Development and Validation of an Instrument for Measuring Language Development in 4 to 5 Year Old Children

Sovia Mas Ayu, Asmaru Dewi, Cahnyio Wijaya Kuswanto, Mardiana

1 UIN Raden Intan Lampung, Indonesia, 2 MIS-YYP Kelapa Tiga, Lampung, Indonesia, 3 UIN Raden Intan Lampung, Indonesia, 4 Universitas Muhammadiyah Lampung, Indonesia

Keywords
Language Development, Children Aged 4-5 Years, Early Childhood Education, Instrument Validation

Abstract
This research aimed to develop and validate an instrument to measure language development in children aged 4 to 5 years in kindergarten. The study followed the Plomp development model, encompassing five phases: initial investigation, design, realisation test, evaluation and revision, and implementation. The initial investigation involved analysing theories and existing early childhood language development research. The design phase produced an initial draft of the instrument, which included a language development measurement tool, a validation questionnaire, and a user response questionnaire. Expert validation was conducted using Aiken’s formula to establish content validity. The instrument was then tested in three kindergartens with a sample size of 194 children. Confirmatory Factor Analysis (CFA) and descriptive analysis assessed the model’s fit and instrument reliability. Results indicated that the instrument had satisfactory content validity and construct reliability (CR > 0.7), though some fit indices, such as the Comparative Fit Index (CFI), were below ideal thresholds. The validated instrument consists of two language development scales: receptive language (e.g., understanding stories, commands, rules) and expressive language (e.g., asking and answering questions, verbal communication). The study’s findings imply significant applications for early childhood education, offering a standardised tool for assessing language development, guiding curriculum design, and enabling targeted interventions to improve language outcomes. Despite its strengths, the study acknowledges limitations, such as the need for better-fit indices and further empirical testing with diverse populations. Future research should focus on longitudinal studies to explore the long-term impacts of the instrument and incorporate technological advancements to enhance its accuracy and relevance. Continuous refinement and validation will ensure the instrument’s effectiveness in measuring early childhood language development.

Introduction
The development of language skills in early childhood is a critical foundation for communication and cognitive growth, impacting educational outcomes and social integration (Peterson, 1996; Kurniati, 2017). Early language development facilitates the acquisition of literacy skills essential for academic success and lifelong learning (Neumann et al., 2019; Sansavini et al., 2021). Language proficiency is crucial for children to effectively express their thoughts, emotions, and needs, enhancing their social interactions and emotional well-being (Sarmini et al., 2020). Recognising the importance of early language development is paramount for fostering a society where individuals can thrive academically and socially (Kang et al., 2021; Lim et al., 2020). This underscores the need for comprehensive assessment tools to monitor and support language development in young children.

In developmental psychology and early childhood education, understanding the intricacies of language development in children aged 4 to 5 years is essential for designing effective educational interventions (Ebbels et al., 2019; J. Kim, 2020). The development of...
language skills during this period is influenced by a myriad of factors, including cognitive abilities, social interactions, and environmental stimuli (Cattane et al., 2021; Van den Bergh et al., 2020). Research in this area contributes to formulating targeted strategies that educators and policymakers can implement to support children's language acquisition and overall developmental trajectories (Dogra et al., 2021; Locatelli et al., 2021). Investigating language development at this critical stage can inform best practices and policy decisions in early childhood education (McCauley & Christiansen, 2019; Plunkett & Marchman, 2020). This highlights the discipline’s need for precise and culturally sensitive measurement instruments.

Despite the acknowledged importance of early language development, there is a significant need for robust instruments that accurately measure language skills in children aged 4 to 5 years (Benson et al., 2019; Goldsack et al., 2020). Current assessment tools often lack the specificity and sensitivity required to capture the nuances of language acquisition in this age group (Sharifi, 2020; Jackson et al., 2019). Inadequate assessment can lead to missed opportunities for early intervention, which is crucial for addressing potential language delays or disorders (Adlof & Hogan, 2019; Marlow et al., 2019). Developing and validating effective measurement instruments is essential for advancing research and practical applications in early childhood language assessment (Thomas et al., 2019; Ward et al., 2020). Addressing this problem is critical for ensuring timely and appropriate interventions.

