Evaluation of Linguistic and Cultural Biases of the Assessment of Comprehension and Expression Tool (ACE 6-11) On Typically Developing Multilingual Kenyan Children in Nairobi City County, Kenya

***[1]Laetecia Kondiek., [2]Dr. Mathew Karia**

[1]Department of Early Childhood and Special Needs Education (Speech and Language Pathology), Kenyatta University - Nairobi, Kenya

[2]Department of Special Surgeries, Kenyatta University - Nairobi, Kenya

DOI: https://doi.org/10.47772/IJRISS.2026.101000487

Received: 21 October 2026; Accepted: 26 October 2026; Published: 05 November 2026

ABSTRACT

Standardized language assessments are commonly developed within particular linguistic and cultural populations. Yet, they may also be used with multilingual children whose language experiences differ from those represented in the normative sample. The Assessment of Comprehension and Expression 6-11 is a standardized, English-language assessment for school-aged children. Evidence concerning its performance among multilingual Kenyan children remains limited. This study examined the performance of typically developing multilingual Kenyan children on the ACE 6-11 relative to the published normative benchmark and potential linguistic and cultural biases of the tool. The study was grounded in Bronfenbrenner's Ecological Systems Theory of Human Development. The research design was convergent parallel mixed methods. The 60 multilingual children were selected purposively in three schools in Nairobi City County. Descriptive and inferential statistics were used to sum up quantitative data, while qualitative observations made during the test administration were analyzed thematically. Overall, the composite standard-score mean for this sample was found to be 96.78 (SD = 22.59) which fell within the ACE 6-11 Average range (85-115). The large standard deviation (SD) and score range (65-138), however, suggested massive individual variation. Inferential analysis revealed statistically significant differences in scores across the three school groups (Kruskal-Wallis, $H = 21.45$, $p < .001$), with pairwise comparisons indicating significant differences between groups. Qualitative observations revealed that there were meaningful responses in Kiswahili and mixed language, which the English-only scoring criteria failed to credit. There were also notable difficulties with lexical concepts, unfamiliar animals, objects, occupations and narrative contexts which generated alternative responses. The study concluded that although the sample's overall mean was within the published normative range, it should not be taken as evidence that the ACE 6-11 is culturally or linguistically equivalent for multilingual Kenyan children. The convergence of score patterns and observed response behavior suggested that lexical familiarity, language of response, and cultural-context familiarity may have contributed to the obtained scores. Multilingual assessment in Kenya should therefore combine standardized scores with detailed language history, contextual and qualitative information.

Keywords: ACE 6-11, multilingual children, typically developing, linguistic bias, cultural bias, normative benchmark.

INTRODUCTION

Language assessment is important in the identification of children who may require speech-language intervention and in distinguishing developmental language difficulties from differences associated with language exposure. Standardized assessments are valuable because they provide structured procedures and normative reference points. However, the validity and clinical usefulness of standardized language assessments depend on the quality of their psychometric evidence and how well they fit the population being assessed (Denman et al., 2017).

The Assessment of Comprehension and Expression 6-11 is a standardized assessment of spoken language for children aged 6-11 years (Adams et al., 2001). Multilingual assessment presents a particular challenge because a child may possess a concept but access its lexical label in another language (McLeod et al., 2017). Similarly, performance on a syntactic or comprehension task can reflect both the underlying language construct and the child's familiarity with the language in which the task is presented. Cross-cultural assessment research asserts that linguistic equivalence alone is insufficient, urging consideration of functional and cultural equivalence. In addition, bilingual assessment research emphasizes consideration of children's development and language experience across all their languages rather than treating performance in only one language as a complete representation of their linguistic knowledge (Paradis et al., 2021).

This study was motivated by the need to examine the ACE 6-11 in the Kenyan multilingual context. The study addressed two questions. That is, how do typically developing multilingual Kenyan children's scores compare numerically with the ACE 6-11 normative benchmark? Second, what linguistic and cultural features become visible when children's actual responses and behaviors during administration are considered alongside their scores?

Study objectives

The study addressed the following objectives:

1. To compare the performance of typically developing multilingual Kenyan children aged 6-11 years on the ACE 6-11 with the published normative benchmark.

2. To identify potential linguistic and cultural biases of the ACE 6-11 tool that may influence the performance of multilingual Kenyan children.

