Examining Language Development Quality in Early Childhood Education

A Systematic Review of Assessment and Monitoring Frameworks in Indonesia

Mengkaji Kualitas Perkembangan Bahasa di PAUD

Tinjauan Sistematis atas Kerangka Asesmen dan Pemantauan di Indonesia

Sinta Maulida Hapsari[1], Nindya Alifia Tittandi[2], Octavian Dwi Tanto[3]

[1, 2, 3] Universitas Jember, Indonesia

Corresponding Author
Sinta Maulida Hapsari
sintamausari@unej.ac.id

Abstract

The quality of language development in early childhood education (PAUD) plays a crucial role in preparing children for primary education and supporting their overall cognitive growth. However, despite the existence of various assessment practices across Indonesia, inconsistencies remain in the use of tools, monitoring frameworks, and their practical application within PAUD institutions. This study aims to critically analyze the types of assessment instruments and monitoring frameworks used to evaluate language development in Indonesian PAUD contexts through a systematic review. To achieve this objective, the study applied a systematic review method guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, ensuring a structured and transparent process in identifying, selecting, and synthesizing relevant studies. A total of 20 empirical studies and literature reviews published between 2019 and 2024 were selected from databases including SINTA, GARUDA, DOAJ, Google Scholar and Scopus. The findings reveal that observational checklists, anecdotal records, and portfolio assessments are the most commonly used methods, while the use of digital tools—such as e-portfolios and language development applications—remains limited and underutilized. Most monitoring frameworks are aligned with the 2013 National Curriculum and Ministerial Regulation No. 137 of 2014, although implementation gaps persist due to variations in teacher competence, institutional capacity, and technological infrastructure. This review underscores the urgent need for a standardized, culturally relevant, and technology-integrated system for assessing language development in early childhood education. Enhancing teacher assessment literacy, promoting the integration of digital solutions, and improving policy coherence are recommended to advance the effectiveness and equity of language monitoring practices in Indonesia's PAUD sector.

Keywords

Language Development; Early Childhood Education; Assessment Tools; Monitoring Frameworks

Abstrak

Kualitas perkembangan bahasa dalam pendidikan anak usia dini (PAUD) memegang peranan penting dalam mempersiapkan anak menuju pendidikan dasar serta mendukung pertumbuhan kognitif secara menyeluruh. Namun, meskipun berbagai praktik asesmen telah diterapkan di Indonesia, masih terdapat ketidakkonsistenan dalam penggunaan instrumen, kerangka pemantauan, serta penerapannya di berbagai lembaga PAUD. Penelitian ini bertujuan untuk menganalisis secara kritis jenis-jenis instrumen asesmen dan kerangka pemantauan yang digunakan untuk mengevaluasi perkembangan bahasa dalam konteks PAUD di Indonesia melalui kajian sistematis. Untuk mencapai tujuan tersebut, penelitian ini menerapkan metode systematic review yang mengacu pada protokol PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), guna memastikan proses yang terstruktur dan transparan dalam mengidentifikasi, memilih, dan mensintesis studi yang relevan. Sebanyak 20 studi empiris dan tinjauan literatur yang diterbitkan antara tahun 2019 hingga 2024 dipilih dari basis data seperti SINTA, GARUDA, DOAJ, Google Scholar dan Scopus. Temuan menunjukkan bahwa checklist observasi, catatan anekdot, dan asesmen portofolio merupakan metode yang paling banyak digunakan, sementara penggunaan alat digital—seperti e-portofolio dan aplikasi pengembangan bahasa—masih terbatas dan belum optimal. Sebagian besar kerangka pemantauan mengacu pada Kurikulum Nasional 2013 dan Peraturan Menteri Pendidikan dan Kebudayaan RI Nomor 137 Tahun 2014, namun pelaksanaannya belum merata akibat perbedaan dalam kompetensi guru, kapasitas lembaga, dan kesiapan teknologi. Kajian ini menyoroti pentingnya pengembangan sistem asesmen perkembangan bahasa yang terstandar, kontekstual secara budaya, dan terintegrasi dengan teknologi. Peningkatan literasi asesmen guru, pemanfaatan solusi digital, serta penguatan koherensi kebijakan direkomendasikan untuk meningkatkan efektivitas dan pemerataan pemantauan perkembangan bahasa dalam pendidikan anak usia dini di Indonesia.

Kata Kunci

Perkembangan Bahasa; Pendidikan Anak Usia Dini; Alat Asesmen; Kerangka Pemantauan

1. Introduction

Language development in early childhood is a fundamental aspect of PAUD (Early Childhood Education) as it plays a vital role in preparing children for formal education and supporting their social interaction skills. In Indonesia, efforts to ensure the quality of language development have been initiated through national policies such as Ministerial Regulation No. 137 of 2014. However, in practice, there is a lack of consistency and clear standardization in the use of assessment instruments and monitoring frameworks for children's language development across PAUD institutions.

Recent studies show that the majority of young children in Indonesia demonstrate age-appropriate language development. A study by Widaningsih and Ramadhena (2024) conducted in Kalangsurya Village, Rengasdengklok Subdistrict, Karawang Regency, found that among 75 children aged 3–6 years, 81.3% had normal language development, while 18.7% showed delays mostly due to limited stimulation in the home environment (Widaningsih & Ramadhena, 2024). This highlights the importance of assessments capable of detecting early delays and providing accurate feedback to both teachers and parents.

One example of assessment instrument development is presented by Liu et al (2024), who designed an early detection tool for language development in children aged 5–6 years. The instrument was validated by assessment experts and tested in two kindergartens in Tasikmalaya City. The results indicated that the tool effectively identified children's language development needs Nonetheless, its implementation remains geographically limited and has not yet been integrated at the national level.