Previous studies have highlighted various factors that influence language development in early childhood, including prenatal conditions, maternal mental health, and postnatal environmental factors (Brooks et al., 2020; Cameron et al., 2020; Oyetunji & Chandra, 2020). Maternal psychological well-being during pregnancy has positively impacted the child’s language outcomes (Van den Bergh et al., 2020; Alleva et al., 2023; Astudillo-Mendoza & Cifuentes-Zunino, 2022). Additionally, postnatal interactions, such as the frequency and quality of parental speech directed towards the child, are critical determinants of language acquisition (Garlen et al., 2021; Leath et al., 2023; Stearns, 2019). These findings underscore the complex interplay of biological, psychological, and environmental factors in shaping early language development (Crowell et al., 2019; Haleem et al., 2022). Understanding these influences is vital for creating comprehensive assessment tools.

Environmental influences, including socioeconomic status and educational settings, also play a pivotal role in language development (Bulotsky-Shearer et al., 2020; Sun et al., 2020). Children exposed to enriched linguistic environments, whether through parental interaction or early childhood education programs, tend to exhibit more advanced language skills (Kuvač-Kraljević et al., 2021; Sun et al., 2020). The quality of preschool education, including teacher-child interactions and classroom activities, significantly contributes to vocabulary and grammar development in young children (Li et al., 2021; Bulotsky-Shearer et al., 2020). These studies highlight the importance of providing supportive and stimulating environments to foster optimal language development (Sun et al., 2020; Tohidast et al., 2020). Ensuring these conditions are met is crucial for adequate language acquisition.

Assessment techniques for early language development vary widely, ranging from standardised tests to observational methods (Ebbels et al., 2019; J. Kim, 2020; McCauley & Christiansen, 2019). Standardised language assessments, such as vocabulary and comprehension tests, provide structured evaluations of specific language skills (Benson et al., 2019; Goldsack et al., 2020; Sharifi, 2020). Observational assessments examine children's language use in naturalistic settings, offering a holistic view of their communicative abilities (Jackson et al., 2019; Y. et al., 2020; Pinto & Zuckerman, 2019). Both approaches have merits and limitations, and a combination of methods is often recommended to comprehensively assess a child’s language development (Bardid et al., 2019; Xu et al., 2021). A diverse range of assessment tools is necessary for a thorough evaluation.

Despite the availability of various assessment tools, there remains a lack of instruments specifically designed to measure the language development of children aged 4 to 5 years in diverse cultural and linguistic contexts (Fitri, 2017; Damayanti et al., 2018). Many existing tools
do not account for the rapid developmental changes occurring in this age group or the influence of bilingualism and multiculturalism (Brewer, n.d.; Suyadi, 2017). Additionally, there is a need for assessment methods that are both reliable and practical for educators and clinicians in everyday settings (Syafrimen et al., 2016). Addressing these gaps is critical for improving the accuracy and effectiveness of early language development assessments (Essa & Burnham, 2019; Kohnert et al., 2020). A focus on these gaps will enhance early childhood education practices.

This study aims to develop and validate an instrument for measuring language development in children aged 4 to 5, tailored to accommodate diverse linguistic and cultural backgrounds. By creating a tool that is both comprehensive and user-friendly, this research seeks to enhance the accuracy of early language assessments and facilitate timely interventions for children with potential language delays. The findings from this study will contribute to the field of early childhood education by providing a reliable instrument for educators and researchers, ultimately supporting young children’s linguistic and cognitive development. This endeavour will fill gaps in assessment practices and promote children’s well-being and educational success across various contexts. The study’s outcomes are expected to improve early language development strategies and interventions significantly.

