



A Literature Review of Language Assessment Scale

Gao, Li^{1*} & Ahmad, Hishamuddin¹

¹Faculty of Human Development, Sultan Idris Education University, 35900 Tanjong Malim, Perak, MALAYSIA

*Corresponding author: 229378915@qq.com

Received 29 May 2024; Accepted 28 June 2024; Available online: 28 June 2024

Abstract: Language assessment scales are essential tools in both educational and clinical contexts, employed to evaluate language proficiency and inform instructional strategies. This literature review seeks to identify and analyze the most highly cited studies related to language assessment scales. The aim of this research is to synthesize past findings and highlight current trends. The methodology involved a systematic search of the Scopus database, selecting articles published between 2014 and 2024 that aligned with the study's focus. The findings revealed several critical themes: first, the importance of validating assessment tools such as the Preschool Language Assessment Scale (PreLAS) and the Expressive One-Word Picture Vocabulary Test (EOWPVT-3: SBE) to ensure they accurately reflect the language abilities of dual language learners (DLLs). Second, the role of communicative competence in teacher education was emphasized, particularly the need for targeted teaching strategies and assessment tools. Third, the review highlighted the distinct challenges of assessing language for specific purposes (LSP) in professional contexts, especially in Asian call centres. Additionally, the review underscored the unique challenges faced by special populations, such as deaf or hard-of-hearing students, and examined the potential of content-based interventions, like shared book reading, to enhance vocabulary development among DLL preschoolers. The implications of these findings suggest that while current tools are valuable, there is a continuous need for refinement to better address the specific linguistic needs of diverse learners. This review contributes to the academic discourse by proposing themes that warrant further investigation, emphasizing the necessity for culturally relevant and psychometrically sound language assessment tools.

Keywords: Literature review, language assessment scale, education

1. Introduction

Language assessment scales are indispensable tools in both educational and clinical environments, serving to evaluate learners' language proficiency and inform interventions or educational strategies (Kjell et al., 2023). The development and implementation of these scales are pivotal for accurate diagnosis and effective language instruction, particularly within diverse and multilingual settings. This literature review aims to identify the most highly cited studies related to language assessment scales published in Scopus and to explore the themes emerging from these findings. By focusing on influential research within this field, this review provided insights into the key areas of interest and the dominant themes that have shaped the discourse around language assessment. The identification and analysis of these high-impact studies helped to elucidate the prevailing trends and challenges in the field, offering a comprehensive understanding of the state of research on language assessment scales. Through this exploration, this review contributed to the broader academic discourse by synthesising past findings and proposing themes that warrant further investigation.

The necessity for reliable and valid language assessment tools is paramount, particularly given the increasing recognition of language proficiency as a crucial determinant of academic success and social integration (Schmitt et al., 2020). Standardised language assessments are widely employed to inform decisions related to educational placement, therapeutic interventions, and eligibility for services. However, the effectiveness of these assessments is contingent upon their psychometric properties, including validity, reliability, and fairness. A systematic review by Denman et al. (2017) underscores the significance of these properties, highlighting that while many language assessments demonstrate evidence of validity and reliability, substantial gaps persist, particularly in structural validity and error measurement.

Following that, language assessment scales must also be adaptable to diverse linguistic and cultural contexts, especially in multilingual environments. The challenges associated with assessing English language learners exemplify the complexities of developing scales that are both accurate and equitable. Stoll et al. (2021) found that traditional teacher assessments often fail to accurately identify reading disabilities among English language learners, underscoring the necessity for more sensitive and culturally responsive assessment tools.

Moreover, linguistic and cultural considerations aside, the role of language assessment in various educational contexts warrants further examination. North & Piccardo (2023) reviews language proficiency scales such as the Common European Framework of Reference (CEFR) and the Canadian Language Benchmarks (CLB), discussing their application in educational settings to standardise language assessment across different regions and learner populations. Additionally, the evaluation of rating quality in language assessments, particularly those requiring rater judgment, represents a critical area of concern. Aryadoust et al. (2021) emphasise the importance of inter-rater reliability and advocate for modern measurement techniques to ensure fairness and accuracy in high-stakes language assessments. Therefore, this review answered two research questions:

- 1) What are the top cited past studies published in Scopus related to language assessment scale?
- 2) What are the themes drawn from the findings of the top cited past studies selected?

