Multimodal Design and Local Wisdom in EFL Recount Text Instructional Materials: A Critical Document Analysis

Teti Susanti[1], Widya Syafitri[2*], Veni Roza[3], Leli Lismay[4]

[1]*Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia: tetisusanti19@gmail.com*

[2*]*Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia: widyasyafitri@uinbukittinggi.ac.id*

[3]*Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia: veniroza@uinbukittinggi.ac.id*

[4]*Universitas Islam Negeri Sjech M. Djamil Djambek Bukittinggi, Indonesia: lelilismay@uinbukittinggi.ac.id*

2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License-(CC-BY-SA) (https://creativecommons.org/licenses/by-sa/4.0/)

DOI : https://doi.org/10.30983/mj.v4i2.11216

Submission: February 18, 2026	*Revised: May 30, 2026*	*Published: June 30, 2026*

Abstract

This study critically analyses an instructional material package for teaching recount text to Grade VIII EFL students at SMP Negeri 4 Bukittinggi, West Sumatra, Indonesia, using two theoretical lenses: Kress and van Leeuwen's multimodal social semiotic theory and the concept of local wisdom as a framework for culturally embedded pedagogy. The document under analysis is a Deep Learning-based instructional material package spanning four meetings, which integrates instructional design, teaching content, student worksheets, and assessment instruments within a single object of study. The document was analysed through critical qualitative document analysis, combining iterative close reading with thematic coding against two code families derived from the theoretical frameworks, supported by an audit trail and negative case analysis to strengthen trustworthiness. The analysis finds that multimodal principles are partially and unevenly applied: visual and kinaesthetic modes are present but function mostly as motivational supplements rather than as epistemologically integrated semiotic resources in the sense theorised by Kress (2010). The auditory mode is similarly present but is restricted to teacher-dominated oral channels rather than distributed multimodal meaning-making tasks. Regarding local wisdom, the analysis finds an absence of intentional integration: the instructional materials make no explicit reference to Minangkabau cultural values, oral narrative traditions, or community-based knowledge as pedagogical resources, despite the materials being designed for students whose primary cultural context is Minangkabau. This gap represents missed pedagogical opportunity given the structural and functional parallels between the Minangkabau oral narrative tradition of Kaba and the recount text genre. The study concludes with recommendations for how both theoretical frameworks can be more substantively and systematically embedded in EFL instructional material design in culturally specific Indonesian contexts.

Keywords: *multimodal theory, local wisdom, recount text, EFL instructional materials, critical document analysis, genre-based pedagogy, Minangkabau*

1. Introduction

The design of EFL instructional materials has attracted growing scholarly attention as researchers and practitioners recognize that the quality of materials determines, in substantial measure, the quality of the learning opportunities made available to students (Tomlinson, 2014). Two theoretical frameworks have emerged as particularly generative for the critical evaluation of such materials: multimodal social semiotic theory, which examines how meaning is made through the orchestration of multiple representational modes including language, image, sound, and gesture (Kress, 2010; Kress & Leeuwen, 2004); and local wisdom, which foregrounds the pedagogical value of embedding culturally situated knowledge, values, and practices into instructional design (Sadri, Bagus, & Bayu, 2025). Hyland (2004) defines genre as staged, goal-oriented social action, which provides the basis for analysing recount as a genre here.

In the Indonesian EFL context, both frameworks hold particular salience. The Merdeka Belajar curriculum framework, introduced by the Ministry of Education, Culture, Research and Technology, explicitly calls for deep learning, learner-centered pedagogy, and the integration of character values into subject teaching. However, national curriculum mandates rarely specify how teachers should operationalize multimodal design or local wisdom integration at the level of actual instructional materials. The result is that individual teacher-developed materials may vary widely in the degree to which they genuinely enact either framework, as opposed to gesturing toward them rhetorically.

This study addresses these gaps: between the theoretical ideal of epistemologically integrated multimodal design and its uneven realization in classroom materials, and between curriculum mandates for local wisdom integration and their actual operationalization at the material level, by subjecting a complete instructional material document for recount text developed for Grade VIII at SMP Negeri 4 Bukittinggi to a critical analytical reading through both multimodal theory and local wisdom as evaluative lenses. The document is selected because it represents a teacher-authored, school-level instructional material that claims to represent deep learning principles, multimodal approaches, and contextually responsive pedagogy, and therefore establishes an appropriate object for critical analysis of the distance between design intentions and theoretical ideals.

The study is guided by three research questions: (1) how do the instructional materials embody the principles of multimodal social semiotic theory, and in what ways is this embodiment partial or uneven? (2) In what ways, if any, do the instructional materials integrate local wisdom, specifically Minangkabau cultural knowledge and practices, as pedagogical resources? (3) What gaps can be identified between the theoretical ideals of multimodal and local wisdom pedagogy and their actualization in the materials, and what recommendations follow from these gaps?

2. Literature Review

2.1 Multimodal Social Semiotic Theory

Multimodal social semiotic theory, as developed by Kress and van Leeuwen (2004) and elaborated by Kress (2010), constitutes the primary theoretical lens of this study. The framework suggests that all communication is inherently multimodal, meaning that meaning is always made through the simultaneous arrangement of multiple semiotic modes, each of which carries specific affordances and limitations for representation. A mode, in this framework, is a socially and culturally shaped resource for making meaning, including but not limited to written language, spoken language, image, color, gesture, layout, music, and three-dimensional objects.