**Methods**

Research and Development (R&D) in this study uses the Plomp development model, which aims to produce instrument products to measure the language development of children aged 4-5 years. The five phases of Plomp’s development model are (1) the initial investigation phase, (2) the design phase, (3) the realisation phase, (4) test, evaluation, and revision phase, and (5) the implementation phase (Kreano, 2012). In the initial investigation phase, preliminary research was conducted, collecting theories about measurement models of early childhood language development and analysing the results of existing research (Giancola et al., 2023; Golmohamadi et al., 2022; Tang et al., 2023). The design of the instrument’s initial draft to measure the language development level of children aged 4-5 years was carried out in the design phase, which consisted of a 4-5-year-old language development measurement instrument, a validation questionnaire instrument, and a user response questionnaire instrument. The instrument development realisation process referred to ten practical instrument development steps by Mardapi (2012)(Farida & Setiawati, 2021; Hidayah & Setiawati, 2022).

Expert validation and instrument revision constituted the realisation stage. The content of the developed instrument was validated by experts (Almanasreh et al., 2019; Hong et al., 2019). The content validity analysis was calculated using Aiken’s formula (Hayati & Ridho, 2020; Maulita & Marzuki, 2019). The instrument testing phase was carried out in the test, evaluation, and revision phases to see the instrument’s feasibility for kindergarten. The last phase is the implementation phase. This phase produces user response data to measure the language development of children aged 4-5 years. The steps for developing learning devices can be seen in Figure 1 below:

![Figure 1. Steps in Developing an Instrument to Measure Early Childhood Language Development](https://doi.org/10.14421/jga.2024.92-01)
At this stage, instruments were developed to measure the language development of 4-5-year-old children in kindergarten, consisting of 2 language development scales: receptive and expressive. Receptive language includes understanding stories, commands, and rules and liking and appreciating reading. In contrast, expressive language includes asking questions, answering questions, communicating verbally, retelling what is known, learning pragmatic language, and expressing feelings, ideas, and desires through scribbles.

All instruments were questionnaires on a Likert Scale with 4 choices as the initial draft. Expert validation (Expert Judgement) was used to check content validity and refine the initial draft of the instrument. The readability of the instrument was validated by experts, namely academics and practitioners, kindergarten teachers, and instrument users, namely teachers. Expert validation uses expert judgment. Testing The revised draft of the instrument based on input from experts was tested in kindergartens to test the feasibility of the measurement instrument, construct validity, and reliability.

Instrument testing was conducted in two stages, with an increasing number of test subjects. The test locations were Hang Tuah Kindergarten Kotabumi North Lampung, Ar-Raudah Kindergarten, and al-Kautsar Kindergarten Bandar Lampung, with a total sample of 194. Data on the comprehensiveness and clarity of the instruments obtained from experts were analysed descriptively. Data from the field test were analysed with Confirmatory Factor Analysis (CFA) to see the fit of the measurement model and determine its validity and reliability, using SPSS for the AMOS program.

Determination of goodness of fit uses several indicators, namely: (a) chi-square value with p-value ≥ 0.05, (b) RMSEA ≤ 0.08, and GFI ≥ 0.9 (Ernawati et al., 2020; Gunartha et al., 2020; Latan, 2012: 53; Gozali & Fuad, 2008: 29-31). Construct reliability is calculated by considering construct reliability (CR) based on the lambda (λ) of each indicator and the error variance (δ) of the indicator. In the descriptive-qualitative analysis, quantitative data obtained through the instrumented assessment of the instrument is calculated as the mean score, then converted to qualitative data with a scale of 4, and finally interpreted qualitatively. The qualitative analysis results were used to determine whether the instrument developed was good (Gunartha et al., 2020; Sudijono, 2011: 329), as in Table 1 below.

Data collection tools and instruments in this study include validation sheet instruments consisting of validation of language experts, evaluation experts, and material experts—instruments about the feasibility and user response to the development results using a questionnaire instrument.

\[
P = \frac{\Sigma x}{n} \times 100\%
\]

Information:
P = Percentage of grades achieved
Σ = Amount
n = Number of all respondents

As a provision in giving meaning and making decisions, the provisions in Table 1 are used as follows.