2. Literature Review

A language assessment scale is an instrumental tool used to evaluate learners' language proficiency across various dimensions, such as comprehension, expression, and fluency (Wu et al., 2020). These scales are critical in educational, clinical, and research contexts for diagnosing language disorders, guiding language instruction, and assessing language development. Language assessment scales vary considerably in their design and application, being tailored to specific populations, age groups, and linguistic contexts (Rolstad & MacSwan, 2024).

At its core, a language assessment scale quantifies language abilities by providing a structured framework for evaluating language skills (Aikens et al., 2020). Typically, these scales include a variety of items or tasks that test different aspects of language, including vocabulary, grammar, phonology, and pragmatics. The items are designed to increase in difficulty, allowing for a detailed assessment of the individual's language abilities. The results are often compared against standardised norms, enabling educators and clinicians to determine whether an individual's language skills fall within the expected range for their age or developmental level.

One of the most widely used language assessment tools is the Preschool Language Scale (PLS), which evaluates language skills in children from birth to six years of age. This scale is designed to assess both receptive (understanding) and expressive (speaking) language abilities. It includes tasks that measure preverbal behaviours, vocabulary, sentence structure, and more advanced language skills. The PLS has been extensively utilised in research and clinical practice to identify language delays and disorders, monitor language development over time, and evaluate the effectiveness of interventions (Zimmerman & Castilleja, 2005).

For language assessment scales to be considered reliable and valid, they must meet rigorous psychometric standards. Reliability refers to the consistency of assessment results, while validity concerns the accuracy with which the scale measures what it is intended to measure. A systematic review by Denman et al. (2017) examined the psychometric properties of various language assessment scales used with children aged four to twelve years. The study found that while many scales demonstrated some level of reliability and validity, few provided comprehensive evidence across all psychometric domains, highlighting the need for ongoing research to enhance these tools (Denman et al., 2017).

Another important aspect of language assessment scales is their application in rating scales used for performance assessments (Knoch et al., 2021). Rating scales, such as those employed in speaking and writing assessments, help evaluate the quality of language use in real-life contexts. These scales may be holistic, providing an overall score based on general language performance, or analytic, breaking down the assessment into specific components like grammar, vocabulary, and fluency. The choice of scale type depends on the purpose of the assessment and the specific language abilities being measured (Botes et al., 2021). Hence, language assessment scales are indispensable tools for assessing and diagnosing language abilities. Their design and application must be carefully considered to ensure they are both psychometrically sound and culturally relevant. As research in language assessment continues to evolve, these tools will undoubtedly become increasingly refined and effective in addressing the diverse needs of language learners.

3. Methodology

To identify relevant and suitable past studies, the Scopus database was utilised to source high-quality articles. The primary keyword used in the search was "language assessment scale." The initial search yielded 34 articles. However, some outdated studies were not considered, as only articles published between 2014 and 2024 were selected to ensure relevance and contemporaneity. This initial selection process filtered the results down to 16 articles. Subsequently, a second round of selection was conducted based on a rigorous quality appraisal. Given that this literature review is focused on an educational context rather than a clinical one, certain articles were excluded. Ultimately, 6 articles were selected for inclusion in this review, based on their alignment with the study's focus and their overall quality.

4. Findings

The findings for each research question were reported in a structured and systematic manner, ensuring clarity and coherence in the presentation of results.

4.1 What Are the Top Cited Past Studies Published in Scopus Related to Language Assessment Scale?

Table 1 presented the top-cited studies published in Scopus related to language assessment scales in education. This table serves as a concise summary of the most influential research in the field. The table summarises main top-cited past studies with related to language assessment, focusing on topics such as dual language learners, teacher training, language performance in specific contexts, and the effects of interventions on language learning. The studies vary in their impact, as reflected by the number of citations each has received. For instance, the study by Rainelli et al. (2017) on the validity of the first two subtests of the Preschool Language Assessment Scale for Spanish-speaking preschool children has been cited 22 times, indicating its significant influence in the field. In contrast, the study by Aikens et al. (2020) on screening approaches for dual language learners has received 8 citations, while the research conducted by Gràcia et al. (2020) and Lockwood (2015) on communicative competence in initial teacher training and language performance assessment in Asian call centres, respectively, have each garnered 7 citations.

