

Deconstructing the Authenticity Crisis: A Critical Analysis of Written Text Bias in Listening Evaluation in Elementary School

Rahmat Iqbal¹, Martinez², Intan Safiah³, Rizki Kurniawati⁴,

^{1,2,3,4}Departement Pendidikan Guru Sekolah Dasar, Syiah Kuala University, Indonesia

¹rahmatiqbal@usk.ac.id, ²martines@usk.ac.id, ³intan.afia@usk.ac.id, ⁴rizkikurniawati@usk.ac.id

ARTICLE INFO

Article History:

Received : 02-03-2026

Revised : 20-06-2026

Accepted : 30-06-2026

Online : 10-07-2026

Keywords:

Listening Evaluation;

Text Authenticity;

Critical Analysis;

Elementary Education.



ABSTRACT

Listening skills assessments in elementary school are often biased because written texts fail to represent the characteristics of authentic oral communication, as the listening tests given are only rigid written texts. This condition is further exacerbated because teachers believe that students' listening skills do not need to be optimally evaluated as they have been continuously developing since birth. This condition will result in biased evaluation of students' listening skills, which is far from the characteristics of authentic assessment, ultimately leading to poor listening skills. This research aims to deconstruct the conditions of listening evaluation thru a critical analysis of the assessment instruments used in elementary schools. Using a qualitative approach with content analysis methods, this study examines three assessment documents (test questions, listening transcripts, and rubrics) at SD Negeri 15 Banda Aceh. The research findings indicate a dominance of rigid scripted speech, high lexical density, and the absence of natural oral linguistic features such as redundancy and pauses in the evaluation instrument. This condition creates a bias where listening tests become "read-aloud reading tests," which cognitively overloads students and has low construct validity. This research recommends reconstructing the evaluation model to shift from static text toward using auditorily designed scripts or dynamic digital media to ensure the authenticity of students' oral communication competence.



Crossref



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

A. INTRODUCTION

The Listening skills in elementary school are a complex cognitive process that requires understanding meaning thru auditory input (Iqbal, 2022). However, there is a wide gap between listening theory and evaluation practice in the classroom. Some researchers highlight that the main issue lies in the use of read-aloud written text, rather than using text that has the characteristics of authentic spoken language (Iqbal et al., 2023; Rossi & Brunfaut, 2021). In Indonesia, this phenomenon is exacerbated by Poerwanti (2012) finding that elementary school teachers still have limitations in designing authentic assessments, so listening evaluations often become mere repetitions of reading tests with the medium shifted to sound.

The use of authentic-sounding texts in listening evaluations in elementary school does not mean using recordings of complex, natural adult speech, but rather the use of texts intentionally designed to resemble oral characteristics. Referring to Rossi & Brunfaut (2021) this text should include planned oral-based features, such as redundancy, the use of strategic pauses, and intonation variation that provides meaning cues.

This is crucial considering the cognitive characteristics of elementary school students, who are in the concrete operational stage (Gathercole et al., 2004). From the perspective of Cognitive Load Theory (CLT), children have limited working memory capacity to process abstract information. Dense and rigid written text forces children's brains to expend excessive decoding effort (extraneous cognitive load), thereby hindering the meaning construction process (McNamara & Magliano, 2009; Paas et al., 2003). An evaluation that ignores this aspect no longer measures listening comprehension, but rather the students' cognitive resilience to an unreasonable information load.

At a macro level, this linguistic distortion reflects a systemic mismatch in the implementation of Indonesia's *Kurikulum Merdeka*. While the curriculum explicitly demands that primary language instruction pivot toward active communicative competency, structural assessment modules that differentiate real-time oral processing from reading mechanics are virtually non-existent for elementary school teachers. Consequently, educators nationwide continuously fall back on textbook-centric item banks, forcing listening tests to rely on rigid visual syntax rather than dynamic acoustic parsing.

At the institutional level, this pedagogical gap severely disrupts daily assessment practices at SD Negeri 15 Banda Aceh. Initial screenings of the school's test blueprints show that listening scripts are drawn directly from unadapted reading textbooks. For young learners who are still developing formal language literacy, being forced to parse highly complex, non-redundant written grammar under real-time testing conditions triggers immense cognitive fatigue, meaning the instrument inadvertently tests rapid reading recall rather than authentic functional listening skills.