A major challenge in the field lies in the limited teacher training on assessment practices, low levels of assessment literacy, and lack of access to digital technology. Several studies, including Aulia (2024) and Siregar et al. (2023), reveal that most assessments used in PAUD remain traditional such as checklists, anecdotal notes, and manual portfolios, while the use of digital applications and e-portfolios is still very limited.

The language development assessment model implemented in Indonesia generally follows the 2013 PAUD Curriculum and is reinforced by the Standar Tingkat Pencapaian Perkembangan Anak (STPPA). The curriculum promotes authentic, ongoing, and context-based assessment (Kemendikbud, 2014). This is reflected in the use of observation, documentation of children's work, and narrative reporting. However, since the implementation of assessments depends heavily on teacher capacity, not all PAUD institutions are able to integrate assessment into planning and reporting in a systematic manner aligned with the curriculum. On the other hand, the recently piloted Kurikulum Merdeka emphasizes formative assessment and the flexibility of developing individualized child learning profiles, offering new opportunities for a needs-based and technology-supported assessment system (Kemendikbudristek, 2022).

Although previous studies have made important contributions in describing assessment tools, no systematic review has yet examined the effectiveness and consistency of various monitoring frameworks for children's language development in Indonesia. This study aims to fill that gap by providing a comprehensive review of the tools and assessment strategies employed and evaluating how well these practices align with the national curriculum and developmentally appropriate principles of early childhood education.

In the results and discussion section, this article presents findings from a systematic review that highlight: (1) the dominant types of language development assessment tools used in Indonesian PAUD settings; (2) the monitoring frameworks and curriculum references underlying their implementation; (3) challenges and gaps in assessment practice; and (4) strategies for strengthening evidence-based and technology-supported assessment systems to ensure the quality of children's language development.

2. Methods

This study employed a Systematic Literature Review (SLR) approach, following the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to ensure transparency, replicability, and

research rigor (Page et al., 2021). The detailed methods used in this study are as follows:

2.1. Rationale for Method Selection

The SLR method was selected due to its relevance in synthesizing existing studies on language development assessment and monitoring in early childhood education (PAUD) in Indonesia. This method enables researchers to map current trends, challenges, and gaps in assessment practices, and to establish a strong foundation for policy recommendations (Tricco et al., 2018).

2.2. Inclusion and Exclusion Criteria

The inclusion criteria for this review were: (1) studies conducted in Indonesia; (2) focus on children aged 0–6 years; (3) discussion of assessment and/or monitoring of language development; (4) publication date between 2019 and 2024; and (5) publication in peer-reviewed journals, scientific proceedings, or institutional research reports. The exclusion criteria included non-scientific works such as opinion articles, essays, or studies lacking an explicit methodology.

2.3. Data Sources and Collection Techniques

Data were collected from five academic databases: SINTA, GARUDA, DOAJ, Google Scholar, and Scopus. The search used both Indonesian and English keywords, including: "asesmen perkembangan bahasa PAUD", "language development assessment in early childhood", and "monitoring framework Indonesia". The search was conducted between January and March 2025.

2.4. Selection and Validation Procedures

Article selection followed the four stages of PRISMA: identification, screening (titles and abstracts), eligibility (full-text review), and inclusion. Validation was conducted independently by two reviewers. In cases of disagreement, the final decision was reached through joint discussion and consensus.

2.5. Data Analysis Technique

The data were analyzed using thematic synthesis, grouping findings into key themes such as types of assessment tools, monitoring frameworks, implementation effectiveness, field-level challenges, and

policy implications. This approach follows the analytical guidelines of Braun and Clarke (2021), which emphasize in-depth exploration of patterns and meanings in qualitative research.

2.6. Informants/Respondents

This study did not utilize primary data and did not involve any informants or respondents. All information was derived from previously published research documents.

2.7. Research Setting and Time Frame

The review covered studies conducted across various regions in Indonesia, making the research scope national in nature. The review process was conducted over a period of three months, from January to March 2024.

3. Results and Discussion

3.1. Result

Based on the results of the literature review and screening process, 20 eligible articles were selected and analyzed. The findings presented in Table 1 demonstrate the diversity of tools and frameworks used to assess and monitor language development in early childhood education institutions in Indonesia. The discussion below elaborates on the key themes derived from the analysis, corresponding to the research questions posed in the introduction.

3.2. Discussion

The systematic review of 20 selected studies reveals that the dominant form of language development assessment in Indonesian PAUD institutions remains traditional and observational. Tools such as anecdotal records, developmental checklists, work samples, performance assessments, and portfolios are used in over 90% of the articles analyzed (Siregar et al., 2023; Nurwahyuni & Mahyuddin, 2021). This pattern reflects a prevailing cultural and pedagogical preference for informal, authentic assessments that align closely with the play-based and holistic learning models commonly practiced in PAUD settings.