METHODS

This study adopted a convergent parallel mixed methods research design. This design was suitable as it addressed both the measurable test performance and the contextual features of the participant's responses (Creswell & Clark, 2018). The sample included 60 typically developing multilingual Kenyan children between 6-11 years from 3 schools in Nairobi City County. The three schools represented different educational and socioeconomic contexts. School A was an international school representing a relatively high socioeconomic context, School B was a private school representing a middle socioeconomic context while School C was public primary school serving a comparatively low income community. Each school contributed 20 children for the study. The inclusion of these three settings provided an opportunity to examine variation in performance among children attending schools with different socioeconomic profiles. Additionally, the inclusion criteria included age, multilingual experience, together with evidence of typical development. The participants' multilingual experience in English and Kiswahili was established through teacher information. However, individual language proficiency, language dominance, age of acquisition and frequency of exposure to each language were not formally measured.

The primary data collection instrument was the ACE 6-11. The tool was administered in its original form without any modification or alteration. Only the five main subtests were examined. That is, Sentence Comprehension, Inferential Comprehension, Naming, Syntactic Formulation, and Semantic Decisions. The researcher conducted the tests as outlined and made observations of the children's responses and behavior during test administration. The observations were recorded on a structured observation checklist. The observation checklist featured indicators for clarification requests, apparent confusion, culturally unfamiliar content, difficulty with verbal expression, code switching, need for repetition/rephrasing, unfamiliarity with test materials, and engagement. Observation notes were analyzed thematically to identify common patterns in the responses that the numerical score could not account for.

Quantitative data were analyzed using descriptive and inferential statistics. Descriptive statistics including means, standard deviations, ranges, and frequencies were used to summarize participants' performance on the ACE 6-11 and describe patterns across the five subtests. The ACE 6-11 composite scores were compared to the

published average standard score range of 85-115, with 100 as the normative mean. The Kruskal-Wallis test was used to examine difference in standard scores among the three groups. Where statistically significant differences were identified, a pairwise Mann-Whitney U test with Bonferroni adjustments were used to determine the school groups that differed. The qualitative observations were analyzed thematically and merged with the quantitative results. Qualitative data from the structured observation checklist and field notes were analyzed using a coding framework developed to identify linguistic and cultural response patterns. The framework included unfamiliar English vocabulary, code-switching, alternative interpretations of assessment items, syntactic formulation difficulties, requests for clarification or repetition, and confusion or unrelated responses. Observations were categorized according to these codes, and recurring patterns were summarized to complement the quantitative findings.

Code	Category	Description
C1	Unfamiliar English vocabulary	Difficulty understanding or identifying English words used in assessment items.
C2	Code switching	Use of Kiswahili instead of, or alongside, English when responding
C3	Alternative interpretation	An interpretation of an item that differs from its intended meaning
C4	Syntactic formulation difficulties	Difficulty formulating a response using the English grammatical structure required by the task.
C5	Clarification or repetition	Requests for clarification or repetition, or observed difficulty that prompted the assessor to repeat or rephrase an item.
C6	Confusion or Unrelated responses	Observable uncertainty or responses that appear unrelated to the question asked.

Table 1 Coding Framework for Linguistic and Cultural Observations

The study was approved by the KU Ethics Committee, National Commission for Science, Technology and Innovation (NACOSTI) and the participating schools before data collection. Participation was voluntary, confidentiality was respected and schools were coded in the research records and reporting. This study does not contain any child-identifying information.

FINDINGS

The 60 children were distributed across the six age bands. That is, 6-11 years. Each school had 20 children participating in the study.

Age (years)	School A	School B	School C
6	3	4	3
7	4	4	3
8	3	4	4
9	3	4	3
10	4	2	3
11	3	2	4
Total	20	20	20

Table 2. Distribution of children by age and school

Normative Performance

Overall mean ACE 6-11 composite standard score was 96.78 (SD = 22.59) with scores ranging from 65 to 138. The mean was thus within the ACE 6-11 average standard-scores range (85-115). However, the score distribution was wide suggesting that the overall mean did not describe all children's performance equally.

Group	N	Mean SS	SD	Minimum	Maximum
School A	20	114.3	19.2	78	138
School B	20	95.05	17.29	65	135
School C	20	81.01	18.11	65	135
Overall	60	96.78	22.59	65	138

Table 3. ACE 6-11 composite standard-score distribution.