The inconsistency of this instrument creates a validity crisis in measuring student competency. Written text-based evaluations fail to capture the prosodic elements, pauses, and redundancy that are characteristic of oral communication. Morales & Fernández (2019) emphasize that listening assessment should reflect how language is used in real-world contexts. Without textual authenticity, we are not measuring listening comprehension, but rather the students' ability to dissect dense syntactic structures. A skill that is closer to written literacy than oral literacy. This aligns with the view of Ajjawi et al., (2024) that authenticity in assessment should be able to represent real-world standards so that evaluation results are useful for students' communicative development.

Failure to present a representative evaluation impacts students' low motivation and understanding ability. Research by Demirbaş & Şahin (2022) proves that when elementary school students are given more dynamic and reality-based input, such as digital stories, their comprehension abilities significantly improve compared to static text. Therefore, this research is important because it explores the extent to which written text bias has distorted the results of listening comprehension evaluations in elementary schools.

To address this gap with high empirical precision, this study is driven by three specific, measurable research questions:

1. To what extent do the listening evaluation instruments at SD Negeri 15 Banda Aceh deviate from spoken language standards in terms of quantitative indices (lexical density, sentence length, and redundancy index)?
2. What are the empirical structural and linguistic disparities between the current classroom test scripts and their reconstructed authentic spoken versions across different text genres?
3. How does the resulting written text bias systematically alter the validity of the assessment and affect the cognitive load thresholds of elementary school students?

By answering these questions, the main focus of this research is not merely to identify errors, but rather to deconstruct existing evaluation patterns toward an assessment model that truly represents the characteristics of oral communication.

B. METHODS

This study employs a descriptive qualitative approach with content analysis methods to dissect the linguistic content and textual characteristics within the listening evaluation instruments for elementary school students. The primary focus of the research is to reveal the extent to which "written text bias" or literacy distortion occurs in assessments at SD Negeri 15 Banda Aceh. The research subjects consist of three assessment documents from SD Negeri 15 Banda Aceh, selected thru purposive sampling techniques. The analysis was conducted microscopically on 25 test items and their answer sheets to gain an in-depth understanding of the literacy bias phenomenon that occurs systematically in the field.

To ensure the sharpness of the analysis, this study uses the Authentic Text Listening Analysis Format (FAOTS) as an auxiliary instrument. This instrument synthesizes authentic spoken text criteria from Rossi & Brunfaut (2021) and the authentic assessment model from Morales & Fernández (2019). The FAOTS instrument dissects documents thru four critical dimensions: spoken linguistic features (redundancy and pauses), syntactic structure and lexical density, genre and situational authenticity, and cognitive suitability for elementary school children based on Demirbaş & Şahin (2022) standards. Researchers set strict cognitive thresholds, where a text is categorized as having "Written Text Bias" if its lexical density exceeds 55%, is dominated by complex sentences with more than 12 words, and has a minimal redundancy index. To ensure the validity of the FAOTS rubric, the instrument underwent an expert judgment validation process involving two applied linguistics experts specializing in language assessment. The validation focused on the clarity of descriptors, the relevance of item indicators, and the appropriateness of the cognitive thresholds, resulting in a Content Validity Index that confirmed the rubric's high readiness for data analysis.

The data collection process was carried out thru a study of documentation and verbatim transcripts of the original listening comprehension evaluation scripts. To maintain the trustworthiness and trustworthiness of the findings, data and instrument triangulation were rigorously applied. Data triangulation was achieved by cross-examining three distinct assessment documents across different thematic units, supported by confirmation from the verbatim oral transcripts. Meanwhile, instrument triangulation was conducted by juxtaposing the FAOTS rubric analysis with theoretical frameworks from multiple established benchmarks (Demirbaş & Şahin, 2022; Morales & Fernández, 2019; Rossi & Brunfaut, 2021). to prevent subjective analytical bias. Next, data analysis followed the Miles, Huberman, and Saldana model combined with text authenticity criteria thru four stages: data condensation to select instructions and content, data presentation in the form of a comparative table of scripted vs. authentic features, critical comparative analysis to dissect cognitive disparities, and conclusion drawing. Thru this procedure, the research aims to formulate the extent to which the evaluation instruments at SD 15 Banda Aceh successfully or unsuccessfully represent the authenticity of functional oral communication for elementary school students.

