

Epistemic Justice as a Framework for Analyzing Early Grade Literacy Assessment Outcomes: A Baseline–Endline Analysis of the Balanced Reading Program in Badung Regency

Ni Komang Dwi Eka Yuliati¹, I Gede Ratnaya², Luhung Achmad Perguna³, I Dewa Ayu Made Budhyani⁴, I Gusti Lanang Parwata⁵

¹²⁴⁵ Universitas Pendidikan Ganesha, Jl. Udayana No.11, Banjar Tegal, Singaraja, Kabupaten Buleleng, Bali 81116

³ University of Pécs, Pécs, Vasvári Pál u. 4, 7622 Hungary

Email: ekayuliati.ylai@gmail.com

Abstrak

Reading assessment in early grades is commonly understood as a technical skill. However, the ability to make meaning from texts is a social practice also shaped by the mother tongue, culture, and pupils' everyday experiences. This study aims to analyze changes in early-grade students' literacy achievements after the implementation of the Balanced Reading Program in Badung Regency. In addition, the study examines cultural representation in the reading assessment instrument and discusses it from the perspectives of epistemic justice and cultural familiarity. The research uses a mixed-methods approach that combines quantitative analysis of baseline–endline data with qualitative document analysis of the assessment instrument. Quantitative data were obtained from baseline and endline tests without a control group. First- to third-grade pupils sampled at baseline numbered 320 and at endline 312. The sample was spread across six elementary schools. A total of 189 traceable paired data were analyzed using paired t-tests, Wilcoxon tests, Cohen's d effect sizes, and rank-biserial correlations. The analysis shows statistically significant improvements in syllables ($p < .001$), nonword reading ($p = .006$), fluency ($p = .017$), and comprehension ($p = .018$), but with small to moderate effect sizes, so their practical significance should be interpreted proportionally. Letter-sound knowledge ($p = .713$) and listening ($p = 1.000$) tended to remain stagnant. Instrument review shows that, in general, the assessment texts are close to children's world. However, the texts have not systematically integrated local languages, home literacy practices, and the socio-cultural landscape of Bali. These findings indicate potential problems of epistemic injustice in the design of reading assessments. Students may be evaluated using texts that do not necessarily provide equal opportunities to construct meaning. As conceptual implications and an agenda for further research, this study recommends developing reading assessments that are more culturally responsive.

Keywords: Reading assessment; Epistemic justice; Cultural familiarity; Early literacy; Sociology of education

Keadilan Epistemik sebagai Kerangka untuk Menganalisis Hasil Asesmen Literasi Kelas Awal: Analisis Baseline–Endline Program Membaca Berimbang di Kabupaten Badung

Abstract

Asesmen membaca di kelas awal umumnya dipahami sebagai keterampilan teknis. Namun, kemampuan membangun makna dari teks merupakan praktik sosial yang juga dibentuk oleh bahasa ibu, budaya, dan pengalaman sehari-hari peserta didik. Penelitian ini bertujuan menganalisis perubahan capaian literasi peserta didik kelas awal setelah pelaksanaan Program Membaca Berimbang di Kabupaten Badung. Selain itu, penelitian ini mengkaji representasi budaya dalam instrumen asesmen membaca dan mendiskusikannya dari perspektif keadilan epistemik dan kedekatan budaya. Penelitian ini menggunakan pendekatan metode campuran yang menggabungkan analisis kuantitatif data baseline–endline dengan analisis dokumen kualitatif terhadap instrumen asesmen. Data kuantitatif diperoleh dari tes baseline dan endline tanpa kelompok kontrol. Peserta didik kelas satu hingga tiga yang menjadi sampel pada baseline berjumlah 320 orang dan pada endline 312 orang. Sampel tersebar di enam sekolah dasar. Sebanyak 189 data berpasangan yang dapat dilacak dianalisis menggunakan uji-t berpasangan, uji Wilcoxon, ukuran efek Cohen's d, dan korelasi rank-biserial. Analisis menunjukkan peningkatan yang signifikan secara statistik pada suku kata ($p < .001$), membaca nonkata ($p = .006$), kelancaran membaca ($p = .017$), dan pemahaman ($p = .018$), tetapi dengan ukuran efek kecil hingga sedang sehingga signifikansi praktisnya perlu ditafsirkan secara proporsional. Pengetahuan bunyi huruf ($p = .713$) dan menyimak ($p = 1.000$) cenderung stagnan. Telaah instrumen menunjukkan bahwa secara umum teks asesmen dekat dengan dunia anak. Namun, teks-teks tersebut belum secara sistematis mengintegrasikan bahasa lokal, praktik literasi di rumah, dan lanskap sosial-budaya Bali. Temuan ini menunjukkan potensi persoalan ketidakadilan epistemik dalam desain asesmen membaca. Peserta didik dapat dinilai menggunakan teks yang belum tentu memberikan kesempatan yang setara untuk membangun makna. Sebagai implikasi konseptual dan agenda penelitian lanjutan, penelitian ini merekomendasikan pengembangan asesmen membaca yang lebih responsif secara budaya