Table 1. Top cited past studies

No.	Author/Year	Title	Citation
1	Rainelli et al. (2017)	Validity of the first two subtests of the preschool language assessment scale as a language screener for Spanish-speaking preschool children	22
2	Aikens et al. (2020)	Screening approaches for determining the language of assessment for dual language learners: Evidence from Head Start and a universal preschool initiative	8
3	Gràcia et al. (2020)	Development and assessment of communicative competence in initial teacher training	7
4	Lockwood (2015)	Language for Specific Purpose (LSP) performance assessment in Asian call centres: strong and weak definitions	7
5	Lau et al. (2019)	Oral Language Performance of Deaf and Hard-of-Hearing Students in Mainstream Schools	3
6	Gonzalez et al. (2024)	The Effects of a Science and Social Studies Content Rich Shared Reading Intervention on the Vocabulary Learning of Preschool Dual Language Learners	0

The work by Lau et al. (2019) on oral language performance in deaf and hard-of-hearing students in mainstream schools has received 3 citations, reflecting a more specialised focus. Lastly, the study by Gonzalez et al. (2024) on the effects of a content-rich shared reading intervention on vocabulary learning in preschool dual language learners has yet to be cited, likely due to its recent or forthcoming publication. These studies collectively provide a diverse overview of research within the field of language assessment, each with varying degrees of recognition and impact within the academic community. Fig. 1 shown for visualisation of the findings.

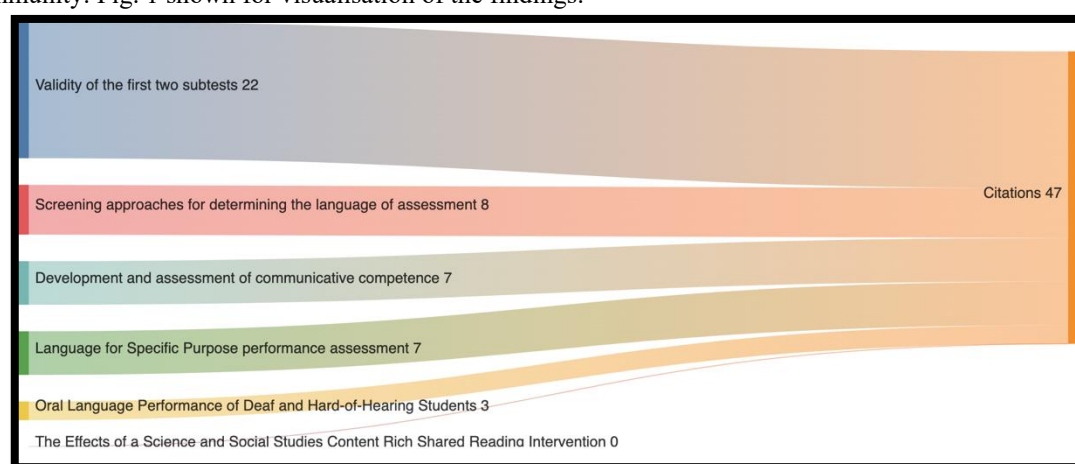


Fig. 1. Visualisation of findings

4.2 What Are the Themes Drawn from The Findings of the Top-Cited Past Studies Selected?

Before identifying the related themes, the findings from each of the selected studies are reported below. To begin with, Rainelli et al. (2017) examined the use of the first two subscales of the Preschool Language Assessment Scale, "Simon Says" and "Art Show", in large-scale early childhood studies. These subscales are employed to guide decisions regarding

the most appropriate language (or languages) researchers should use when directly assessing the academic skills of dual language learner (DLL) children. Specifically, large-scale studies utilise a cut-score derived from the total score on these two PreLAS subscales in English and/or Spanish, in conjunction with parent or teacher reports of children's language abilities, to determine the most suitable language for assessment. However, limited research supports the use of these cut-scores as part of a language routing procedure for Spanish-speaking DLL preschool children from low-income backgrounds. This study investigated the validity of the two English PreLAS subscale scores using a local sample of children enrolled in Head Start ($N = 872$) and Hispanic children from the national FACES 2006 sample ($N = 935$). Rasch and DIF analyses supported the invariance of item difficulty values across the three- and four-year-old age groups in the overall sample. For a subsample of Spanish-speaking DLL children, receiver operating characteristic (ROC) analyses identified the most appropriate cut-scores on the PreLAS screener for both age groups. The findings provided evidence supporting the validity of the English PreLAS language screener score as part of a more comprehensive language routing procedure.