Table 1. Respondents' Preferences on AI in ELL

No	Author(s) & Year	Title (Simplified)	Type of Assessment Tool	Sample / Age Group	Main Findings	Notes / Limitations
1	Aulia (2024)	Asesmen Bahasa di TK Pertiwi Duwet	Checklist, Anecdotal Notes	5–6 years; 1 school	Conducted regularly, but not used to adjust learning strategies	Weak linkage to instruction; no data analysis
2	Hulme et al. (2015)	Early Language Screening	LanguageScreen test consisting of 77 dichotomously scored items across four subscales.	Children aged 3-8 years.	The screening tool is effective in identifying language difficulties and can derive profiles across different language subscales for individual children.	Effectiveness may vary based on age and demographic factors.
3	Dockrell & Marshall (2015)	Assessing Language Skills in Young Children	Various assessment tools including observational methods and standardized tests.	Preschool children aged 3-5 years.	The study emphasizes the importance of using a variety of assessment methods to accurately gauge language skills in young children, highlighting the role of both qualitative and quantitative measures.	The effectiveness of assessments can vary based on individual child characteristics and environmental factors.
4	Adistia (2022)	Asesmen perkembangan bahasa anak usia dini 4–5 tahun di TA Az-Zahrah Rifaiyah Bojong.	Qualitative assessment tools including observation sheets, reading portfolios, and rubrics for speaking and reading skills.	Children aged 4-5 years at TA Az-Zahrah Rifaiyah Bojong (20 children).	The assessment indicates that language development is progressing well based on the techniques used by teachers, highlighting the importance of gradual assessments over time to monitor children's language growth after learning activities.	The study focuses on a specific age group and setting; results may not be generalizable to other contexts or populations.
5	Finders et al. (2023)	Early childhood education language environments: Considerations for research and practice.	Observational and environmental language assessment tools, including LENA (Language ENvironment Analysis) and teacher-child interaction analyses.	Children aged 3–5 years in early childhood education (ECE) settings. The study also observed and analyzed teacher behaviors and classroom environments.	Standardized assessments may not capture authentic language use. Emphasis on responsive, meaningful conversations improves outcomes.	Limited generalizability due to sample size and context (specific classrooms). Suggests more culturally responsive tools are needed for diverse populations.
6	Conti-Ramsden & Durkin (2012)	Language Development and Assessment in the Preschool Period	Various standardized tests for assessing preschool children's language abilities including vocabulary and grammar measures.	Preschool children aged under 5 years old from diverse backgrounds.	Highlights that comprehensive assessment requires a multi-method approach involving both formal tools and informal observations to accurately evaluate a child's language skills; stresses early identification for intervention purposes can improve long-term outcomes for children with language impairments.	Variability in early language acquisition makes it challenging to identify difficulties accurately; no universal policy exists for screening young children.
7	Simanjuntak et al. (2019)	Asesmen Formatif Perkembangan Bahasa Anak	Qualitative assessment tools including interviews, observations, checklists, and portfolios.	Children aged 4-5 years at various early childhood education settings.	The research emphasizes the importance of formative assessments in monitoring language development through various effective techniques that align with children's learning needs.	Limited to specific age groups; findings may vary based on different educational environments.
8	Karem et al. (2019)	A Scoping Review Using the International Classification of Functioning, Disability and Health–Children and Youth Version	Literature-based review using ICF-CY framework	325 peer-reviewed studies	Most tools focus on semantics and activity; few assess participation in social context.	Conceptual review; lacks empirical validation of tool effectiveness.
9	Khairunnisa & Mahyuddin (2022)	Keterbatasan Asesmen Digital di PAUD	E-portfolio (Pilot)	5–6 years; 1 school	Teachers interested but constrained by digital literacy	No institutional support; lack of training

10	Knauer et al. (2019)	Multilingual language assessment in Kenyan children	Receptive & expressive vocabulary tests in 3 languages	505 children aged 2–6 years (Kenya)	Mother tongue vocabulary (Luo) strongly predicts English development; parental literacy positively influences outcomes.	Focus on multilingual setting; cultural adaptation required for replication.
11	Liu et al. (2024)	The use of digital technologies to develop young children's language and literacy skills: A systematic review.	Systematic review of empirical, peer-reviewed journal articles published over the past two decades.	Involving young children in early childhood settings, not children directly.	The study concludes that language development assessments can be effectively conducted using a variety of methods tailored to the needs of children and educators; it emphasizes the importance of continuous monitoring through formative assessments.	The findings are based on literature review from multiple sources; results may vary depending on specific educational contexts.
12	Maghfiroh et al. (2024)	Permasalahan Perkembangan Bahasa pada PAUD	Literature review: the article analyzes existing literature related to language development problems in early childhood.	No primary subjects: the study is based on previous journal articles, and research related to children's language difficulties.	Untreated, they may impact reading, social behavior, and academics. Early detection using various sources is crucial for accurate diagnosis and timely intervention.	Findings are general and not contextualized to specific age groups or local PAUD settings.
13	Siregar et al. (2023)	Asesmen Perkembangan Bahasa Pendidikan Anak Usia Dini	Various assessment methods including observation, anecdotal records, checklists, portfolios, and interviews.	Children aged 4-6 years in early childhood education settings.	The study concludes that language development assessments can be effectively conducted using a variety of methods tailored to the needs of children and educators; it emphasizes the importance of continuous monitoring through formative assessments.	The findings are based on literature review from multiple sources; results may vary depending on specific educational contexts.
14	Nurwahyuni & Mahyuddin (2021)	Penilaian perkembangan bahasa anak usia dini umur 5–6 tahun pada masa new normal di Taman Kanak-Kanak Ridhotullah Padang	Anecdotal, Rubrics	4–6 years; multiple sites	Low assessment literacy; tools underused	Limited training support
15	Oguz et al. (2019)	Comparison of language disorder assessments: Turkey vs USA	Observational tools, toys, and books	184 speech-language therapists in Turkey & USA	U.S. therapists use book-based assessments; Turkish therapists use toy-based tools in rehabilitation centers.	Limited to clinical settings; not specific to early childhood education institutions.
16	Salehuddin & Mahmud (2024)	Assessing Children's Language Development: A Systematic Literature Review on Early Language Milestone Scales.	Standardized Reading Readiness Tool	Reviewed studies comprised 19 language milestone scales used in early childhood	ELM scales are validated, reliable tools for assessing receptive and expressive language in young children.	Language and geographic bias: Most studies focus on English and Western contexts—other languages underrepresented.
17	Weadman et al. (2021)	The development and psychometric properties of a shared book reading observational tool: The Emergent Literacy and Language Early Childhood Checklist for Teachers (ELLECCT)	Observational Assessment Tool (ELLECCT)	12 professionals across early childhood education and speech-language pathology.	Effectively captures teacher use of extratextual dialogue prompts, vocabulary strategies, responsiveness, print awareness, phonological elements, and paralinguistic engagement during shared reading.	Developed within Australian early childhood contexts generalizability to other cultural or educational settings is limited.
18	Widaningsih & Ramadhena (2024)	Pemantauan Bahasa di Karawang	STPPA-aligned Checklist	75 children; 3–6 years	81.3% normal, 18.7% delayed; rarely followed up for pedagogical action	No interpretation model applied