Kata Kunci: Asesmen Membaca; Keadilan Epistemik; Kedekatan Budaya; Literasi Kelas Awal; Sosiologi Pendidikan

INTRODUCTION

Reading ability in the early grades is a crucial foundation for later schooling (Suggate et al., 2018). At this stage, pupils begin to recognize written symbols, connect letters with sounds, read words fluently, and build meaning from simple texts (National Reading Panel, 2000). In the lens of the sociology of education, however, reading is more than a neutral technical and cognitive skill. Reading is a social practice closely related to culture (Heath, 1983; Street, 1984). Children's literacy develops through interaction with the use of their mother tongue, family stories, and community practices. The ability to make meaning from texts is always bound up with pupils' cultural and social experiences (Gay, 2018; Paris & Alim, 2017). Assessments that ignore these social dimensions tend to measure children's proximity to the dominant language, so parts of pupils' ability to make meaning from texts risk being only partially captured (Shohamy, 2022).

In Indonesia, strengthening foundational literacy remains an important agenda. Baseline and endline reports show that many early-grade classrooms have not yet reached minimum competency (Yayasan Literasi Anak Indonesia, 2025). In the context of this study, the intervention in place is the Balanced Reading Program, implemented by Yayasan Literasi Anak Indonesia (YLAI) in collaboration with the Badung Regency Government, Bali. The program aims to strengthen literacy foundations through the provision of quality books, teacher training, structured programming, and ongoing coaching. It was evaluated using a baseline–endline design in six primary schools. Badung Regency was selected because it is a major program site and

represents the Balinese context, which is rich in local languages and cultural practices. Badung is thus relevant as a context for examining the extent to which reading assessment instruments accommodate pupils' linguistic and cultural diversity. The central issue is that a literacy program that is technically effective does not automatically become socially just if the same instrument is used to measure the ability to make meaning—a process deeply rooted in pupils' social backgrounds.

Comprehension cannot be separated from the use of the mother tongue, prior knowledge, and socio-cultural experiences that shape how pupils construct meaning. As highlighted in various studies (Gay, 2018; Scarborough, 2001; Snow, 2002), the concept of cultural familiarity in this study is defined as the degree of closeness between the content of texts, characters, situations, settings, and language used in the assessment and pupils' everyday lives. The higher the cultural familiarity of a text, the greater the likelihood that children can access meaning without having to infer unfamiliar contexts far removed from their lived experience (Paris & Alim, 2017). Conversely, if an assessment ignores cultural familiarity, its outcomes risk widening existing social inequalities and failing to measure pupils' authentic reading abilities fully (Walker et al., 2023).

Epistemic justice in this study draws on (Fricker, 2007) framework, which distinguishes two forms of epistemic injustice: testimonial injustice and hermeneutical injustice. Testimonial injustice occurs when a person's contribution as a knower is undervalued because of prejudice toward their identity. Hermeneutical injustice occurs when a person is disadvantaged because the available collective interpretive resources are inadequate to capture their experience. In the context of reading assessment, this framework becomes relevant. If texts and questions are designed from the perspective of a dominant culture, pupils' responses shaped by different cultural experiences may be judged as less credible (testimonial injustice). Meanwhile, the absence of local cultural representation in the instrument limits the interpretive resources available to children for constructing understanding (hermeneutical injustice). This prompts reflection on whose experiences are normalized in texts, whose experiences are taken as benchmarks, and who is structurally disadvantaged by instrument designs that neglect pupils' social constructions.