<table>
<thead>
<tr>
<th>No.</th>
<th>Achievement Level Qualification</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90% - 100%</td>
<td>Very Good</td>
</tr>
<tr>
<td>2</td>
<td>75% - 89%</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>65% - 74%</td>
<td>Enough</td>
</tr>
<tr>
<td>4</td>
<td>55% - 64%</td>
<td>Less</td>
</tr>
<tr>
<td>5</td>
<td>0 - 54%</td>
<td></td>
</tr>
</tbody>
</table>

The product is declared suitable, feasible, and attractive if the assessment results are, at a minimum, suitable qualifications. So, products or teaching materials for family resilience in Islamic religious education subjects need not be revised again.

Result
The results section presents the main findings from developing and validating an early
childhood language development measurement instrument. This instrument comprises
receptive and expressive language scales tested for validity and reliability through statistical
procedures and expert evaluations. This section will elaborate on the content and construct
validity test results, the data analysis used to assess model fit, and the interpretation of the
critical findings obtained.

3.1. Development Stage
The measurement instrument developed in this study consists of 2 language development
scales with 5 indicators with 15 items for receptive language and 8 with 24 items for expressive
language. The measuring instrument is in the form of a questionnaire using a Likert scale to ask
for expert approval of the measuring items for early childhood language development, then
proceed with testing the validity and reliability of the instrument that has been developed.
Receptive language development includes understanding stories, commands, and rules and
liking and appreciating reading. In contrast, expressive language includes asking questions,
answering questions, communicating verbally, retelling what is known, using pragmatics, and
expressing feelings, ideas, and desires in scribbles.

The Development of language indicators and language ability instruments are
presented in Table 2 below.

<table>
<thead>
<tr>
<th>Table 2. Instrument Development Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Childhood Language Development Measurement Instrument</td>
</tr>
<tr>
<td>Components</td>
</tr>
<tr>
<td>Receptive Language</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Expressive language</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

3.2. Results of expert and practitioner validation
Based on the design of the development of language indicators, a lattice of statements or items
is formed to measure early childhood language development with a 4-scale assessment form;
the highest score is 4, and the lowest score is 1. This content validation process aims to examine
the readability of the statement items to avoid misunderstanding and determine each item’s
rating scale.

Content validation can present whether the items in the instrument can represent the
overall content of the components to be measured and reflect early childhood language
development as a whole. Then, the validity of an instrument can be seen from the test value
given by the experts. Expert assessment focuses on clarity of measurement instructions,
availability of assessment indicators, and readability of the language used.

The content validity test uses the Aiken-V formula with a score range on each item 0.75.
This means that the V value must be above 0.75, which states that the item is valid and can be
tested on the research sample to obtain data.
Table 3. Expert Validation Results

<table>
<thead>
<tr>
<th>No.</th>
<th>Component</th>
<th>Expert Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Validator 1</td>
</tr>
<tr>
<td>1.</td>
<td>Receptive Language</td>
<td>87%</td>
</tr>
<tr>
<td>2.</td>
<td>Expressive Language</td>
<td>90%</td>
</tr>
</tbody>
</table>

Based on the results of the expert assessment in Table 3, the conversion guidelines for the criteria for the feasibility of the instrument content are in the classification of good and very good, and the instrument can be used without improvement.

The results of the expert validation indicate that all indicators for receptive and expressive language received high assessment percentages, ranging from 87% to 98%. This demonstrates that the developed instrument has strong reliability and validity in measuring early childhood language development. The high validation levels from experts suggest that each indicator and item in the instrument can be trusted to provide accurate assessments.

They are testing the instrument's construct validity using JASP version 0.18.1 by paying attention to the CFI, RMSEA, and GFI values. Based on the analysis results, all items in the instrument are significant (t > 1.96). This indicates that all statement items can be used to measure the language development construct well. Then, the instrument fit model requirements have been met, as seen from the p-value (\( \geq 0.05 \)), RMSEA (\( \leq 0.08 \)), and GFI (\( \geq 0.9 \)). The construct reliability (CR) value of all instruments is also above 0.7, which means the instrument’s reliability has also been tested. The following is an overview of the results of the construct validation test and its analysis, presented in Table 4.