19	Gayatri et al. (2020)	Expanding Language Development in Preschool Age Children	Informal assessments including observations through checklists and anecdotal notes.	Preschool children aged 4-6 years.	Highlights the importance of formal and informal assessments and teachers' role in tracking language development.	Focused on specific educational settings; results may not be generalizable.
20	Aulia (2024)	Assessment of Early Childhood Language Development at TK Pertiwi Duwet	Anecdotal notes, checklists, photo series.	Children aged 4-5 years at TK Pertiwi Duwet.	Outlines language assessment techniques that help educators tailor stimuli and interventions for children's development.	Focused on a specific educational setting; results may not be generalizable to other contexts.

Source: *Compiled by the author (2025).*

3.2.1. Types of Language Development Assessment Tools in PAUD

Teachers commonly use checklists adapted from the STPPA, marking children's progress in receptive and expressive language skills across developmental stages. Anecdotal records typically written during free play or structured activities help document spontaneous language use, vocabulary growth, and comprehension indicators. Portfolios are employed to gather artifacts such as drawings with labels, dictated stories, and language related group work products. According to Siregar et al. (2023), these tools are valued for their adaptability and ease of integration into everyday pedagogical routines. However, the review also reveals significant limitations in practice.

Twenty studies indicated that teachers often treat assessment merely as a formality, focusing on filling checklists for reporting rather than using them to inform instruction (Aulia, 2024; Widyastuti, 2020). The depth of observation is often limited, and the interpretation of findings is rarely documented systematically.

The studies consistently emphasize that most teachers do not conduct follow-up assessments or use the data to adjust lesson plans, thus limiting the formative value of these practices.

In terms of innovation, only two articles (less than 10%) report the use of digital-based tools for assessing language development. Liu et al (2024) introduced a prototype of a digital observation app for language development in children aged 5–6, validated by expert reviewers and tested in two kindergartens. The tool showed potential in capturing more structured and time-stamped data through image, video, and score-based observation formats. However, limitations in digital literacy among teachers, infrastructure constraints, and lack of institutional support hinder broader adoption.

A promising development has emerged in the form of a reading readiness assessment instrument tailored for Indonesian early childhood learners. With a reliability index of 0.88, the tool demonstrated high potential for early detection of phonological and decoding challenges, suggesting that more targeted tools could complement the current observational methods.

In summary, while traditional assessment tools are widely used and culturally accepted in PAUD, they often fall short in terms of standardization, interpretative depth, and instructional integration. Digital innovation is still emerging, though promising. The findings highlight the urgency of enhancing teachers' assessment literacy and investing in both low-tech and high-tech assessment instruments that are accessible, valid, and pedagogically meaningful.

3.2.2. Monitoring Frameworks and Curriculum Integration

Indonesia's early childhood education policy has mandated structured monitoring of child development through national regulations such as Ministerial Regulation No. 137 of 2014 and the 2013 PAUD Curriculum. These frameworks introduce the STPPA as a reference for monitoring six developmental domains, with language both receptive and expressive playing a central role. Nearly all reviewed studies confirm that PAUD teachers refer to the STPPA in their assessment practices (Siregar et al., 2023; Nurwahyuni & Mahyuddin, 2021).

Despite its widespread use, implementation of STPPA based assessment is inconsistent and often superficial. A key finding from the review is that many teachers utilize the indicators merely as checklists for administrative

Table 2. Comparison Between Traditional and Digital Assessment Practices in PAUD

No.	Aspect	Traditional Assessment	Digital Assessment
1	Tool Used	Checklists, Anecdotal Notes, Portfolios	Mobile Apps, E-Portfolios, Digital Rubrics
2	Dominance in Practice	Very high (used in >90% of reviewed studies)	Very low (<10% implementation across PAUD)
3	Ease of Implementation	High – requires no tech	Low – requires infrastructure & training
4	Pedagogical Integration	Limited – often procedural	Potentially high – if integrated with planning
5	Documentation Quality	Basic – mostly qualitative	Rich – photos, videos, timestamps
6	Training Requirements	Moderate – assumed knowledge	High – most teachers need digital literacy
7	Limitations	Often lacks follow-up analysis	Not yet scalable; dependent on school resources
8	Potential for Scalability	High – easy to replicate	Moderate – needs systemic support

Source: *Compiled by the author based on reviewed literature (Aulia, 2024; Siregar et al., 2023; Karem et al, 2019; Khairunnisa & Mahyuddin, 2022).*

reporting rather than as dynamic tools to monitor and respond to children's developmental trajectories. For instance, Widaningsih and Ramadhena (2024) documented that in a sample of 75 children, while teachers regularly recorded developmental milestones, they rarely engaged in further reflection or curricular adjustment based on those records.

The 2013 PAUD Curriculum emphasizes holistic, child-centered learning and stresses the importance of continuous and contextual assessment. Yet, research shows that this ideal is not fully realized in practice. Karem et al. (2019) and Khairunnisa (2022) both observed that teachers often do not see the relevance of language assessment in shaping lesson design or learning outcomes. Assessment is treated more as a compliance requirement than a tool for pedagogical reflection.

In contrast, the Kurikulum Merdeka pilot trials offer a more flexible framework by encouraging teachers to develop individualized child learning profiles. This approach supports narrative-based monitoring and reflective practice, with potential for richer documentation of children's language progress. However, the review found that only a handful of institutions had adopted this model, and even fewer had integrated it meaningfully into daily planning and parent communication (Widaningsih & Ramadhena, 2024).