Next, Aiken et al. (2020) conducted analyses to evaluate the performance of the Preschool Language Assessment Scale (preLAS) and the Expressive One-Word Picture Vocabulary Test (EOWPVT-3: SBE), a conceptually scored vocabulary measure, in determining DLL children's language path through a direct assessment battery. The study drew on data from two programmes serving linguistically diverse children: the Head Start Family and Child Experiences Survey (FACES) and the Universal Preschool Child Outcomes Study (UPCOS), both of which utilised the preLAS for routing purposes. Several key findings emerged that should inform future language routing procedures. The results suggested that beyond its use for language screening purposes, the preLAS could help differentiate DLL children's scores on other developmental measures. Additionally, the items were not ordered by difficulty, which has implications for how preLAS scores should be interpreted. The study also found that the language of response on the EOWPVT-3: SBE indicated Spanish-English DLLs' readiness to respond to language-specific assessments in English. The research suggests that when using the preLAS to appropriately route children, assessment procedures should consider total rather than consecutive errors. Furthermore, other assessment measures, such as the EOWPVT-3: SBE, may be useful in language routing, specifically for Spanish-English DLLs.

Gràcia et al. (2020) explored the importance of developing communicative competence in the education of future teachers, given its fundamental role as an instrument of communication, knowledge production and transformation, and as a means to enhance learning quality and critical thinking, both in relation to studies and future professional performance. Despite its importance, few studies have focused on this area. This study contributes to filling this gap by presenting the results of a study aimed at improving the construction of argumentative oral texts and metalinguistic reflection on their elements through the development and introduction of specific teaching strategies and instruments. A mixed-method approach was used, incorporating both quantitative and qualitative analyses. The instruments employed included the Evaloe, or Escala de Evaluación de la Lengua Oral en Contexto Escolar (i.e., Oral Language Assessment Scale in a School Context), and a rubric specifically designed for this purpose. The results demonstrated that with the introduction of appropriate teaching strategies and instruments, students could learn to construct texts involving one or more interlocutors and reflect on these texts.

Lockwood (2015) addressed the growing focus on LSP (Language for Specific Purposes) performance assessment, where language testers have debated how testing tools and processes can be refined to more accurately and validly assess professional communication at work. Several suggestions have been made to enhance LSP performance assessment validity, including ethnographic studies of target language situations, authentic discourse analyses of relevant texts, and the involvement of subject matter experts (SMEs) as informants in defining 'successful' workplace communication. These steps are considered crucial for building validity in LSP performance assessments. Some language testing researchers have also proposed distinguishing between 'weak' and 'strong' LSP performance, although clear lines of distinction have yet to be established in LSP assessment studies. In this study, Lockwood proposes a distinction between 'weak' and 'strong' versions of LSP performance assessment, based on empirical data collected over ten years of developing and embedding the Business Performance Language Assessment Scales (BUPLAS) in Asian call centres. The article argues that a workplace LSP spoken assessment used for recruitment (predictive of work success) constitutes a 'weak' version, while a workplace LSP spoken assessment used for quality assurance (observed success at work) represents a 'strong' version. This distinction is particularly important in the Asian call centre industry, where the purposes of recruitment and quality assurance differ significantly.

Lau et al. (2019) examined the language outcomes of Cantonese-speaking deaf or hard-of-hearing children attending primary schools within the Hong Kong educational system. In Hong Kong, students are expected to be fluent in Cantonese, Putonghua, and English, yet the curriculum does not include Cantonese studies, as it is assumed that children have acquired Cantonese by school entry. The study utilised the Hong Kong Cantonese Oral Language Assessment Scale, which comprises six subtests, to assess 98 children with mild to profound hearing loss. Regression analysis was employed to examine the influence of various factors on oral language performance in these children. The results revealed that 41% of the participants had achieved age-appropriate oral language skills, while 18% and 41% exhibited mild-to-moderate or severe oral language impairments, respectively. The degree of hearing loss and the use of speech therapy were identified as significant negative predictors of oral language performance. The study highlighted issues such as late diagnosis and

delayed fitting of hearing devices, as well as poor oral language outcomes, emphasising the need for policymakers to reconsider the existing educational approaches and support for deaf or hard-of-hearing children.