Researchers used the main supporting instrument, the Listening Text Authenticity Analysis Format, which was developed by the researchers based on a synthesis of Rossi & Brunfaut (2021) theory. This instrument was validated using the Listening Text Authenticity Analysis Rubric (RAOTS), which includes three critical dimensions:

1. Oral Linguistic Features: Measures the presence of redundancy, natural pauses, and interaction.
2. Lexical Density: Measures the ratio of content words to cognitive threshold (ideally 40-45%).
3. Cognitive Load: Measures syntactic complexity based on sentence length and compound structures.

Table 1. Rubric for Analyzing the Authenticity of Listening Texts

Assessment Aspect	Score 1 (High Written Text Bias)	Score 2 (Transitional)	Score 3 (Authentic/Ideal Spoken)
Spoken Linguistic Features	The text is rigid, with no natural pauses, no redundancy, and the language is one-directional (formal).	There is an attempt to provide simple pauses or repetitions, but the word choice (diction) is still textbook-like	Therefore, there will be redundancy, there are fillers or natural pauses, the use of discourse markers (Well, So, Yes), and there is pseudo-interaction.
Lexical Density	High (>55%) The ratio of content words as nouns or verbs is very high without adequate function words to allow for cognitive processing.	Dense (46-55%) Information begins to be broken down into multiple sentences, but there is still an accumulation of key information in close proximity.	Low (40-45%) Information is distributed gradually through chunking techniques that facilitate decoding.
Cognitive Load	High. Dominated by long (>12 words) complex sentences. Forcing students to do heavy grammatical analysis.	Currently. There is variation between complex and simple sentences. The average sentence length ranges from 9-12 words.	Low. Dominated by simple single sentences (<9 words) that go straight to the core information (to the point).

C. RESULT AND DISCUSSION

To provide empirical evidence of the linguistic distortion across the analysed documents, the researchers quantified the characteristics of each text using the Authentic Text Listening Analysis Format (FAOTS) rubric. The evaluation focused on three primary assessment documents from SD Negeri 15 Banda Aceh. The strict cognitive thresholds established in the methodology specify that a text is categorized as having "Written Text Bias" if its Lexical Density (LD) exceeds 55%, the Mean Length of Utterance (MLU) is dominated by complex sentences exceeding 12 words, and the Redundancy Index (RI) is minimal (scored 1 on a 1–4 scale).

Table 2. Quantification of FAOTS Rubric Scores Per Analyzed Text

Text Category	Lexical Density (%)	Mean Sentence Length (Words)	Redundancy Index (1–4)	Final Categorization
Narrative Text	61.5%	25 words	1.0 (Stiff/None)	Written Text Bias
Instruction Text	56.8%	15 words	1.2 (Minimal)	Written Text Bias
Announcement Text	63.2%	21 words	1.0 (Stiff/None)	Written Text Bias
Informative Text	59.4%	19 words	1.5 (Minimal)	Written Text Bias
News Text	66.8%	20 words	1.1 (Minimal)	Written Text Bias
Scientific Text	57.4%	13.8 words	1.0 (Stiff/None)	Written Text Bias

Analysis of the listening evaluation documents at SD Negeri 15 Banda Aceh reveals a total absence of natural spoken language features, with all scripts relying on rigid, scripted speech. Authentic listening texts must possess intrinsic oral features rather than being merely written scripts read aloud (Rossi & Brunfaut, 2021). By fully adopting written literacy structures without redundancy or fillers, these instruments fail to measure actual listening skills; instead, they test the students' ability to process written text through hearing. This omission of strategic pauses and repetitions eliminates the "breathing room" necessary for auditory processing, significantly escalating students' cognitive load.

This high lexical density is critical considering that children's working memory capacity specifically the central executive and phonological loop develops linearly from ages 4 to 15 and

is still undergoing functional expansion (Gathercole et al., 2004). Consequently, utilizing rigid and dense listening scripts biologically exceeds the temporary storage capacity of primary school students. As Bachman & Palmer (1996) warned, excessive processing demands that misalign with real-world language usage severely invalidate test scores, failing to reflect true language ability.

Therefore, the use of rigid and dense listening scripts not only triggers bias but also biologically exceeds the child's temporary storage capacity. As Bachman & Palmer (1996) warned, if the processing demand on a language test is too high and does not align with the characteristics of real-world language use, the test score will not be valid in reflecting true language ability.