A significant issue highlighted in several studies is the disconnect between national policy and classroom application. While the STPPA provides generalized indicators, it does not offer structured instruments or training modules to guide their interpretation and use. As a result, teachers often apply their own discretion or rely on senior colleagues for guidance practices that are vulnerable to subjective bias and inconsistencies across institutions (Aulia, 2024).

Moreover, integration between assessment and curriculum planning remains weak. Less than 30% of reviewed studies report that language development monitoring data are used to inform short-term or long-term instructional decisions. In most cases, assessments are archived or submitted for supervisor review without influencing classroom strategies. This lack of feedback loops significantly reduces the impact of monitoring on learning outcomes.

In conclusion, while Indonesia has developed robust policy frameworks for language development monitoring in PAUD, the implementation is hindered by limited teacher capacity, insufficient instructional linkage, and a lack of practical tools. Bridging the gap between curriculum goals and classroom practices will require intensive capacity building, supportive leadership, and the co-construction of monitoring tools that are not only standardized but also practical and adaptable.

3.2.3. Implementation Challenges

Despite the formal inclusion of language development assessment in national early childhood policies, the implementation in PAUD institutions across Indonesia faces numerous systemic and practical challenges. This review identifies five interrelated obstacles: limited teacher competence, insufficient access to standardized tools, unequal technological infrastructure, administrative overload, and policy–practice disconnect.

The first and most frequently reported challenge is the low assessment literacy among PAUD teachers. In 18 of the 20 reviewed studies, teachers admitted difficulty in interpreting developmental indicators, differentiating between receptive and expressive language skills, or connecting assessments to instructional strategies. Siregar et al. (2023) and Nurwahyuni & Mahyuddin (2021) both reported that most teachers view assessment as a task to be completed rather than a process to support learning. Consequently, data collected are often superficial, under-analyzed, and not translated into pedagogical actions.

Second, the lack of contextually adapted and validated instruments presents another barrier. Although the STPPA provides general developmental milestones, it does not offer operational tools or criteria for measuring specific aspects of language development. This gap forces many institutions to rely on internally developed checklists or copy instruments from other contexts, which may not be culturally or linguistically relevant. As shown by Aulia (2024), the instruments used in TK Pertiwi Duwet were not validated and lacked clear scoring guidelines, resulting in inconsistent interpretations among teachers.

Third, digital assessment tools remain largely inaccessible or underutilized, especially in rural or lower-resourced settings. Only 2 out of 20 reviewed studies reported the use of digital media such as e-portfolios or apps for monitoring language skills (Siregar, 2023; Khairunnisa, 2022). Even where hardware is available, usage is hampered by lack of teacher training, unclear assessment models, and concerns about data privacy and system complexity. Teachers in these contexts often prefer paper-based documentation, which may be more familiar but less efficient and less rich in capturing language development progress.

Fourth, the administrative burden associated with early childhood assessment is a persistent complaint. Teachers are expected to document each child's progress weekly or even daily, often across multiple domains. Yet, there are few mechanisms in place to synthesize or utilize this information in a meaningful way. Alfatikah (2022) observed that in classrooms with more than 20 children, teachers spend significant time on documentation rather than interaction, thus undermining the very development they aim to support.

Lastly, the gap between policy expectations and classroom realities reduces the effectiveness of language assessment initiatives. Supervisory visits from education officials tend to emphasize completeness of documentation rather than its instructional value. Moreover, in many schools, there is no follow-up support such as mentoring, collaborative reflection, or data-driven planning sessions. This lack of institutional infrastructure results in assessment becoming performative rather than transformative.

In summary, these challenges point to the urgent need for a systemic response one that includes national-level standardization of instruments, continuous professional development, digital equity investments, and reform in supervision and reporting structures. Without addressing these foundational issues, the potential of language assessment to drive quality learning in PAUD will remain largely unrealized.

3.2.4. Synthesis and Implications

The systematic analysis of 20 articles on language development assessment in PAUD reveals a critical paradox: although national frameworks and assessment instruments are established, their fragmented and inconsistent implementation substantially undermines their intended pedagogical impact. This section synthesizes key findings from the review and critically discusses their implications for early childhood education policy, practice, and system development in Indonesia.

The review affirms that traditional assessment tools, such as checklists, anecdotal notes, and portfolios continue to dominate the landscape. These tools align
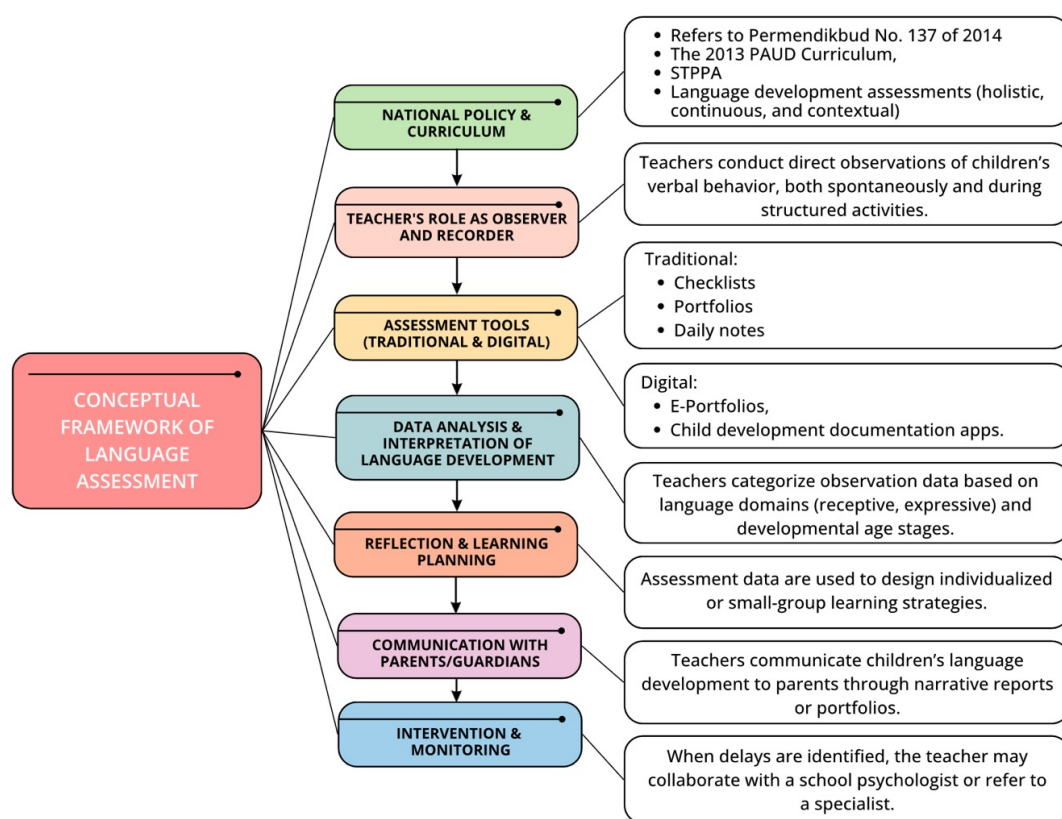
Figure 1. Conceptual Flow of Language Assessment Implementation in PAUD (Source: Adapted and interpreted by the author).