The mean for school A and B were within the average range while the mean for school C was below the average range. Notably, the overall mean fell within the normative range but occurred alongside a substantial individual variation as shown by the large standard deviation

The study further examined whether standard scores differed across the three school groups. A Kruskal-Wallis test was used to compare the groups, followed by Bonferroni-adjusted pairwise Mann-Whitney U tests.

Test statistic	Value
Kruskal-Wallis H	21.45
Degrees of freedom (df)	2
p-value	<.001
Decision	Statistically significant

Table 4 Kruskal-Wallis test results

The Kruskal-Wallis test revealed a statistically significant difference in the ACE 6-11 standard scores among the 3 groups.

$H(2) = 21.45$, $P < .001$

However, the test did not tell which specific groups differed. To put this into context, Pairwise Mann-Whitney U test with Bonferroni adjustment were conducted to identify the specific group differences.

School Comparison	U	Unadjusted p	Bonferroni-adjusted p
School C vs B	105.0	.01051	.03154
School C vs A	46.0	.00003	.00010
School B vs A	92.5	.00372	.01115

Table 5 Pairwise Mann-Whitney U test Comparison of scores across groups (Bonferroni Adjusted)

All pairwise comparisons were statistically significant as the Bonferroni-adjusted p values were below the (.05) significance level. These results show a pattern of higher standard scores with increasing SES.

Performance Across ACE 6-11 Subtests

ACE 6-11 Subtest	Maximum Raw Score	Mean Raw Score	SD	Scoring Below 50%
Sentence Comprehension	31	18.92	5.49	26.7%
Inferential Comprehension	14	5.37	2.35	63.3%
Naming	25	10.33	4.99	71.7%
Syntactic Formulation	20	5.45	3.95	86.7%
Semantic Decisions	20	7.32	3.69	66.7%

Table 6. Performance across the five ACE 6-11 subtests.

Sentence Comprehension

Sentence Comprehension produced the strongest mean performance among the five subtests, but specific items remained difficult. Observations indicated difficulty with items requiring children to distinguish habitual from ongoing actions in English. For example, where a sentence contrasted a habitual activity with an activity occurring at the moment, such as *'Jim plays cricket after school* and *Jim is playing tennis after school'* and subsequently asked to identify '*Jim's usual game after school',* some children selected the action represented most prominently in the picture rather than the answer determined by the grammatical tense or aspect. Spatial terms such as left and right also proved to be difficult for most of the children.

Inferential Comprehension

Inferential Comprehension was among the weaker domains. During the *burglary picture-and-story task,* some children described the story as difficult or confusing, requested repetition, paused for extended periods or produced responses that did not clearly connect with the presented narrative. The task required children to integrate verbal information, visual information, vocabulary and inferential reasoning.

Naming

Naming elicited some of the clearest multilingual lexical-access patterns. For example, most of the participants provided useful Kiswahili equivalents for English target words such as *nanasi* for *pineapple*, *shoka* for *axe* and *hoho* for *pepper.* In other test items their responses reflected different interpretations of the pictured stimuli such as identifying a *factory* as a *church*, an *armadillo* as a *crocodile*, a *judge* as a *lawyer* and a *dentist chair* or *setting* as a *barber*. These responses are crucial as they demonstrate that an incorrect English label can coexist with the correct recognition of the pictured object, social context or category. Where such responses were made the participant's were not credited because of the tool's English only scoring criteria.

Syntactic Formulation

The lowest scoring subtest was Syntactic Formulation with most scores ranging below 50%. In this subtest most children responded to the test questions entirely in Kiswahili or code-switched their responses. Though the responses were formulated meaningfully, they were not credited because only English productions were required. Therefore, the resulting score of this subtest reflected two demands that were occurring simultaneously. That is, the ability to produce the required syntactic structure and the ability to produce it in the correct language which is English. In multilingual populations these demands are impractical as multilingual children often use different languages flexibly to communicate their knowledge and understanding. Findings therefore suggest that

this subtest was sensitive to the test's language requirements and underestimated the participant's underlying language competence when expressive production was difficult.

Semantic Decisions

Most participants had difficulties with specific semantic concepts. In this domain the participants were required to identify a word that was nearly synonymous with the target word. Words such *meadow, maze,* and *monsoon and fair* had relatively high error rates. The high frequency of errors around the same words is an indication that that lexical unfamiliarity may have contributed to the poor performance. Thus a child requires vocabulary recognition before semantic reasoning demand can be expressed.