Systematically, the analysed documents are dominated by the narrative-literacy genre, completely lacking functional genres like oral instructions, announcements, or problem-solving dialogues. Because assessment authenticity depends on how well a task represents real-world standards (Ajjawi et al., 2024), relying solely on monotonically read fairy tales fails to foster functional communicative competency. This practice trains students to be "passive recipients" of written literature rather than active responders to spoken social situations. Ultimately, this dominance of static text exposes a serious policy gap: while the *Kurikulum Merdeka* mandates interactive communication outcomes, the actual instruments at SD 15 Banda Aceh still measure passive literacy, forcing students to merely "read through their ears".

To provide a concrete illustration of the bias in the written text, the researcher presents a comparison between the text used in the instrument at SD 15 Banda Aceh and the text that has been reconstructed based on the principles of oral communication authenticity. The results of the research are focused on 6 topics that represent the category of listening texts. Each category of findings was assessed for the level of linguistic features, lexical density, and level of cognitive load. Assessment of the text based on the assessment rubric in table 1.

Table 3. Analysis of Narrative Text

Analysis Aspect	Written Text Bias Version	Authentic Version
Listening Manuscript	"Pada suatu pagi yang sangat cerah, seekor kancil yang cerdik sedang berjalan di pinggir sungai yang jernih untuk mencari timun guna memuaskan rasa laparnya."	"Nah, suatu pagi... hari itu cerah sekali. Ada seekor kancil, kalian tahu kan kancil yang cerdik itu? Dia jalan di pinggir sungai. Dia sedang cari apa? Timun. Iya, dia lapar... lapar sekali."
Linguistic Features	Using multi-tiered, rigid, and formal compound sentences (<i>scripted speech</i>).	Using short sentences, there are pauses, redundancies (repetition of words "lapar"), and interaksi.
Lexical Density	Height. Information is condensed into one long sentence.	Low. Information is provided gradually (per piece/chunking).
Cognitive Load	Height. Students must dissect complex grammatical structures while listening.	Low. Giving students time to process information through pauses and repetitions.

The absence of authentic spoken elements in this narrative-themed listening text highlights the gap between theory and practice in the field. The use of complex sentences like *"Pada suatu pagi yang sangat cerah..."* is a representation of stiff written language that is rarely found in natural spoken communication. This contrasts sharply with the principle of authenticity in oral communication, which emphasizes redundancy, pauses, and interaction, such as using the phrase *"Nah, suatu pagi..."* or repeating the word *"lapar"* to reinforce information retention. A text that lacks these oral characteristics is not actually testing students' listening comprehension, but is merely "reading aloud."

Theoretically, high lexical density in the biased version forces students' working memory to simultaneously parse complex grammatical structures while listening. This is relevant to Cognitive Load Theory (Sweller, 2010, 2024), where excessive intrinsic cognitive load hinders information assimilation. Conversely, the authentic version provides time for the brain to process information thru pauses and repetition, thus supporting a deeper understanding of meaning.

Table 4. Analysis of Instruction Text

Analysis Aspect	Written Text Bias Version	Authentic Version
Listening Manuscript	<i>"Siswa diinstruksikan untuk segera mengumpulkan lembar jawaban ke meja guru setelah bel tanda berakhirnya ujian berbunyi."</i>	<i>"Anak-anak, dengarkan ya... kalau bel sudah bunyi, ujiannya selesai. Nah, langsung kumpulkan kertas jawabannya ke depan, ke meja Ibu. Ya, segera ya."</i>
Linguistic Features	Passive sentences, stiff, without oral pauses.	Direct imperative sentences, with pauses, and discourse particles ("Nah").
Lexical Density	High (information is condensed into one clause).	Low . Information is provided gradually (per piece/chunking).
Cognitive Load	High (students must process a rigid time sequence).	Low (straight to the core action)

In terms of instruction, the field script uses stiff, passive sentences without verbal pauses, such as *"Siswa diinstruksikan untuk segera mengumpulkan..."*. The absence of this oral feature eliminates students' opportunity to interact with real-world linguistic features full of repetition. The authentic version using the discourse particle *"nah"* and direct commands is easier to process because it goes straight to the core of the action. This relates to the Modality Principle in learning (Aslan et al., 2024; Çeken & Taşkın, 2022; Cromley & Chen, 2025), where effective auditory communication should avoid overloading the verbal processing channel by using more functional language.