with the philosophy of the 2013 PAUD Curriculum, which promotes contextual, authentic, and holistic learning. However, their pedagogical potential remains underutilized. As revealed by Siregar et al. (2023), most teachers use checklists simply as documentation for compliance, not as a diagnostic tool to inform instruction. This points to a disconnect between the intended formative nature of assessment and its application in classrooms.

The current situation also reflects the absence of standardized, validated assessment instruments specifically designed for the Indonesian PAUD context. While the STPPA provides important developmental benchmarks, it does not supply practical tools that guide daily classroom assessment. This leaves educators with the responsibility to create or adapt their own instruments, often without proper validation or training. The result is a landscape marked by heterogeneity and lack of reliability, which hinders data-driven decision-making at the institutional and system levels.

Moreover, the integration of assessment outcomes into instructional planning is minimal. Less than one-third of the studies reviewed reported that language development data influenced curricular adjustments. Teachers rarely conduct analysis or reflection on the collected data, and when they do, it is not systematically documented or linked to individualized learning trajectories (Aulia, 2024; Widaningsih & Ramadhena, 2024). This indicates that the assessment cycle remains incomplete and assessment is not fully functioning as a tool for improving teaching and learning.

The limited use of digital assessment systems is another missed opportunity. Despite positive trials in several urban settings using mobile apps and e-portfolios (Tan et al., 2023), the adoption of such innovations remains rare, primarily due to unequal infrastructure and a lack of teacher digital literacy. A national-level push to develop open-access, culturally grounded digital assessment platforms could significantly enhance the accessibility, efficiency, and consistency of language assessment across Indonesia's diverse educational

settings. Importantly, the review reveals that teacher capacity and institutional support are the linchpins of effective assessment. Policies alone are insufficient; they must be accompanied by structured professional development, mentoring, and supervisory support that shifts the focus from compliance to pedagogy. Without this, assessment is likely to remain a compliance task rather than a catalyst for pedagogical improvement.

In sum, the implications of this review suggest that improving the quality of language development assessment in PAUD requires a systemic, multi-layered response. This includes:

a) Developing standardized yet flexible assessment tools validated for diverse PAUD settings;

b) Embedding assessment training into pre- and in-service teacher development programs;

c) Promoting the integration of assessment data into planning and parent communication;

d) Facilitating the digital transformation of monitoring systems;

e) Realigning supervision systems to support reflective practice and continuous improvement.

By addressing these areas, Indonesia can move toward an equitable and evidence-based early childhood system, where language development assessment truly contributes to the holistic growth of every child.

3.2.5. Conceptual Flow of Language Assessment Implementation in PAUD

A conceptual flow model is developed in this section to illustrate the ideal sequence of language assessment practices in PAUD, grounded in both empirical findings and national curriculum standards. This flow outlines the ideal cycle of assessment practice, as informed by empirical evidence from the reviewed studies and aligned with national educational frameworks in Indonesia. It emphasizes the dynamic interplay between policy, teacher agency, assessment tools, and pedagogical decision-making.

The process begins with the foundation of national policy and curriculum, particularly the 2013 PAUD Curriculum and the STPPA, as stipulated in Ministerial Regulation No. 137 of 2014. These regulatory frameworks mandate that assessments must be holistic encompassing cognitive, linguistic, and socio-emotional domains continuous in nature, and contextualized to the lived experiences of the child. In this framework, teachers are expected to use developmental indicators, especially for language competencies, as guiding references in their daily classroom practices (Siregar et al., 2023).

Building upon this policy base, the teacher's role as an observer and recorder becomes central to assessment enactment. Teachers are positioned not merely as transmitters of knowledge, but as reflective practitioners who systematically observe and document children's language behavior both spontaneous and elicited across a range of learning scenarios. This includes tracking expressive and receptive abilities such as vocabulary acquisition, sentence formation, comprehension, and participation in verbal interactions. Observations are typically recorded using anecdotal notes and structured checklists, which serve as primary data sources (Nurwahyuni & Mahyuddin, 2021).

The selection and application of assessment tools constitute the next stage in the flow. In most PAUD institutions, traditional tools such as portfolios, developmental checklists, and narrative documentation continue to be prevalent due to their accessibility and familiarity. However, in technologically equipped environments, teachers are beginning to adopt digital instruments, such as e-portfolios and mobile-based child observation apps which offer improved efficiency, richer data formats (e.g., photo, video), and opportunities for real-time documentation (Khairunnisa & Mahyuddin, 2022; Simanjuntak, 2019). These tools are particularly promising for tracking complex, process-based language development phenomena.