What makes the sociology of education perspective particularly relevant for assessing literacy and reading is that these practices occur within social relations. The development of teaching media and learning strategies based on local culture has proven effective in internalizing community values and strengthening cultural capital (Widianto & Lutfiana, 2021). Learning designs that connect content with pupils' lives and experiences can reinforce social meaning and make learning more relevant (Widianto et al., 2020). During the pandemic, the success of learning transformations was strongly influenced by socio-economic disparities and pupils' cultural access (Firmansyah et al., 2021). Reading literacy attainment data through a sociology of education lens thus becomes a way to situate pupils' reading achievements within networks of language, family, and community. Literacy attainment is not merely an individual achievement separate from its social order.

Viewed from this perspective, YLAI's baseline–endline findings show an interesting pattern. Dimensions related to accuracy, such as decoding and reading fluency, are relatively strong, while reading comprehension grows more slowly and some basic skills, such as letter-sound and listening, tend to stagnate (Yayasan Literasi Anak Indonesia, 2026). This gap raises the question of whether the reading assessment instrument used is adequate to capture the process of meaning-making. Initial instrument review suggests that some items ask about the language used at home, preferred stories, and family storytelling practices. However, this information is not yet systematically connected to comprehension scoring or text design.

Based on this background, the study has two aims. First, to analyze changes in early grade pupils' reading literacy attainment after the implementation of the Balanced Reading Program in Badung Regency. Second, to examine cultural representation in the reading assessment instrument and discuss it through the perspectives of epistemic justice and cultural familiarity. In doing so, the study formulates conceptual implications for the development of more culturally responsive early-grade reading assessments. Epistemic justice and cultural familiarity are positioned as analytical lenses for interpreting patterns of attainment and instrument representation, rather than as variables that are directly measured and tested.

METHOD

This study uses a mixed-methods approach with an explanatory sequential design. Baseline and endline assessments without a control group were used to measure changes in literacy scores before and after program implementation. The qualitative phase consisted of document analysis of the assessment instrument to interpret the representation of language and culture in the texts. The quantitative and qualitative stages complement each other. Quantitative data depict patterns of attainment change, while qualitative documents help interpret these patterns through the lens of epistemic justice and cultural familiarity. The quantitative results are interpreted as changes in literacy outcomes after implementation, not as causal evidence of program effectiveness, because no control group was used (Yayasan Literasi Anak Indonesia, 2026).

The study was conducted in six primary schools in Badung Regency, Bali, namely SDN 3 Werdi Bhuwana, SDN 5 Carangsari, SDN 1 Bongkasa, SDN 5 Pelaga, SDN 6 Ungasan, and SDN 2 Benoa. The sample comprised pupils in Grades 1, 2, and 3 who participated in both baseline and endline assessments. At baseline, 320 pupils were assessed, and at endline, 312 pupils. There was a decline of about 3–4% (attrition) due to pupil absence at one of the data collection points, for reasons such as illness, leave, or transferring schools. Pupils included in the analytical sample were those officially enrolled in Grades 1–3, present at both assessment points, and able to complete at least 50% of the subtasks. Inferential analysis was conducted on 189 pupil data pairs traceable from baseline to endline (Yayasan Literasi Anak Indonesia, 2026). Because only complete cases were analyzed, there is a possibility of attrition bias: pupils absent at one of the assessments may have different attainment characteristics than those with complete data. Therefore, findings are generalized cautiously and interpreted as patterns among pupils whose data could be traced, not the entire program population.