Finally, Gonzalez et al. (2024) investigated the effects of a content-based shared book reading (SBR) intervention on the receptive and expressive vocabulary outcomes of DLL preschool children enrolled in two school districts in South Texas. The study focused on 50 preschool teachers and 298 preschoolers, who were randomly assigned at the class level to either a well-specified and scripted SBR condition or a comparison SBR condition. The children were selected based on their scores on the school district-administered Preschool Language Assessment Scales (PreLAS®), which determined them to be at the Limited English Speaker (LES) level of English proficiency. Teachers in the intervention condition implemented the curriculum over 18 weeks in five-day instructional cycles, with approximately 20 minutes of instruction per day. The results indicated significant and robust effects on proximal measures of expressive and receptive vocabulary, although no significant effects were found on standardised measures. Table 2 shown for the themes drawn.

Table 2. Themes drawn from findings

Theme	Key Studies	Summary
Validity and Utility of Language Assessment Tools	Rainelli et al. (2017); Aiken et al. (2020)	Emphasizes the importance of validating tools like PreLAS and EOWPVT-3: SBE to ensure accurate reflection of DLLs' language abilities.
Assessment and Development of Communicative Competence	Gràcia et al. (2020)	Highlights the role of specific teaching strategies and assessment tools, such as Evaloe scale, in enhancing argumentative oral texts among future teachers.
Language for Specific Purposes (LSP) Assessment	Lockwood (2015)	Addresses challenges of assessing professional communication within LSP, particularly in Asian call centers. Distinguishes between 'weak' and 'strong' LSP performance assessments.
Language Outcomes for Special Populations	Lau et al. (2019)	Focuses on the oral language performance of deaf or hard-of-hearing children in multilingual educational settings. Emphasizes early intervention and tailored educational support.
Impact of Content-Based Interventions on Language Learning	Gonzalez et al. (2024)	Examines the effects of content-rich shared book reading interventions on vocabulary development of DLL preschoolers. Suggests need for additional strategies to achieve broader language development goals.

A detailed examination of the selected studies reveals several overarching themes in language assessment research, particularly concerning dual language learners (DLLs), language for specific purposes (LSP), and communicative competence. A prominent theme is the validity and utility of language assessment tools. Studies by Rainelli et al. (2017) and Aiken et al. (2020) underscore the critical importance of validating tools such as the Preschool Language Assessment Scale (PreLAS) and the Expressive One-Word Picture Vocabulary Test (EOWPVT-3: SBE) to ensure they accurately reflect DLLs' language abilities. These tools play a pivotal role in guiding language routing procedures in large-scale educational settings, highlighting the necessity for meticulous consideration of their application and cut-scores. Another significant theme pertains to the assessment and development of communicative competence. Gràcia et al. (2020) emphasise the role of targeted teaching strategies and assessment tools, such as the Evaloe scale and custom rubrics, in enhancing the construction and reflection of argumentative oral texts among future teachers. This underscores the broader need for developing communicative skills as an essential component of professional training, particularly in fields where language is fundamental to communication and learning. The theme of Language for Specific Purposes (LSP) assessment is explored in Lockwood's (2015) study, which addresses the complexities of assessing professional communication within LSP, particularly in Asian call centres. The study introduces a distinction between 'weak' and 'strong' LSP performance assessments, tailored to specific purposes such as recruitment versus quality assurance, thereby underscoring the importance of aligning assessments with real-world professional demands.

In the context of special populations, the theme of language outcomes for these groups emerges as another critical area of focus. Lau et al. (2019) examine the oral language performance of deaf or hard-of-hearing children within multilingual educational settings. The study identifies challenges such as late diagnosis and delayed access to hearing devices, which adversely affect language outcomes, thereby emphasising the need for early intervention and tailored educational support. Lastly, the theme of the impact of content-based interventions on language learning is investigated in Gonzalez et al. (2024), which examines the effects of content-rich shared book reading interventions on the vocabulary development of DLL preschoolers. While the study demonstrates significant benefits in expressive and receptive vocabulary, the absence of significant effects on standardised measures suggests that these interventions may need to be supplemented with additional strategies to achieve broader language development objectives. Collectively, these themes highlight the complexity of language assessment research, emphasising the importance of validating assessment tools,

developing communicative competence, tailoring assessments to specific purposes, addressing the needs of special populations, and exploring the efficacy of targeted interventions to optimise language assessment across various contexts. A visualisation was shown below for better comprehension.