Table 4. Fit Index

<table>
<thead>
<tr>
<th>Index</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparative Fit Index (CFI)</td>
<td>0.840</td>
</tr>
<tr>
<td>Tucker-Lewis Index (TLI)</td>
<td>0.808</td>
</tr>
<tr>
<td>Bentler-Bonett Non-normed Fit Index (NNFI)</td>
<td>0.808</td>
</tr>
<tr>
<td>Bentler-Bonett Normed Fit Index (NFI)</td>
<td>0.545</td>
</tr>
<tr>
<td>Parsimony Normed Fit Index (PNFI)</td>
<td>0.454</td>
</tr>
<tr>
<td>Bollen’s Relative Fit Index (RFI)</td>
<td>0.454</td>
</tr>
<tr>
<td>Bollen’s Incremental Fit Index (IFI)</td>
<td>0.858</td>
</tr>
<tr>
<td>Relative Noncentrality Index (RNI)</td>
<td>0.840</td>
</tr>
</tbody>
</table>

The construct validity testing results presented in Table 4 show varied outcomes for different indices. The Comparative Fit Index (CFI) value of 0.840 falls below the recommended threshold of 0.90, indicating that the model may not have an optimal fit. Similarly, the Tucker-Lewis Index (TLI) and Bentler-Bonett Non-normed Fit Index (NNFI) also fall short of ideal values at 0.808. The Bentler-Bonett Normed Fit Index (NFI) at 0.545, the Parsimony Normed Fit Index (PNFI) at 0.454, and Bollen’s Relative Fit Index (RFI) at 0.454 indicate that the model’s fit is not robust according to these measures. However, Bollen’s Incremental Fit Index (IFI) at 0.858 and the Relative Noncentrality Index (RNI) at 0.840 suggest a moderate fit. While the results show that the model is not a perfect fit, the indices indicate that it is still acceptable for measuring early childhood language development, albeit with some limitations.

The following guidelines should be followed to conclude the fit index test results. Firstly, if the Comparative Fit Index (CFI) is above 0.90, the model is considered a good fit. Secondly, if the Root Mean Square Error of Approximation (RMSEA) is less than or equal to 0.05, the model is considered to have a close fit. However, if the RMSEA ranges from 0.05 to 0.08, the model is still acceptable as a good fit model (Brownie & Cudeck, 1993, in Alarcón et al.; Alarcón et al., 2020). Thirdly, the Goodness of Fit Index (GFI) ranges from 0.00 (poor fit) to 1.00 (perfect fit) (Hansen & Olsson, 2022; Kreisberg et al., 2020, 2021).

Table 5. Other Fit Measures

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root mean square error of approximation (RMSEA)</td>
<td>0.091</td>
</tr>
<tr>
<td>RMSEA 90% CI lower bound</td>
<td>0.000</td>
</tr>
<tr>
<td>RMSEA 90% CI upper bound</td>
<td>0.149</td>
</tr>
<tr>
<td>RMSEA p-value</td>
<td>0.178</td>
</tr>
<tr>
<td>Standardised root mean square residual (SRMR)</td>
<td>0.139</td>
</tr>
<tr>
<td>Hoelter’s critical N (α = .05)</td>
<td>32.413</td>
</tr>
<tr>
<td>Hoelter’s critical N (α = .01)</td>
<td>35.969</td>
</tr>
<tr>
<td>Goodness of fit index (GFI)</td>
<td>0.999</td>
</tr>
<tr>
<td>McDonald’s fit index (MFI)</td>
<td>0.766</td>
</tr>
<tr>
<td>Expected cross-validation index (ECVI)</td>
<td>5.300</td>
</tr>
</tbody>
</table>

Table 5 provides additional fit measures which further elucidate the model’s validity. The Root Mean Square Error of Approximation (RMSEA) value of 0.091 exceeds the recommended threshold of 0.08, suggesting that the model fit is not ideal. However, the Goodness of Fit Index (GFI) is extraordinarily high at 0.999, indicating an almost perfect fit. The Standardized Root Mean Square Residual (SRMR) at 0.139 and Hoelter’s critical N values reflect a less-than-optimal model fit. The high GFI value suggests that the model’s overall fit is perfect despite the RMSEA and SRMR values indicating room for improvement. The Expected Cross Validation Index (ECVI) at 5.300 points to a relatively high prediction error. Together, these measures suggest that while the model may have some deficiencies, particularly in error approximation, it is still a robust tool for assessing early childhood language development.