Quantitative data were obtained from YLAI's Reading Literacy Test, developed by adapting the Early Grade Reading Assessment (EGRA) framework (Dubeck & Gove, 2015; RTI International, 2009) to early grade literacy instruction in Indonesia. The instrument consists of ten subtasks, including letter recognition, letter-sound knowledge, syllable reading, meaningful word reading, nonword reading, listening, reading fluency (34-word and 78-word texts), and comprehension. Fluency scores are expressed as correct words per minute, while other subtasks are expressed as the number of correct responses. For interpretive purposes, letter recognition, letter-sound, syllable reading, meaningful word reading, and nonword reading are grouped as phonics and decoding components, while fluency and comprehension are treated as higher-level reading skills. This grouping is used as a framework for reading the results in line with early literacy skill development. The interval between baseline and endline was approximately three months. Baseline took place on 10–12 November 2025 before full implementation, while endline took place between 16–18 December 2025 and January 2026. Each pupil was assessed individually by trained enumerators for about 15–20 minutes, with field supervision and spot checks to ensure data quality.

Quantitative data analysis was conducted in two stages. Descriptive analysis summarized baseline and endline attainment for each subtask using means, medians, and standard deviations. Inferential analysis tested differences in scores between baseline and endline for paired pupil data. Before difference testing, Shapiro–Wilk normality tests were applied to score differences. If differences were normally distributed, paired t-tests were used; if normality assumptions were not met, Wilcoxon signed-rank tests were used as a non-parametric alternative. Effect sizes were reported as Cohen's *d* for parametric analyses and rank-biserial correlations for non-parametric analyses, using conventional thresholds of approximately 0.2 (small), 0.5 (medium), and 0.8 (large) (Cohen, 1988). All statistical analyses were performed using Jamovi version 2.7.

The qualitative phase consisted of document analysis of the Individual Pupil Assessment Sheet, carried out by the researcher following traceable procedures. The units of analysis were all reading texts, listening texts, comprehension questions, and items related to reading habits embedded in the instrument. Analysis used a categorization framework derived from the concepts of cultural familiarity and epistemic justice, operationalized through three indicators of cultural representation: (1) the closeness of story settings and situations to pupils' everyday lives (home, school, community, customary practices, local games, and food); (2) the use and recognition of

local languages or the language used by pupils at home; and (3) the connection between information on pupils' reading habits and the design of texts or the interpretation of scores. Coding procedures involved reading each unit of analysis, assigning codes for the presence of the three indicators (present, generically present, or absent), and summarizing cross-unit patterns to assess the degree to which the instrument integrated cultural familiarity. Because the document analysis was descriptive and conducted by a single reviewer, its findings are positioned as descriptive qualitative analysis of cultural representation in the instrument, not as an empirical test of the impact of cultural familiarity on pupils' reading attainment.

Interpretation in this article is constrained by the nature of the quantitative data, which are secondary and collected through a design without a control group, and by the descriptive nature of the document analysis. These constraints arise because the study utilizes existing program evaluation data and focuses on re-reading these data through the lens of epistemic justice rather than designing a new experiment. The study did not conduct in-depth interviews on pupils' cultural experiences while taking the test; hence, the phrase "changes after implementation" is understood as programmatic patterns, not definitive causal evidence. Claims regarding epistemic justice are thus positioned as theoretical interpretations of data patterns and instrument representation.

RESULT AND DISCUSSION

Descriptive analysis shows mean score increases in most foundational literacy subtasks between baseline and endline, though the magnitude and statistical significance vary. Table 1 summarizes the difference test results for the 189 data pairs.

Table 1. Difference test results for baseline–endline by subtask (N = 189)

Subtask	Baseline Mean	Endline Mean	Test	p	Effect Size
Letter sounds	1,70	1,58	$t = 0,368$	0,713	not significant
Syllables	9,04	9,43	$t = -3,53$; Wilcoxon	<0,001	d -0,257; rb -0,697
Meaningful words	8,87	9,06	$t = -1,21$; Wilcoxon	0,229	d -0,088 (n.s.)
Nonwords	—	—	$t = -2,76$	0,006	Cohen's d -0,201
Fluency	2,21	2,35	$t = -2,41$	0,017	Cohen's d -0,176
Listening	2,43	2,43	$t = 0,00$	1,000	Cohen's d 0,00
Comprehension	2,83	3,12	$t = -2,39$	0,018	Cohen's d -0,174

Source: processed from the YLAI Baseline–Endline Report (2026)

Overall, results show statistically significant improvements in syllable reading, nonword reading, fluency, and comprehension. However, effect sizes are mostly small (Cohen's d around 0.17–0.26; rank-biserial correlations up to about 0.70 for syllable reading), so claims about change must be framed proportionally. The data more strongly indicate early improvement in decoding skills and reading fluency than substantial changes in comprehension. Accordingly, the discussion focuses on the pattern of development across early literacy components (decoding–fluency–comprehension), and sociological interpretations are positioned as lenses for framing these patterns, not as standalone empirical claims.