Table 5. Analysis of the Announcement Text

Analysis Aspect	Written Text Bias Version	Authentic Version
Listening Manuscript	<i>"Sehubungan dengan adanya perbaikan pipa air, seluruh warga sekolah dilarang menggunakan toilet di lantai dua hingga pemberitahuan lebih lanjut."</i>	<i>"Pemberitahuan ya... toilet di atas, lantai dua itu, jangan dipakai dulu. Kenapa? Karena pipanya lagi diperbaiki, sedang dibetulkan. Jadi, jangan ke sana dulu ya."</i>
Linguistic Features	Using the formal conjunction "with regard to."	Use of interrogative intonation ("Why?").
Lexical Density	High (dominated by abstract nouns).	Low (standard daily language).
Cognitive Load	High (causing cognitive fatigue)	Low (giving time for thought).

The findings in the announcement text indicate a dominance of formal conjunctions such as *"Sehubungan dengan..."* and abstract nouns with high lexical density. The use of overly formal language can lead to cognitive fatigue in students due to the long processing time. In the authentic version, the use of interrogative intonation like *"Kenapa?"* serves as an attention anchor that activates students' thinking schemas. From a cognitive psychology perspective, the thinking breaks provided in the authentic version allow for the consolidation of short-term memory into meaningful information.

Table 6. Analysis of Informative Texts

Analysis Aspect	Written Text Bias Version	Authentic Version
Listening Manuscript	<i>"Konsumsi buah-buahan secara rutin sangat disarankan bagi anak-anak karena kandungan vitamin yang tinggi mampu meningkatkan sistem kekebalan tubuh."</i>	<i>"Ayo, siapa yang suka makan buah? Jeruk, apel... itu bagus lho buat badan. Ada vitaminnya, banyak sekali. Biar apa? Biar kita sehat, biar tidak gampang sakit."</i>
Linguistic Features	Complex sentences with intricate subordinate clauses.	Benefit repetition (redundancy).
Lexical Density	High (dense medical/health term)	Low (child-friendly vocabulary).
Cognitive Load	High (difficult to filter out key information).	Low (emphasizing the keyword).

Health-related texts are often filled with dense medical terminology and complex compound sentences with intricate subordinate clauses. This makes it difficult for students to filter out key information due to the high cognitive load. The authentic version simplifies this density with child-friendly vocabulary and emphasizes keywords thru benefit redundancy. Verbal repetition of benefits strengthens information assimilation into long-term memory, consistent with the Law of Repetition in learning theory.

Table 7. Analysis of News Text

Analysis Aspect	Written Text Bias Version	Authentic Version
Listening Manuscript	<i>"Pemerintah daerah mengimbau masyarakat untuk melakukan kerja bakti guna mencegah terjadinya wabah penyakit demam berdarah di musim penghujan."</i>	<i>"Nah, sekarang kan sering hujan. Pak Camat bilang, 'Yuk, kita kerja bakti, bersihkan lingkungan'. Supaya apa? Supaya tidak ada nyamuk, biar tidak ada demam berdarah."</i>
Linguistic Features	The formal SPOK structure is very long.	Contextualization with the current situation.
Lexical Density	High (many bureaucratic terms)	Low (focus on real action) Cognitive Load
Cognitive Load	High (working memory load increases)	Low (information is presented in chunks).

Field news texts use a very long and formal SPOK structure and bureaucratic terms, which increases working memory load. This characteristic indicates the instrument's failure to represent real communication situations. Contextualizing with current situations and breaking down information into chunks in an authentic version makes the text more relevant for students. This supports Contextual Teaching and Learning (Mana et al., 2020; Rahman et al., 2023; Yonanda et al., 2025), where students' understanding improves when the material is connected to real-life situations around them.

Table 8. Analysis of Scientific Texts

Analysis Aspect	Written Text Bias Version	Authentic Version
Listening Manuscript	"Air yang dipanaskan hingga mencapai suhu seratus derajat celcius akan mengalami proses penguapan dan berubah wujud menjadi gas atau uap air."	"Coba lihat air yang dimasak Ibu. Kalau panas sekali, airnya mendidih. Terus keluar asapnya, kan? Nah, airnya berubah jadi uap... uap itu seperti gas."
Linguistic Features	Technical terminology without the aid of spoken context.	Simple verbal analogies and process repetition.
Lexical Density	High (concept density per sentence).	Low (explaining concepts step by step).
Cognitive Load	High (students fail to grasp the core process).	Low (allowing the brain space to process abstract data).