The number and percentage of children who demonstrated each observed response pattern were calculated using the total sample of 60 participants. The results are presented in the table below.

Code	Observed Response	Frequency (n)	Percentage (%)
C4	Challenges differentiating habitual from ongoing activities	56	93.3%
C2	Used nanasi for "pineapple"	28	46.7%
C2	Used shoka for "axe"	26	43.3%
C2	Used hoho for "pepper"	49	81.7%
C3	Interpreted "factory" as a church	51	85.0%
C3	Interpreted "judge" as a lawyer	53	88.3%
C3	Interpreted "armadillo" as a crocodile	59	98.3%
C3	Interpreted "dentist" as a barber	54	90.0%
C1	Demonstrated difficulty with unfamiliar English vocabulary with words like fair, monsoon, maze, meadow	57	95.0%
C4	Codeswitching, responding entirely in Kiswahili	52	86.7%
C5, C6	Paused, required repetition, or gave unrelated responses during the burglary story	47	78.3%

Table 6. Observed Linguistic and Cultural Response Patterns

As shown in the table above, the frequency of observed response patterns varied across the sample, with some responses occurring in nearly all participants. The counts represent the number of children who demonstrated each response pattern and each child was counted once per response category.

DISCUSSION

The overall ACE 6-11 composite standard score mean was (96.78) which was within the published average range of (85-115). Comparing the sample's performance to the normative band on a numerical basis, then the score was within the normative band. This is significant as it showed that using an English standardized assessment with a sample of multilingual Kenyan children does not automatically produce uniformly low scores.

Concurrently, the high standard deviation of (22.59) and the performance range of (65-138) indicated the need for careful interpretation of the overall mean. Group means are a measure of central tendencies but they do not indicate how close individual children are to the mean. In this sample scores were spread throughout the full

range of standard scores. Therefore, it is not possible to say that the children 'performed normally' or 'performed poorly' as the performance pattern was highly heterogeneous. Some participants performed significantly higher while others had significantly lower scores.

Further analysis into the scores showed statistically significant differences across the three groups. School A had the highest mean of standard score(114.30), school B (95.05) while school C had the lowest mean (82.00). Comparison between school C and B was significant (adjusted p = .03154) indicating that their score distributions differed. The difference between school C and A was also significant (adjusted p = .00010), as was with School B and A (.01115). From the findings, there is a clear pattern of ACE 6-11 standard scores increasing with socioeconomic status. This therefore indicates that there is a direct association between SES and performance on the ACE 6-11.

The subtest performance provided a deeper understanding of the normative comparison. A composite score combines several domains into a single score, thus concealing specific processes within individual subtests that were challenging, making it hard to determine what processes were difficult (McLeod et al., 2017). For instance, through the subtest analysis, we are able to see that Sentence Comprehension was relatively easy while Syntactic Formulation, Inferential Comprehension and Semantic Decisions were relatively difficult.

Syntactic formulation particularly generated low scores because the task required the participants to produce a language rather than just recognize it. In multilingual contexts selected languages for response directly impact expressive performance. An individual child might have the conceptual and grammatical means to produce a response, but it is easier for them to retrieve and/or construct it in a language they know, such as Kiswahili. If only responses produced in English are awarded, then the obtained score may underestimate the child's overall language abilities.

The findings from Semantic Decisions raised a concern from a receptive and lexical point of view. High error frequencies around *meadow, maze, fair,* and *monsoon* indicate that errors may have occurred because of the lack of familiarity with the words themselves, which may have affected the ability to show the required semantic relationship in the test item. Standardized language tests usually include vocabulary items, but the difference between vocabulary knowledge and semantic reasoning becomes notable when the vocabulary reflects experiences not evenly distributed across cultures (Villiers, 2004).

Observations from the naming subtest provided a clear example of multilingual lexical access. Some of the responses given, such as *nanasi* for *pineapple, shoka* for *axe,* and *hoho* for *pepper*, are not coincidental but significant lexical forms in Kiswahili. The assessment required participants to provide responses of those images in their English forms. In such instance the Kiswahili responses were marked as incorrect although the participants showed evidence of access to the referent and a lexical representation for it in another language.