Changes in phonics and decoding foundations

The most consistent changes appear in phonics and decoding components, especially syllable reading and nonword reading. For syllables, mean scores increased from 9.04 to 9.43 ($t = -3.53$; $p < 0.001$; Cohen's $d = -0.257$), and because Shapiro–Wilk tests indicated violations of normality, interpretation relies on the Wilcoxon test, which is also significant ($p < 0.001$) with a rank-biserial correlation of -0.697 , suggesting a medium-to-large effect size (Yayasan Literasi Anak Indonesia, 2026). In nonword reading, the mean increased by 0.49 points ($t = -2.76$; $p =$

0.006; Cohen's $d = -0.201$). Practically, these gains suggest that pupils became more able to process sound units longer than single letters and began to activate letter–sound rules when encountering unfamiliar letter strings, rather than merely recognizing memorized word forms. Nonetheless, given the predominantly small effect sizes, these changes are better interpreted as indications of developmental direction than as major leaps in attainment.

By contrast, meaningful word reading only shows a small, non-significant increase (8.87 to 9.06; $p = 0.229$; Cohen's $d = -0.088$). This pattern can be interpreted sociologically: meaningful words typically come from vocabulary circulating in children's everyday literacy practices, so scores are relatively high at baseline and room for improvement is limited. In other words, change is more evident in demanding phonological decoding tasks than in recognizing already familiar words from children's social environments (Widianto & Lutfiana, 2021).

Stagnant dimensions: letter-sound and listening

Letter-sound scores did not improve and even slightly decreased from 1.70 to 1.58 ($t = 0.368$; $p = 0.713$), while listening scores remained unchanged at 2.43 ($p = 1.000$; Cohen's $d = 0.00$). These two dimensions are the most stagnant. From a sociology of education perspective, this stagnation cannot be read merely as a weakness in phonics or oral comprehension. Letter-sound knowledge and listening are dimensions closest to oral language and home literacy practices; children's interactions with sound, symbols, and oral stories are shaped by family and community habits that do not always align with the assumptions of formal instruments in standard Indonesian. Stagnant listening scores especially suggest that the foundation of oral comprehension—an arena where children can practice constructing meaning without the burden of decoding—has not yet been fully cultivated, possibly because listening and meaningful discussion activities have not received sufficient or consistent emphasis across classrooms. This aligns with literature emphasizing the importance of connecting learning materials with learners' lived experiences to strengthen relevance (Widianto et al., 2020).

Fluency and comprehension: significant but small effects

Reading fluency increased significantly from 2.21 to 2.35 ($t = -2.41$; $p = 0.017$; Cohen's $d = -0.176$), and comprehension increased from 2.83 to 3.12 ($t = -2.39$; $p = 0.018$; Cohen's $d = -0.174$), both with small effect sizes. Practically, this indicates slight average gains distributed thinly across the group: most pupils read somewhat faster and answered comprehension questions slightly better, but the magnitude of change remains modest. Paying attention to practical significance helps avoid overinterpreting statistical significance as a major substantive improvement. The relatively slower growth of comprehension compared to decoding is consistent with the literature that situates comprehension as a skill built on decoding and fluency, requiring time and more complex instructional strategies (Snow, 2002).

This decoding–fluency–comprehension pattern is the clearest quantitative finding of the study. The sociology of education lens is then used not to prove but to interpret why comprehension grows more slowly. It is important to underscore that the available quantitative data do not directly measure text cultural familiarity or pupils' cultural experiences; hence, the link between text cultural characteristics and reading scores is not empirically tested but proposed as an interpretive hypothesis for future studies.