Recent findings in science texts indicate the use of technical terminology without adequate oral context, leading students to often fail to grasp the core processes. The extremely high density of concepts per sentence forces the brain to work beyond its auditory capacity. The authentic version uses simple oral analogies (e.g., "*air yang dimasak Ibu*") and step-by-step explanations. This represents the Mental Model theory, where students build an understanding of abstract concepts thru concrete analogies that are already familiar in everyday life.

The empirical dominance of written text bias across the analysed listening instruments creates a critical pedagogical anomaly that blurs the boundary between reading and listening comprehension. The quantification results confirm that the field scripts systematically rely on dense, lexically rich structures with an absolute minimum redundancy index. From the perspective of Cognitive Load Theory, the absence of natural breaks and repetition in the "scripted" text causes the students' working memory to become overloaded, leading them to often fail to grasp the core message of the material presented. Conversely, ideal conditions demand simple and communicative sentence structures so that the brain has transitional space to process information step by step (Paas et al., 2003; Sweller, 2010). Consequently, instead of evaluating authentic auditory processing, these instruments merely test the students' cognitive resilience against excessive text-based information load.

This systemic absence of authentic spoken elements epitomized by rigid textbook phrases like "*Pada suatu pagi yang sangat cerah...*" exposes a severe gap between language theory and assessment practices. Natural oral communication inherently relies on strategic redundancy, prosodic cues, and interactive markers to facilitate information retention. When an evaluation instrument strips away these oral features, it ceases to function as a valid test of listening comprehension; instead, it reduces the assessment activity to a mere visual reading test shifted into an aloud auditory medium. This operational mismatch directly violates the Modality Principle (Aslan et al., 2024; Çeken & Taşkın, 2022) and Contextual Teaching and Learning frameworks (Rahman et al., 2023), as the testing materials fail to represent the functional language standards children encounter in real-world communicative environments.

Ultimately, this ongoing assessment misalignment severely invalidates student scores and undermines the core mandates of the *Kurikulum Merdeka*. Students may score highly by relying on rote written memory, yet remain untrained to navigate the loose, dynamic structures of everyday spoken interaction. To restore instrument validity, test developers and educators must systematically reconstruct listening assessments by applying text chunking, integrating pseudo-interactions, and lowering lexical density through child-friendly spoken analogies. Shifting the evaluative focus from rigid "scripted" text to authentic spoken communication is an absolute imperative to ensure that assessment outcomes truly reflect functional, real-world communicative readiness.

D. CONCLUSION AND SUGGESTIONS

Based This study concludes that the listening evaluation instruments at SD Negeri 15 Banda Aceh are heavily distorted by written text bias, transforming an auditory assessment into a mere "read-aloud reading test." Microscopic analysis reveals that the current instruments rely entirely on rigid scripted speech with high lexical density and zero natural oral features. This structural misalignment triggers severe systemic consequences: it induces low construct validity by measuring passive written literacy recall rather than genuine listening comprehension, while simultaneously overloading children's working memory and causing rapid cognitive fatigue due to the omission of conversational pauses and structural redundancies. Furthermore, by relying exclusively on static textbook narratives instead of real-world functional genres, these assessments fail to evaluate students' readiness to respond to dynamic, everyday spoken interactions.

Ultimately, these findings demand a radical pedagogical shift among educators and test developers. To restore assessment authenticity and align with children's cognitive limits, traditional listening evaluations must be systematically reconstructed moving away from static text toward auditorily designed scripts, simple verbal analogies, or dynamic digital media that truly capture functional oral communication.

ACKNOWLEDGEMENT

The author states that there is no conflict of interest, either financially or personally, with any party or organization that can be considered to affect the objectivity of this research. All data obtained from three assessment documents at Banda Aceh City Elementary School were analysed independently without interference from the school or related education authorities.