Following data collection, analysis and interpretation are conducted. Here, teachers categorize documented observations based on core domains of language development typically distinguishing between receptive (e.g., listening comprehension) and expressive (e.g., verbal output) abilities—and compare them to age-appropriate benchmarks outlined in national standards. This interpretative process enables teachers to identify both strengths and developmental delays, thereby laying the groundwork for responsive instructional strategies (Widaningsih & Ramadhena, 2024).

The outcomes of assessment then inform the reflection and planning of instruction. Teachers utilize the interpreted data to design learning experiences that are tailored to individual or group needs. For instance, a child with limited vocabulary expression might be supported through targeted games, rich storytelling experiences, or structured peer interactions that stimulate verbal engagement (Aulia, 2024). This stage embodies the pedagogical function of assessment as a feedback mechanism to enhance learning.

An essential, often under-emphasized component in the flow is communication with parents or guardians. Teachers are encouraged to share children's language development progress through formal channels such as portfolio reviews, narrative progress reports, and parent-teacher meetings. This collaborative communication not only fosters home-school alignment but also empowers families to support language-rich environments at home (Siregar et al., 2023).

Finally, the flow culminates in intervention and follow-up monitoring. When assessment reveals significant delays or atypical development, teachers may engage in additional targeted observation, consult with specialists, or initiate referrals to professional services. Continued monitoring ensures that interventions are responsive and appropriately adjusted based on the child's evolving needs (Asthiningsih, 2021).

This conceptual flow model underscores the dual role of assessment as both a diagnostic and pedagogical tool. It positions assessment not as an isolated activity, but as an integral component of a cyclical, evidence-informed instructional process. While this model is theoretically supported by policy, its practical realization demands ongoing teacher professional development, access to reliable assessment tools, supportive institutional structures, and sustained investment in capacity building across the PAUD ecosystem. By addressing these areas, Indonesia can move toward an equitable and evidence-based early childhood system, where language development assessment truly contributes to the holistic growth of every child.

4. Conclusion

This systematic review provides a comprehensive synthesis of the current practices, challenges, and opportunities in assessing language development within Indonesia's early childhood education (PAUD) sector. The evidence confirms that, although national policies such as the 2013 Curriculum and the STPPA monitoring framework provide normative guidance, there remains a notable discrepancy between policy expectations and actual classroom practices.

Firstly, the review reveals that assessment tools predominantly consist of traditional instruments, including observational checklists, anecdotal records, and portfolio assessments. While these tools are widely employed across PAUD institutions, their use tends to be limited to fulfilling administrative requirements rather than functioning as formative assessments that actively inform instruction and support individualized learning pathways for children. This highlights a critical pedagogical gap, where assessment data is underutilized in driving meaningful curriculum adaptation.

Secondly, the monitoring frameworks currently implemented align broadly with national curriculum standards and ministerial regulations. However, the review identifies significant variations in the fidelity of framework application, largely influenced by disparities in teacher competence, institutional capacity, and the availability of technological resources. These factors contribute to inconsistent monitoring quality and undermine the potential benefits of structured language development evaluation.

Thirdly, the challenges impeding effective language assessment in PAUD include the absence of standardized, validated assessment instruments tailored to the Indonesian early childhood context, limited teacher training in assessment literacy, and unequal access to digital tools and infrastructure. Although digital assessment approaches such as e-portfolios and language development applications show promise in some urban and well-resourced settings, their adoption remains sporadic and inaccessible in many rural and under-resourced areas.

Finally, the review emphasizes the urgent need for a paradigm shift toward developing culturally contextualized and standardized assessment instruments that accommodate diverse learning environments. Equally important is the implementation of professional development programs aimed at enhancing teacher

capacity in administering, interpreting, and utilizing assessment results effectively. Furthermore, supervisory and monitoring models should prioritize reflective practices and meaningful documentation to bridge the gap between assessment outcomes and instructional planning.

The findings presented in this review directly address the research questions by detailing the dominant assessment tools, the structure and application of monitoring frameworks, the critical challenges faced in practice, and evidence-based strategies for strengthening language assessment systems. This study advocates for further research exploring the scalability and contextual adaptability of digital and hybrid assessment models, as well as the role of collaborative teacher learning communities in embedding assessment as an integral component of early childhood pedagogy.

By responding systematically to these issues, this review contributes to the advancement of equitable, effective, and sustainable language development assessment practices in Indonesia's PAUD sector, ultimately supporting better educational outcomes for young learners.

Acknowledgment

References

Adistia, F. (2024). Asesmen perkembangan bahasa anak usia dini 4–5 tahun di TA Az-Zahrah Rifaiyah Bojong. *SINAU: Seminar Nasional Pendidikan Anak Usia Dini, 1*(1), 178–186. Crossref | Google Scholar

Alfatikah, A. F. (2022). Eksplorasi pelaksanaan asesmen perkembangan anak usia 5–6 tahun di Taman Kanak-Kanak Al-Falah Jakarta (Skripsi, Universitas Islam Negeri Syarif Hidayatullah Jakarta). *Repository UIN Jakarta*. Google Scholar

Asthiningsih, N. W. W. (2021). Deteksi dini perkembangan anak menggunakan instrumen DDST. *Universitas Muhammadiyah Kalimantan Timur*.