Cultural Representation in the Instrument

Document review shows that the assessment texts already attempt to use situations close to children, such as birthday presents, going home from school, playing with dolls, or being helped by adults (Yayasan Literasi Anak Indonesia, 2026). However, based on the three indicators of cultural representation described in the Method section, the cultural representation is still largely generic and does not explicitly incorporate the local socio-cultural landscape of Badung and Bali, such as local family practices, traditional games, customary activities, or relational patterns more familiar to children. In other words, the reading contexts are “generally close” but not necessarily “culturally familiar” for all pupils. Theoretically, this may engender hermeneutical injustice in (Fricker, 2007) sense: when the interpretive framework embedded in texts does not provide relevant cultural resources, children who can meaningfully interpret stories from their own worlds

may appear less competent when characters, situations, or question styles do not align with their cultural schemas.

The instrument actually holds potential to move toward greater justice. In the reading habits section, children are asked which language is most often used during reading or storytelling, whether local languages are used, what kinds of stories they enjoy, who reads to them, and how they share stories with family or friends (Yayasan Literasi Anak Indonesia, 2026). These questions reveal instrument designers' awareness of variations in home literacy cultural capital. Unfortunately, the third indicator—linking reading habit information with text design and score interpretation—has not yet been fulfilled: this information remains disconnected from text design, comprehension scoring, or reader profile grouping. If data on home language and story preferences were integrated into text development and results analysis, assessments could conceptually evolve from merely measuring outcomes to understanding why a child succeeds or struggles in meaning-making.

Sociological Implications: From Technical Measurement to Epistemic Justice

Overall, over roughly three months, patterns of results show the clearest changes in mid-level phonics components (syllable and nonword reading) and early shifts in fluency and comprehension with small effect sizes, while letter-sound and listening lag behind. From a sociology of education perspective, the gap between mechanical reading and comprehension can be read as a reflection of the relationship between school culture and home culture. Decoding is mostly taught in school and relatively independent of specific cultural capital, so it responds quickly to intervention. In contrast, comprehension requires pupils to connect texts with background knowledge and social experiences, making it more sensitive to the alignment between the world of texts and children's worlds. This interpretation is offered as a theoretical reading, with full awareness that the study did not provide direct empirical evidence on the impact of text cultural familiarity on comprehension scores.

If reading assessment texts predominantly represent the dominant culture, comprehension scores risk reproducing inequality: children from linguistic and cultural backgrounds farther from the texts may be rated low not because they lack reading ability, but because they are less familiar with the context being measured. This is the core of the epistemic justice concern raised in the study. Culturally responsive reading assessments do not mean lowering standards; rather, they ensure that what is measured is genuinely the ability to read and make meaning, not the ability to guess unfamiliar contexts (Gay, 2018; Paris & Alim, 2017). The development of learning media and strategies grounded in local wisdom and context has proven effective in enhancing meaningful learning and strengthening community cultural capital (Widianto et al., 2020; Widianto & Lutfiana, 2021) and similar principles can conceptually be applied to assessment design.

Based on this reading, more culturally responsive early-grade reading assessments could consider at least four principles. First, comprehension texts should be developed from domains of experience genuinely present in pupils' lives, including home environments, communities, relevant customary and religious practices, local games, food, and social relations. Second, text language should bridge standard Indonesian used at school and the language children use at home, without compromising the goal of mastering the standard language. Third, comprehension questions should not only assess literal recall but also allow inferencing grounded in pupils' experiences. Fourth, information on reading habits, home language, and story preferences should be treated as explanatory variables in result analysis, not merely as background data (Yayasan Literasi Anak Indonesia, 2026). These four principles are offered as conceptual implications and development agendas rather than as conclusions already empirically tested in this study.

CONCLUSION

Based on analysis of YLAI's baseline–endline data and review of the reading assessment instrument, the study concludes that following the implementation of the Balanced Reading Program, there were statistically significant but small-to-moderate changes in mid-level decoding, fluency, and comprehension, while letter-sound and listening remained relatively stagnant. Patterns of early literacy attainment show greater improvement in decoding and fluency than in comprehension. Viewed through the sociology of education and epistemic justice frameworks, this pattern can be interpreted not merely as a technical progression issue but as an indication of potential epistemic injustice: the instrument currently presents children's life contexts in generic

ways and has not systematically integrated cultural familiarity, home language, and local literacy practices into text design and score interpretation. It is important to underline that conclusions regarding epistemic justice are theoretical interpretations of data patterns and instrument representation, not empirical findings derived from direct measurements of cultural familiarity or pupils' experiences.