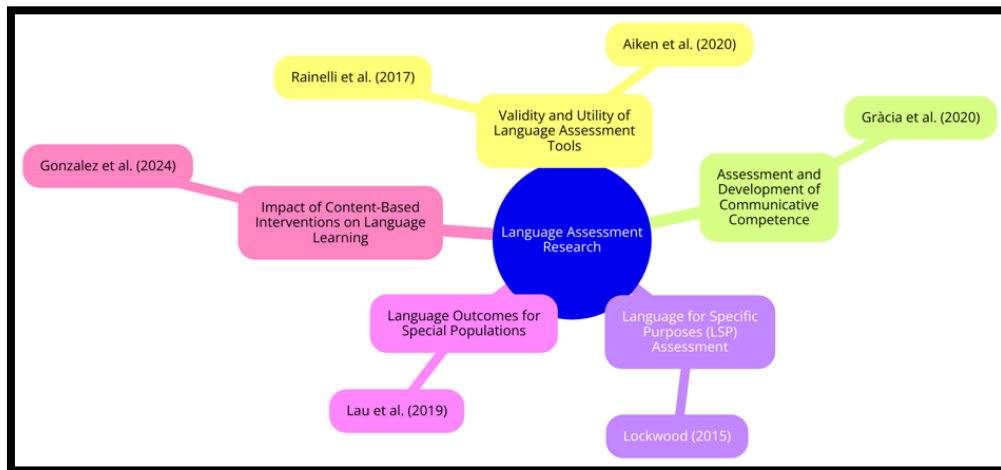


Fig. 2. Visualisation of themes withdrawn

5. Discussion

The findings from this literature review on language assessment scales underscore both significant advancements and enduring challenges within the field of language proficiency evaluation. The research by Rainelli et al. (2017) and Aiken et al. (2020) provides crucial validation for widely used tools such as the Preschool Language Assessment Scale (PreLAS) and the Expressive One-Word Picture Vocabulary Test (EOWPVT-3: SBE). These studies are consistent with earlier research (Abedi et al., 2020; Schmitt et al., 2020) emphasising the necessity of rigorous validation processes for language assessment tools to ensure they accurately reflect the language abilities of dual language learners (DLLs). Similar to the findings of Kjell et al. (2023), these studies reaffirm that without proper validation, language assessments can lead to misdiagnosis or inappropriate educational placements, particularly for linguistically diverse populations. However, what distinguishes these studies is their focus on the specific application of these tools in large-scale educational settings, highlighting the need for precise cut-scores and culturally sensitive assessment procedures—an area that has been underexplored in prior research.

In contrast, the study by Gràcia et al. (2020) on developing communicative competence in teacher education diverges from earlier research, which has traditionally concentrated more on the technical aspects of language assessment rather than its application in fostering critical communicative skills. This study offers a fresh perspective by integrating assessment tools such as the Evaloe scale into teacher training, demonstrating that language assessment is not solely about measuring proficiency but also about cultivating the skills necessary for effective communication in professional settings. This finding is particularly significant as it underscores a shift in the field towards recognising the broader educational implications of language assessment, extending beyond mere proficiency to encompass the development of essential communication skills for future educators.

Lockwood's (2015) exploration of Language for Specific Purposes (LSP) assessment introduces a novel approach by distinguishing between 'weak' and 'strong' performance assessments, a distinction that has not been thoroughly addressed in earlier studies. This approach underscores the need for assessments tailored to specific professional contexts, ensuring that they are both relevant and effective in predicting job performance. This finding is critical as it challenges the one-size-fits-all approach often seen in language assessment (Lee et al., 2021) and calls for more nuanced tools that cater to the specific demands of different professional environments, particularly in industries such as Asian call centres.

The study by Lau et al. (2019) on the oral language performance of deaf or hard-of-hearing students adds another layer of complexity to the discussion, emphasising the unique challenges faced by this population. The findings align with previous research highlighting the need for early intervention and tailored educational support for special populations (Colizzi et al., 2020; Tabone et al., 2020). However, this study goes further by demonstrating the significant impact of delayed diagnosis and lack of access to hearing devices on language outcomes, reinforcing the urgent need for policy changes to address these gaps in support.

Finally, the investigation by Gonzalez et al. (2024) into content-based interventions, such as shared book reading, provides critical insights into the potential of these interventions to enhance vocabulary development among DLL preschoolers. While the study's findings on expressive and receptive vocabulary gains are promising, the absence of significant effects on standardised measures suggests that these interventions, while beneficial, may not be sufficient on their own. This highlights the need for additional strategies or more comprehensive approaches to achieve broader language development goals, an area that previous studies may not have fully addressed.