Aulia, Z. E. (2024). Asesmen perkembangan bahasa anak usia dini di TK Pertiwi Duwet. *Awladuna: Jurnal Pendidikan Islam Anak Usia Dini, 2*(2), 12–19. Crossref | Google Scholar

Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. SAGE Publications. Google Scholar

Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychology Review, 22*(4), 384–401. Crossref | Google Scholar

Dockrell, J. E., & Marshall, C. R. (2015). Measurement issues: Assessing language skills in young children. *Child and Adolescent Mental Health, 20*(2), 116–125. Crossref | Google Scholar

Finders, J., Wilson, E., & Duncan, R. (2023). Early childhood education language environments: Considerations for research and practice. *Frontiers in Psychology, 14*, 1202819. Crossref | Google Scholar

Gayatri, Y., Dewi, I. P., & Wijayanti, S. (2020). Expanding language development in preschool age children. *International Journal of Early Childhood Education, 6*(2), 110–125. Google Scholar

Gillian West, Margaret J. Snowling, Arne Lervag, Elizabeth Buchanan-Worster, Mihaela Duta, Alexandra Hall, Henrietta McLachlan, Charles Hulme (2021). Early language screening and intervention can be delivered successfully at scale: Evidence from a cluster randomized controlled trial. *Journal of Child Psychology and Psychiatry, 62*(12), 1425–1434. Crossref | Google Scholar

Hulme, C., Nash, H. M., Gooch, D., Lervag, A., & Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychological science, 26*(12), 1877–1886. Crossref | Google Scholar

Karem, R. W., Washington, K. N., Crowe, K., Jenkins, A., Leon, M., Kokotek, L., Raisor-Becker, L., & Westby, C. (2019). Current Methods of Evaluating the Language Abilities of Multilingual Preschoolers: A Scoping Review Using the International Classification of Functioning, Disability and Health–Children and Youth Version. *Language, Speech, and Hearing Services in Schools, 50*(3), 434-451. Crossref | Google Scholar

Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2014). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 137 Tahun 2014 tentang Standar Nasional Pendidikan Anak Usia Dini. Retrieved from https://repositori.kemdikbud.go.id/12860/1/Permendikbud%20No.%20137%20Tahun%202014%20-%20SN-PAUD.pdf.

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2024). Pedoman Implementasi Kurikulum Merdeka pada PAUD Tahun 2024. Retrieved from https://kurikulum.kemdikbud.go.id/file/1711507788_manage_file.pdf.

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2022). Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 16 Tahun 2022 tentang Standar Proses pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah. Retrieved from https://peraturan.bpk.go.id/Details/224238/permendikbudriset-no-16-tahun-2022.

Khairunnisa, & Mahyuddin. (2022). Asesmen Autentik Di Lembaga PAUD pada Era Revolusi Industri 4.0. *Jurnal Family Education, 2*(4), 319-330. Crossref | Google Scholar

Knauer, H., Mammo, A., Ssebo, F., & Ochen, F. (2019). Multilingual language assessment in Kenyan children. *International Journal of Early Childhood, 51*(1), 43–59. Crossref | Google Scholar

Liu, S., Lin, T.-J., Chai, C. S., & Wang, H. (2024). The use of digital technologies to develop young children's language and literacy skills: A systematic review. *SAGE Open, 14*(1). Crossref | Google Scholar

Maghfiroh, L. P., & Sulianto, J. (2024). Permasalahan perkembangan bahasa pada anak PAUD. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang, 10*(2), 258 - 265. Crossref | Google Scholar

Nurwahyuni, E., & Mahyuddin, N. (2021). Penilaian perkembangan bahasa anak usia dini umur 5–6 tahun pada masa new normal di Taman Kanak-Kanak Ridhotullah Padang. *Jurnal Cikal Cendekia, 2*(1), 11–23. Crossref | Google Scholar

Oguz, A., Yavuz, S., & Dogan, M. (2019). Comparison of language disorder assessments: Turkey vs USA. *International Journal of Language & Communication Disorders, 54*(5), 831–845. Google Scholar

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. Crossref | Google Scholar

Salehuddin, K., & Mahmud, F. N. (2024). Assessing Children's Language Development: A Systematic Literature Review on Early Language Milestone Scales. *GEMA Online Journal of Language Studies, 24*(4) 17-46. Crossref | Google Scholar

Simanjuntak, I. A., Akbar, S., & Mudiono, A. (2019). Asesmen Formatif Perkembangan Bahasa Anak. *Jurnal Pendidikan: Teori, Penelitian dan Pengembangan, 4*(8), 113–126. Crossref | Google Scholar

Siregar, M., Indryani, I., Asmara, E. D., Sarniya, A., & Sulistiani, S. (2023). Asesmen Perkembangan Bahasa Pendidikan Anak Usia Dini. *Jurnal Pendidikan Anak Usia Dini, 7*(1), 45–59. Google Scholar

Tan, M., Kilani, H., Markov, I., Hein, S., & Grigorenko, E. L. (2023). Assessing Cognitive Skills in Early Childhood Education Using a Bilingual Early Language Learner Assessment Tool. *Journal of Intelligence, 11*(7), 143-158. Crossref | Google Scholar

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., Lewin, S., Godfrey, C. M., Macdonald, M. T., Langlois, E. V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, O., & Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. Crossref | Google Scholar

Weadman, T., Serry, T., & Snow, P. C. (2021). The development and psychometric properties of a shared book reading observational tool: The Emergent Literacy and Language Early Childhood Checklist for Teachers (ELLECCT). *First Language, 41*(4), 552–578. Crossref | Google Scholar

Widaningsih, C., & Ramadhena, M. P. (2024). Analisis perkembangan bahasa pada anak usia dini di Desa Kalangsurya Kecamatan Rengasdengklok Kabupaten Karawang. *Malahayati Nursing Journal, 6*(3), 123–130. Crossref | Google Scholar

Widyastuti, P., & Pramono, Y. G. H. (2020). An analysis of English teachers' written feedback in primary students' report cards. *LLT Journal: A Journal on Language and Language Teaching, 23*(2), 123–135. Crossref | Google Scholar

.