This finding aligns with the bilingual language-development approaches that focus on children's language knowledge and use in all their languages and not just assuming that performance in one language is a representation of their entire linguistic abilities (Paradis et al., 2021).

Similarly, in Syntactic Formulation correct Kiswahili responses and code-switching instances were observed. However, these responses were not credited as only grammatically correct English productions were considered. The performance on this subtest created a distinction between demonstrating the target linguistic concept and demonstrating it through the language specified by the test, and for multilingual populations these are totally different tasks. This distinction is in line with cross-cultural assessment approaches that advocate for separate consideration of linguistic and functional equivalence (Peña, 2007).

Potential cultural and contextual influences were also identified. For instance, some children identified an *armadillo* as *a crocodile*, a *factory* as *church*, a *judge* as *lawyer,* and a *dentist setting* as *barber*. Usually, pictures can be ambiguous, children can make basic errors when naming a picture, and concepts can be known to them but get weakly retrieved. However, if similar patterns are observed in conjunction with other patterns of unfamiliarity, then they should be critically analyzed, as the assessment may involve children identifying objects, occupations, or settings that are less familiar in their everyday setting. Similarly, cross-cultural assessment

studies have cautioned that differences in familiarity and context may affect the equivalent functioning of assessments across populations (Johnson & Clark, 2008).

In Inferential Comprehension, the *burglary story* provided a different picture. The participants were required to integrate a narrative with visual information and there after prompted to draw an inference. Some of the observations made during this subtest were reports that the story was confusing, requests for repetitions and unrelated responses. This is an indication that contextual and linguistic processing demands may have interacted based on the responses provided. A culturally unfamiliar scenario does not automatically discredit an inference task but it directly increases the background knowledge and comprehension demands placed on a child. The validity of an inference task does not depend on its cultural familiarity, but can make the task more challenging for the child in terms of background knowledge and language comprehension. (McLeod et al., 2017).

Previous literature on multilingual assessment reinforces the idea that assessment adaptation requires attention to cultural appropriateness as well test translation. Additionally, according to Peña (2007), linguistic equivalence on its own does not establish functional or cultural equivalence. This is supported by studies in the Kenyan context by Abubakar et al. (2013) who reported challenges related to cultural relevance when assessment materials were directly translated to Kenyan languages without cultural considerations.

The convergence of the numerical and observational findings helped contextualize this study. The low Syntactic Formulation score was observed alongside the use of Kiswahili and code-switching instances. Difficulties with the naming activity occurred alongside meaningful local lexical substitutions and provision of alternative responses to test items. Challenges experienced with the Semantic Decisions subtest occurred alongside widespread failure on specific low-frequency English words. Inferential Comprehension difficulty occurred alongside reports of confusion with the narrative task and requests for constant repetition. Quantitative scores therefore helped to identify areas of low performance while the qualitative observations helped explain why a particular task may have been difficult.

Limitations

First, the sample comprised 60 children from three schools in Nairobi City County, limiting the generalizability of the findings to other Kenyan populations. Second, although participants were identified as multilingual in English and Kiswahili, individual language proficiency, language dominance and patterns of exposure were not formally assessed. Third, the school groups represented different educational and socioeconomic contexts, but household socioeconomic status was not directly measured. Therefore, differences in performance might not be attributed to socioeconomic status alone. In addition, the linguistic and cultural response patterns were documented through researcher observations and field notes, without independent second-rater coding. Meaning, the possibility of subjectivity in the classification and interpretation of observed responses cannot be excluded. The qualitative findings should therefore be interpreted cautiously as exploratory indications of potential linguistic and cultural influences on assessment performance.

Clinical Implication

The findings of this study emphasizes the need for caution when interpreting standardized language assessment scores in multilingual children. Low scores may partly reflect unfamiliarity with the language or cultural content of assessment items rather than an underlying language disorder. This could potentially result in false-positive diagnoses. Clinicians should therefore consider children's language exposure, developmental history and communication abilities across their languages, and avoid relying on ACE 6-11 scores alone when making diagnostic decisions.

CONCLUSION

The study found that children in this sample performed within the published average standard score range, with a mean of (96.78). However, there were significant variations. The group means did not fully represent the distribution of individual scores. Observational findings indicated several factors that might have influenced performance including English-only scoring criteria, multilingual lexical access, unfamiliar vocabulary, objects

and occupations as well as comprehension demands that were context-dependent. For multilingual children standardized assessment should be understood as an interplay between the child's linguistic abilities and the assessment's linguistic, cultural and contextual demands.