This review is essential as it synthesises current research, identifies gaps in the literature, and proposes directions for future inquiry. It contributes to the ongoing academic discourse by emphasising the need for continuous refinement of language assessment tools to ensure they are both culturally relevant and psychometrically sound. Moreover, by drawing connections between the reviewed studies and existing literature, this review underscores the complexity of language assessment and the necessity for tailored, context-specific approaches to meet the diverse needs of learners across various settings.

The literature review on language assessment scales reveals a multifaceted and evolving field, marked by ongoing efforts to enhance the reliability, validity, and applicability of these tools across diverse educational and clinical settings. The studies reviewed underscore the critical importance of validating language assessment tools to ensure they accurately reflect the language abilities of dual language learners (DLLs) and other special populations. The research conducted by Rainelli et al. (2017) and Aiken et al. (2020) highlights the necessity of rigorous validation processes for tools such as the Preschool Language Assessment Scale (PreLAS) and the Expressive One-Word Picture Vocabulary Test (EOWPVT-3: SBE), which are pivotal in guiding language routing procedures in large-scale educational contexts. These findings indicate that while current tools are valuable, there is a continuous need for refinement to better address the specific linguistic needs of diverse learners.

Another significant contribution of this review is the emphasis on developing communicative competence, particularly within the context of teacher education. Gràcia et al. (2020) demonstrate the effectiveness of targeted teaching strategies and assessment tools, such as the Evaloe scale, in enhancing the construction and reflection of argumentative oral texts among future educators. This focus on communicative competence is vital, as it underscores the broader role of language as a tool for communication, knowledge production, and critical thinking—skills that are essential in both educational and professional settings.

The review also illuminates the complexities of Language for Specific Purposes (LSP) assessment, particularly in professional environments such as Asian call centres. Lockwood (2015) introduces the distinction between 'weak' and 'strong' LSP performance assessments, tailored to specific professional demands such as recruitment and quality assurance. This distinction is crucial, as it highlights the need for assessments that are aligned with real-world professional requirements, ensuring that language assessments are both relevant and effective in predicting and evaluating job performance. In the context of special populations, the study by Lau et al. (2019) on the oral language performance of deaf or hard-of-hearing children within multilingual educational settings underscores the significant challenges these students face, such as late diagnosis and delayed access to hearing devices. These findings emphasise the urgent need for early intervention and tailored educational support to improve language outcomes for these vulnerable groups. Finally, the impact of content-based interventions on language learning is critically examined in Gonzalez et al. (2024), which explores the effects of shared book reading interventions on the vocabulary development of DLL preschoolers. While the study demonstrates significant benefits in expressive and receptive vocabulary, the lack of significant effects on standardised measures suggests that these interventions, although beneficial, may require supplementation with additional strategies to achieve broader language development goals.

6. Conclusion

In conclusion, this review highlights the complexity and significance of language assessment research, emphasising the need for ongoing validation of assessment tools, the development of communicative competence, the tailoring of assessments to specific purposes, the support of special populations, and the exploration of effective interventions. These insights contribute to a deeper understanding of how language assessments can be optimised to meet the diverse needs of learners across various contexts. Future research should continue to explore these areas, with particular attention to refining assessment tools to ensure they are both culturally relevant and psychometrically sound.

Acknowledgement

The authors would like to thank fellow authors and organizations whose intellectual property was utilized for this study.

Conflict of Interest

The authors declare no conflicts of interest.

References

- Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence-based computer accommodation decision system. *Educational Measurement: Issues and Practice*, 39(4), 41-52. <https://doi.org/10.1111/emip.12328>
- Aikens, N., West, J., McKee, K., Moiduddin, E., Atkins-Burnett, S., & Xue, Y. (2020). Screening approaches for determining the language of assessment for dual language learners: Evidence from head start and a universal preschool initiative. *Early Childhood Research Quarterly*, 51, 39-54. <https://doi.org/10.1016/j.ecresq.2019.07.008>

- Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 38(1), 6-40. <https://doi.org/10.1177/0265532220927487>
- Botes, E., Dewaele, J., & Greiff, S. (2021). The development and validation of the short form of the foreign language enjoyment scale. *The Modern Language Journal*, 105(4), 858-876. <https://doi.org/10.1111/modl.12741>
- Colizzi, M., Lasalvia, A., & Ruggeri, M. (2020). Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care?. *International journal of mental health systems*, 14, 1-14. <https://doi.org/10.1186/s13033-020-00356-9>
- Denman, D., Speyer, R., Munro, N., Pearce, W. M., Chen, Y. W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4–12 years: A systematic review. *Frontiers in psychology*, 8, 1515. <https://doi.org/10.3389/fpsyg.2017.01515>
- Gonzalez, J. E., Kim, H., Anderson, J., & Pollard-Durodola, S. (2024). The effects of a science and social studies content rich shared reading intervention on the vocabulary learning of preschool dual language learners. *Early Childhood Research Quarterly*, 66, 34-47. <https://doi.org/10.1016/j.ecresq.2023.08.011>
- Gràcia, M., Jarque, M. J., Astals, M., & Rouaz, K. (2020). Desarrollo y evaluación de la competencia comunicativa en la formación inicial de maestros. *Revista iberoamericana de educación superior*, 11(30), 115-136. <https://doi.org/10.22201/issue.20072872e.2020.30.591>
- Kjell, O. N., Kjell, K., & Schwartz, H. A. (2023). Beyond rating scales: With targeted evaluation, language models are poised for psychological assessment. *Psychiatry Research*, 115667. <https://doi.org/10.31234/osf.io/yfd8g>
- Knoch, U., Deygers, B., & Khamboonruang, A. (2021). Revisiting rating scale development for rater-mediated language performance assessments: Modelling construct and contextual choices made by scale developers. *Language Testing*, 38(4), 602-626. <https://doi.org/10.1177/0265532221994052>
- Lau, T. H., Lee, K. Y., Lam, E. Y., Lam, J. H., Yiu, C. K., & Tang, G. W. (2019). Oral language performance of deaf and hard-of-hearing students in mainstream schools. *The Journal of Deaf Studies and Deaf Education*, 24(4), 448-458. <https://doi.org/10.1093/deafed/enz012>
- Lee, P., Liu, Y., & Tseng, W. (2021). One size fits all? In search of the desirable caption display for second language learners with different caption reliance in listening comprehension. *Language Teaching Research*, 25(3), 400-430. <https://doi.org/10.1177/1362168819856451>
- Lockwood, J. E. (2015). Language for Specific Purpose (LSP) performance assessment in Asian call centres: strong and weak definitions. *Language Testing in Asia*, 5, 1-11. <https://doi.org/10.1186/s40468-014-0009-6>
- North, B., & Piccardo, E. (2023). Aligning language frameworks: An example with the CLB and CEFR. *Language Assessment Quarterly*, 20(2), 143-165. <https://doi.org/10.1080/15434303.2023.2184266>
- Rainelli, S., Bulotsky-Shearer, R. J., Fernandez, V. A., Greenfield, D. B., & López, M. (2017). Validity of the first two subtests of the preschool language assessment scale as a language screener for Spanish-speaking preschool children. *Early Childhood Research Quarterly*, 38, 10-22. <https://doi.org/10.1016/j.ecresq.2016.08.001>
- Rolstad, K., & MacSwan, J. (2024). 5 bilingual language assessment: A persistent case of bias. *Equity in Multilingual Schools and Communities*, 54-66. <https://doi.org/10.21832/9781800417199-007>
- Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 109-120. <https://doi.org/10.1017/s0261444819000326>
- Stoll, M. M., Bergamo, N., & Rossetti, K. G. (2021). Analyzing modes of assessment for children with autism spectrum disorder (ASD) using a culturally sensitive lens. *Advances in Neurodevelopmental Disorders*, 5(3), 233-244. <https://doi.org/10.1007/s41252-021-00210-0>
- Tabone, J. K., Rishel, C. W., Hartnett, H. P., & Szafran, K. F. (2020). Examining the effectiveness of early intervention to create trauma-informed school environments. *Children and Youth Services Review*, 113, 104998. <https://doi.org/10.1016/j.childyouth.2020.104998>
- Wu, H., Liu, Q., Yu, B., Zhang, Y., Ren, D., Serdarevic, M., Liang, Z., Wang, Y., Chen, S., Zhang, K., & Chen, S. (2020). Psychometric properties of the Mandarin clinical evaluation of language for preschooler's core scale. *Journal of Communication Disorders*, 87, 106000. <https://doi.org/10.1016/j.jcomdis.2020.106000>
- Zimmerman, I. L., & Castilleja, N. F. (2005). The role of a language scale for infant and preschool assessment. *Mental Retardation and Developmental Disabilities Research Reviews*, 11(3), 238-246. <https://doi.org/10.1002/mrdd.20078>