REFERENCES

- Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D., & Bearman, M. (2024). From authentic assessment to authenticity in assessment: broadening perspectives. *Assessment and Evaluation in Higher Education*, 49(4), 499–510. <https://doi.org/10.1080/02602938.2023.2271193>
- Aslan, S., Durham, L. M., Alyuz, N., Okur, E., Sharma, S., Savur, C., & Nachman, L. (2024). Immersive multi-modal pedagogical conversational artificial intelligence for early childhood education: An exploratory case study in the wild. *Computers and Education: Artificial Intelligence*, 6. <https://doi.org/10.1016/j.caeai.2024.100220>
- Bachman, L. F., & Palmer, A. F. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press. https://books.google.co.id/books?hl=id&lr=&id=E0yH0NdySrQC&oi=fnd&pg=PP9&dq=Language+Testing+in+Practice:+Designing+and+Developing+Useful+Language+Tests.&ots=CEBXeBIMCL&sig=KBrenjwHLnJ SizOj kH3MksCbLL0&redir_esc=y#v=onepage&q=Language%20Testing%20in%20Practice%3A%20Designing%20and%20Developing%20Useful%20Language%20Tests.&f=false
- Çeken, B., & Taşkın, N. (2022). Multimedia Learning Principles In Different Learning Environments: A Systematic Review. *Smart Learning Environments*, 9(1). <https://doi.org/10.1186/s40561-022-00200-2>
- Cromley, J. G., & Chen, R. (2025). A meta-analysis of Richard Mayer's multimedia learning research: Searching for boundary conditions of design principles across multiple media types. In *Educational Research Review* (Vol. 49). Elsevier Ltd. <https://doi.org/10.1016/j.edurev.2025.100730>
- Demirbaş, İ., & Şahin, A. (2022). The Effect of Digital Stories on Primary School Students' Listening Comprehension Skills. *Participatory Educational Research*, 9(6), 380–397. <https://doi.org/10.17275/per.22.144.9.6>
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, 40(2), 177–190. <https://doi.org/10.1037/0012-1649.40.2.177>
- Iqbal, R. (2022). *Learning to Listen in the Perspective of Learning Concepts and Objectives in Elementary School* [Universitas Pendidikan Indonesia]. <https://repository.upi.edu/85664/>

- Iqbal, R., Fakhurrizi, & Maulana, Y. (2023). Traditions and Habits of Listening Skills in the World of Education at Banda Aceh Elementary Schools. *Journal of Education Science (JES)*, 9(2), 176–182. <https://doi.org/https://doi.org/10.33143/jes.v9i2.3436>
- Mana, L. H. A., Yusandra, T. F., Atmazaki, A., & Ramadhan, S. (2020). Development of a Listening Skills Textbook Based on Contextual Teaching and Learning. *Jurnal Kependidikan: Penelitian Inovasi Pembelajaran*, 4(1). <https://doi.org/10.21831/jk.v4i1.13594>
- McNamara, D. S., & Magliano, J. (2009). Toward a Comprehensive Model of Comprehension. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 51, pp. 297–384). Elsevier Inc. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- Morales, H. S., & Fernández, L. C. (2019). Authentic Assessment in the Listening Comprehension Classroom: Benefits and Implications. *Gist Education and Learning Research Journal*, (19), 6–30.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1
- Poerwanti, J. I. S. (2012). Development of an Authentic Assessment Model for Indonesian Language Subjects in Elementary Schools. *Tahun*, 21, 152–158.
- Rahman, A. A., Zulkifli, Z., Kamaruddin, I., Azhari, D. S., & Supriyadi, A. (2023). The Effect of Contextual Teaching Learning (CTL) Model on Students' Achievement in Elementary School. *Edunesia: Jurnal Ilmiah Pendidikan*, 4(1), 146–157. <https://doi.org/10.51276/edu.v4i1.282>
- Rossi, O., & Brunfaut, T. (2021). Text Authenticity in Listening Assessment: Can Item Writers Be Trained to Produce Authentic-sounding Texts? *Language Assessment Quarterly*, 18(4), 398–418. <https://doi.org/10.1080/15434303.2021.1895162>
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, 110. <https://doi.org/10.1016/j.lindif.2024.102423>
- Yonanda, D. A., Patimah, T. S., Efastri, S. M., & Vioreza, N. (2025). Influence Of Contextual Teaching and Learning Model On Elementary School Students' Critical Thinking Ability. *PUSAKA: Journal of Educational Review*, 2(2), 22–33. <https://doi.org/10.56773/pjer.v2i2.3>