The findings highlight potential linguistic and cultural influences on ACE 6-11 performance among a sample of multilingual children in Nairobi. However, further research involving larger and more diverse samples, systematic assessment of language backgrounds and independent coding of qualitative observations is needed to establish the extent of these influences. Additionally, more research should be done on culturally responsive adaptation, development of local normative data and creation of local assessment tools that are sensitive to this population's linguistic abilities and cultural realities.

REFERENCES

1. Abubakar, A., Van de Vijver, F. J. R., Van Baar, A., Mbonye, M., & Newton, C. R. J. (2013). Challenges in diagnosing developmental language disorders in multilingual children: Evidence from Kenya.
2. Adams, C., Cooke, R. E., Crutchley, A., Hesketh, A., & Reeves, D. (2001). Assessment of Comprehension and Expression (6–11). NFER-Nelson/GL Assessment.
3. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.
4. 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
5. Developmental Science, 22(5), e12875. https://doi.org/10.1111/desc.12875
6. Girolamo, T., Ghali, S., Campos, I. & Ford, A. (2022). Interpretation and Use of Standardized Language Assessments for Diverse School-Age Individuals. Perspectives of the ASHA Special Interest Groups 7(4), pp. 981-994. https://doi.org/10.1044/2022_PERSP-21-00322
7. Johnson, C. J. & Clark, J. M. (2008). Children's picture naming difficulty and errors: Effects of age of acquisition, uncertainty, and name generality. Applied Psycholinguistics. https://doi.org/10.1017/S0142716408080300
8. Jullien, S. (2021). Screening for language and speech delay in children under five years. BMC Pediatrics, 21(Suppl 1), 362. https://doi.org/10.1186/s12887-021-02817-7
9. Kaindio, M. P. (2014). Integrating information communication technology skills in preschool education in Kenya (Vol. 5, No. 5). Mediterranean Journal of Social Sciences, 5(5), 89–96. https://doi.org/10.5901/mjss.2014.v5n5p89 studylib.net+11
10. Khoza-Shangase, K., & Mophosho, M. (2021). Language and culture in speech-language and hearing professions in South Africa: Re-imagining practice. South African Journal of Communication Disorders, 68(1), Article a793. https://doi.org/10.4102/sajcd.v68i1.793
11. KK Nair, V., Clark, G. T., Siyambalapitiya, S., & Reuterskiöld, C. (2023). Language intervention in bilingual children with developmental language disorder: A systematic review. In International Journal of Language and Communication Disorders (Vol. 58, Issue 2, pp. 576–600). John Wiley and Sons Inc. https://doi.org/10.1111/1460-6984.12803
12. Knauer, H. A., Kariger, P. K., Jakiela, P., Ozier, O., & Fernald, L. C. H. (2019). Multilingual assessment of early child development: Analyses from repeated observations of children in Kenya.
13. Kohnert, K. (2013). Language disorders in bilingual children and adults (2nd ed.). San Diego, CA: Plural
14. Korade, C., Nicoladis, E. & Charest, M. (2024). Monolingual and bilingual children's production of complex syntactic structures. Journal of Child Language 51(2), pp. 454-469. https://doi.org/10.1017/S0305000922000411
15. McLeod, S., Verdon, S. & Speech, T. I. (2017). Tutorial: Speech Assessment for Multilingual Children Who Do Not Speak the Same Language(s) as the Speech-Language Pathologist. American Journal of Speech-Language Pathology 26(3), pp. 691-708. https://doi.org/10.1044/2017_AJSLP-15-0161
16. ncbi.nlm.nih.gov+8ncbi.nlm.nih.gov+8pmc.ncbi.nlm.nih.gov+8
17. Paradis, J., Genesee, F., & Crago, M. B. (2021). Dual language development & disorders: A handbook on bilingualism and second language learning (3rd ed.). Brookes Publishing.

18. Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. Child Development, 78(4), 1255–1264. https://doi.org/10.1111/j.1467-8624.2007.01064.x
19. Villiers, J. G. (2004). Cultural and Linguistic Fairness in the Assessment of Semantics. Seminars in Speech and Language 25(1), pp. 73-90. https://doi.org/10.1055/s-2004-824827