The study proposes three main implications, positioned as conceptual contributions and agendas for further research. For program developers and instrument designers, it is necessary to build a bank of comprehension texts layered by degrees of cultural familiarity and linguistic difficulty, empirically tested with diverse pupil groups, accompanied by explicit strengthening of letter-sound instruction and meaningful listening activities. For early grade teachers, assessment results should be used not only to identify pupils who are technically behind but also to examine whether comprehension difficulties relate to unfamiliar texts. For future researchers, studies are needed that directly test the impact of differing levels of text cultural familiarity on comprehension scores—through experimental or quasi-experimental designs with comparison groups—incorporating home language, family reading experiences, and school location as analytical factors. In this way, early grade reading assessments can evolve into instruments that are not only psychometrically sound but also socially just.

REFERENCE

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Dubeck, M. M., & Gove, A. (2015). The Early Grade Reading Assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, 40, 315–322. <https://doi.org/10.1016/j.ijedudev.2014.11.004>
- Firmansyah, R., Putri, D. M., Wicaksono, M. G. S., Putri, S. F., Widiyanto, A. A., & Palil, M. R. (2021). Educational transformation: An evaluation of online learning due to COVID-19. *International Journal of Emerging Technologies in Learning (iJET)*, 16(7), 61–76. <https://doi.org/10.3991/ijet.v16i07.21201>
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Gay, G. (2018). *Culturally Responsive Teaching: Theory, Research, and Practice* (3rd ed.). Teachers College Press.
- Heath, S. B. (1983). *Ways with Words: Language, Life, and Work in Communities and Classrooms*. Cambridge University Press.
- National Reading Panel. (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. National Institute of Child Health and Human Development.
- Paris, D., & Alim, H. S. (2017). *Culturally Sustaining Pedagogies: Teaching and Learning for Justice in a Changing World*. Teachers College Press.
- RTI International. (2009). *Early Grade Reading Assessment Toolkit*. RTI International.
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading disabilities: Evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of Early Literacy Research* (pp. 97–110). Guilford Press.
- Snow, C. E. (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. RAND.
- Shohamy, E. (2022). Critical Language Testing, Multilingualism and Social Justice. *TESOL J*, 56: 1445–1457. <https://doi.org/10.1002/tesq.3185>
- Street, B. V. (1984). *Literacy in Theory and Practice*. Cambridge University Press.
- Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading

- comprehension. *Cognitive Development*, 47, 82-95.
<https://doi.org/10.1016/j.cogdev.2018.04.005>
- Walker, M.E., Olivera-Aguilar, M., Lehman, B., Laitusis, C., Guzman-Orth, D. and Gholson, M. (2023), Culturally Responsive Assessment: Provisional Principles. ETS Research Report Series, 2023: 1-24. <https://doi.org/10.1002/ets2.12374>
- Widianto, A. A., & Lutfiana, R. F. (2021). Kearifan lokal Kabumi: Media internalisasi nilai-nilai karakter masyarakat Tuban Jawa Timur. *Satwika: Kajian Ilmu Budaya Dan Perubahan Sosial*, 5(1), 118–130. <https://doi.org/10.22219/satwika.v5i1.16277>
- Widianto, A. A., Purwasih, J. H. G., & Perguna, L. A. (2020). Promoting social cohesion: The development of e-learning management system materials through life based learning for sociology of religion course. *International Journal of Emerging Technologies in Learning (iJET)*, 15(7), 196–205. <https://doi.org/10.3991/ijet.v15i07.13228>
- Yayasan Literasi Anak Indonesia. (2025). *Lembar Penilaian Individu Siswa: Dokumen Instrumen Asesmen Membaca Kelas Awal YLAI* [Unpublished manuscript].
- Yayasan Literasi Anak Indonesia. (2026). *Laporan Baseline–Endline YLAI: Studi Evaluasi Baseline–Endline Periode September–Desember 2025 Program Membaca Berimbang Yayasan Literasi Anak Indonesia*. Yayasan Literasi Anak Indonesia.