.18483.

[41] A. Setiawan, W. Cendana, M. Ayres, A. A. Yuldashev, dan S. P. Setyawati, "Development and validation of a self-assessment-based instrument to measure elementary school students' attitudes in online learning," *REID (Research Eval. Educ.*, vol. 9, no. 2, hal. 184–197, Des 2023, doi: 10.21831/reid.v9i2.52083.

[42] D. Liljequist, B. Elfving, dan K. Skavberg Roaldsen, "Intraclass correlation – A discussion and demonstration of basic features," *PLoS One*, vol. 14, no. 7, hal. e0219854, Jul 2019, doi: 10.1371/journal.pone.0219854.

[43] N. C. Valentini, G. C. Nobre, L. W. Zanella, K. G. Pereira, M. R. Albuquerque, dan M. E. Rudisill, "Test of Gross Motor Development–3 Validity and Reliability: A Screening Form," *J. Mot. Learn. Dev.*, vol. 9, no. 3, hal. 438–455, Des 2021, doi: 10.1123/jmld.2020-0061.

[44] J. S. Y. Tan dan C. B. S. Lim, "The development of gross motor skills in children: Insights from the Motor Assessment Test for Children," *J. Exp. Child Psychol.*, vol. 256, hal. 106255, Agu 2025, doi: 10.1016/j.jecp.2025.106255.