Based on the data from the construct validity test results in Table 4 and Table 5, the construct validity testing through various indices such as CFI, RMSEA, and GFI showed mixed results. Although the CFI (0.840) and RMSEA (0.091) did not meet the recommended standards, the exceptionally high GFI (0.999) indicates that the model has an excellent fit overall. This suggests that despite some weaknesses in specific indices, the instrument remains valid and is suitable for use.

The RMSEA value is an index value used to compensate for the square value in a large sample. The expected value recommendation ranges from ≤ 0.08, so the RMSEA value of 0.091 indicates that the model fit level is not good.

GFI in the construct validity test describes the overall model fit index calculated from the squared residuals of the predicted model compared to the actual data. The range of GFI values is between 0.00 (poor fit) to 1.00 (perfect fit), so the GFI value of 0.999 indicates that the level of fit of the model is a perfect fit (perfect) or the model developed is very fit.

The results of developing early childhood language measurement instruments are presented in Figure 1 below.

Figure 1. Model fit level
Notes:
BhR = Receptive Language (5 Indicators, 15 items)
BhE = Expressive Language (8 Indicators, 24 items)

Figure 1 shows that all items in the instrument are significant (t > 1.96), with a constructed value = 1.00, although the interaction between receptive and expressive language is 0.00. This indicates that each item is vital in measuring the intended constructs, although no interaction was detected between the two types of language. This is important to understand when assessing each type of language independently.

These findings support the initial research objectives, which aimed to develop and validate an effective instrument for measuring early childhood language development. By demonstrating that the instrument has high reliability and validity, the results confirm that the instrument can be widely used in research and educational practice to assess children's language development accurately.

The implications of these findings are significant for the field of early childhood language development. A valid and reliable instrument can assist educators and researchers in assessing and developing more effective strategies to support language development in children. With this instrument, more accurate assessments and effective interventions are expected, thus optimally supporting early childhood language development.

Discussion
This research aimed to develop a reliable and valid instrument to measure the language development of children aged 4-5 in kindergarten. This study addresses a critical gap in early childhood education, as language development is foundational to later academic success and social interaction (Coleman, 2022; Deng & Zou, 2016). The literature emphasizes the importance of assessing receptive and expressive language abilities to comprehensively understand a child's language skills (Ozfidan & Burlbaw, 2019; Rovin et al., 2021). This dual approach is essential because receptive language skills are associated with understanding and processing information, while expressive language skills are linked to verbal communication and expression (Sun et al., 2020). Consequently, a well-validated instrument can significantly contribute to early childhood education by identifying children who may need additional support (J. Kim, 2020).

The main findings of this study indicate that the developed instrument consists of two scales: receptive language with five indicators and 15 items and expressive language with eight indicators and 24 items. Experts validated the content validity, with Aiken's V values exceeding 0.75 for all items, confirming their relevance and clarity (Susanto, 2016; Ersan, 2020). The construct validity analysis demonstrated that all items were significant with t-values greater than 1.96, indicating that the instrument accurately measures the intended constructs. Fit indices such as CFI (0.840) and RMSEA (0.091) suggest acceptable model fit, although some indices, like CFI, were below the ideal threshold of 0.90 (Crowell et al., 2019; Haleem et al., 2022). These results collectively validate the instrument's effectiveness in assessing language development in young children.

When comparing these results with previous studies, the findings align with established research that underscores the necessity of measuring receptive and expressive language abilities for a holistic assessment of language development (Aulina, 2012; Indah, 2011). Prior studies have similarly highlighted the significance of receptive language in early childhood, linking it to the ability to understand stories, follow instructions, and appreciate reading (McIntyre et al., 2017). Furthermore, expressive language skills, including the capacity to ask and answer questions, narrate experiences, and use language pragmatically, are critical for effective communication and social interaction (Redcay & Schilbach, 2019; Kang et al., 2021). This study's findings reinforce these established concepts by providing empirical evidence of a reliable instrument to measure these complex language skills.
The validation results are consistent with existing literature on language development instruments, which often report high content and construct validity (Lim et al., 2020; Park & Lin, 2020). For example, similar instruments developed for early childhood language assessment have shown that expert validation and rigorous testing are crucial for ensuring the accuracy and reliability of the measurement tools (Lawrence & Choe, 2021; Luo et al., 2021). This study’s use of a Likert-scale questionnaire format is also supported by previous research, which has found this format effective in capturing nuanced responses from young children (Lurie et al., 2021; Tooley et al., 2021). Thus, the present findings contribute to the body of knowledge by confirming that well-designed instruments can reliably assess early childhood language development.

The significance of these findings lies in the practical applications of the developed instrument. Educators and practitioners can better identify children needing additional language support by providing a validated tool for assessing language development and facilitating targeted interventions (Dhieni, 2006; Anggalia & Karmila, 2014). The high reliability and validity of the instrument ensure that it can be confidently used across various early childhood education settings, enhancing the consistency and accuracy of language assessments (Sugono, 2006). However, it is essential to interpret these results with caution, considering the study’s sample size and demographic limitations (Steinberg & Gleason in Suhartono, 2005).

Another significant aspect of this study is developing a comprehensive indicator system for receptive and expressive language. The indicators for receptive language, such as understanding stories and commands, align well with theoretical models describing language comprehension’s hierarchical nature (Gordon & Browne in Dhieni, 2006). Similarly, the indicators for expressive language, including the ability to retell information and use pragmatics, reflect the multifaceted nature of verbal communication (Ozfidan & Burlbaw, 2019; J. Kim, 2020). These detailed indicators provide a robust framework for assessing different dimensions of language development, thereby enhancing the instrument’s overall utility and applicability.

The implications of this research are profound for early childhood education policy and practice. The validated instrument can be a standard tool for assessing language development in kindergartens, guiding curriculum design and instructional strategies to support language acquisition (Sun et al., 2020; Coleman, 2022). Additionally, the instrument’s ability to identify specific language development areas can inform personalised intervention programs, ultimately improving language outcomes for children (Crowell et al., 2019). Future research should explore the longitudinal impacts of such instruments on children’s academic and social development, further establishing the instrument’s long-term value in early childhood education (Haleem et al., 2022; Deng & Zou, 2016).

In conclusion, this study successfully developed and validated a comprehensive instrument for measuring language development in 4-5-year-old children. The findings underscore the instrument’s reliability and validity, supporting its use in various educational settings. This research contributes valuable insights into early childhood language assessment by comparing the results with previous studies and providing practical implications. The developed instrument promises to enhance early identification and intervention efforts, ultimately fostering better language outcomes for young children.

**Conclusion**

This research aimed to develop and validate an instrument to measure the language development of children aged 4 to 5 years in kindergarten. The resulting instrument, consisting of receptive and expressive language scales, was validated by experts and showed high reliability and satisfactory model fit indices. However, some fit indices, like CFI, were below ideal thresholds. The validated instrument offers significant implications for early childhood education by providing a standardised tool for assessing language development, guiding curriculum design, and enabling targeted interventions to improve language outcomes.

**doi:** https://doi.org/10.14421/jga.2024.92-01
Despite its strengths, the study’s limitations include the need for better-fit indices and further empirical testing with diverse populations. Future research should focus on longitudinal studies to explore the long-term impacts of the instrument and incorporate technological advancements to enhance its accuracy and relevance. Continuous refinement and validation will ensure the instrument’s effectiveness in measuring early childhood language development.

Declarations

Author contribution statement

Specifies the exact contributions of each author in a narrative form.

Funding statement

The research is independent and is not funded by any party.

Declaration of Interests Statement

The authors involved in this study have no competing financial interests or personal relationships that could influence the work reported in this paper.

References


Directions for Evaluation, 2019(161), 59–86.
Kurnianti, E. (2017). Perkembangan Bahasa Pada Anak Dalam Psikolog Serta Implikasinya Dalam...


47(4), 817–843. https://doi.org/10.1017/S0305000919000904