Multimodal social semiotic theory does not treat non-linguistic modes as mere decorative supplements to written text. Jewitt (2009) provides the standard toolkit for analysing such multimodal configurations, bringing together different approaches to how modes work together. Rather, it positions each mode as epistemologically equal in principle, carrying meaning that other modes cannot fully duplicate. Kress (2010) argues that in contemporary communicational environments, image has taken the central communicational role in many domains and that to continue treating writing as the default or dominant mode in instructional design is to misrepresent both how meaning is made in the world and how learners engage with meaning. From this perspective, a pedagogically suitable multimodal design requires that images, digital interfaces, and

audio resources be assigned specific semiotic tasks in the construction of knowledge, not merely deployed to decorate or motivate. Lankshear and Knobel (2011) describe such practices as new literacies, in which meaning is made through participation in digital and multimodal practices, not only through print

This framework has been extended to pedagogical contexts by Jewitt (2008), Sahiruddin et al. (2025), and Vaish and Towndrow (2006), who view classroom interaction and instructional materials as always multimodally constituted, so that what is made available for learning is shaped by the semiotic choices embedded in the materials. The New London Group's (1996) notion of multiliteracies stresses that contemporary learners need to develop the capacity to design meaning across modes, so instructional materials should ideally create space for learners to participate in such designing rather than merely consuming monomodal written texts. Cope and Kalantzis (2009) later reframed this pedagogy as learning by design, in which learners move through experiencing, conceptualising, analysing, and applying. Kalantzis and Cope (2012) further argue that literacies today involve designing meaning with available resources, rather than simply reproducing fixed rules.

2.2 Local Wisdom as a Pedagogical Framework

Local wisdom refers to the body of knowledge, values, practices, and cultural orientations that are generated, maintained, and transmitted within a specific community across generations (Sadri et al., 2025). In educational theory, local wisdom has been theorised as a pedagogical resource that can humanise and contextualise learning by connecting formal educational content to the experiential and cultural world of the learner. Gay and Helsel's (2020) and Patag's (2024) culturally responsive teaching frameworks together provide a foundational articulation of this principle in the Western literature, arguing that students learn more effectively when their cultural identity, community knowledge, and home language practices are recognized and mobilized as legitimate resources in the classroom. Barton and Hamilton (1998) similarly treat literacy as rooted in local practices, where what counts as reading and writing grows out of community life

In the Indonesian context, local wisdom pedagogy has been formally endorsed through the Merdeka Belajar curriculum, which includes local content as a curriculum component and positions the cultivation of national and local identity as a core educational goal. Scholars such as Wagiran (2012) have argued that local wisdom integration is not merely additive but transformative: it repositions the learner as a knower whose cultural experience constitutes a valid epistemic resource, rather than as a cultural blank slate to be filled with universalized curriculum content. Street (1995) makes the same argument for literacy, showing that reading and writing are always shaped by social and cultural practices, not by neutral skills alone.

In West Sumatra, where the Minangkabau ethnic group constitutes the majority of the population, local wisdom encompasses the philosophical system of *Adat Basandi Syarak, Syarak Basandi Kitabullah* (Custom is founded on Islamic law, Islamic law is founded on the Quran), communal values expressed through the customary village system, collective practices such as mutual cooperation, and a rich oral narrative tradition including the Kaba, a form of storytelling that recounts past events, ancestral histories, and moral exempla in a structured, chronological, and evaluative narrative format. The Kaba tradition is of particular theoretical relevance to this study because of its structural and functional parallels with recount text: both genres retell past events sequentially, both address a reader or audience, and both typically conclude with an evaluative or

reflective statement. Rose and Martin (2012) describe this same pattern as orientation, record of events, and reorientation, which gives the recount its recognisable shape. This generic overlap creates a natural and theoretically motivated bridge for local wisdom integration in recount text instruction. Martin (1992) defines genre as a staged, goal-oriented social process, which is why the Kaba and the recount can be read as two cultural realisations of the same social purpose.

3. Method

This study adopts a critical qualitative document analysis design (Bowen, 2009; Prior, 2003). This design is understood here as a case-based inquiry in the sense described by Creswell and Poth (2018), who treat case study as a means of gaining depth of understanding of a bounded system. Document analysis as a qualitative method involves the systematic examination of the content, organisation, and implicit assumptions of documents in order to generate interpretive insights about the phenomena they represent. In the tradition of critical discourse analysis (Fairclough, 2003), critical document analysis additionally attends to what is absent or marginalised in a text, not only to what is explicitly present. This dual attention to presence and absence is particularly important for this study, as the identification of gaps between theoretical ideals and material realisations constitutes a central analytical goal. Atkinson and Coffey (2004) remind researchers that documents are not neutral records but social products, made for particular purposes and audiences, which is why they need critical reading.

The primary data source is the Instructional material for Recount Text, Grade VIII, second semester, Academic Year 2025/2026, and developed at SMP Negeri 4 Bukittinggi. The document encompasses four instructional meetings and integrates lesson design documentation, substantive teaching content on recount text, student worksheets with reading and writing tasks, a diagnostic pre-assessment instrument with answer key and remediation matrix, formative assessment rubrics aligned to five Graduate Profile dimensions, and summative assessment instruments. In its integration of planning, content, and assessment, the document functions as a complete instructional material package rather than a conventional lesson plan, and is therefore analysed as such.

The site for this study, SMP Negeri 4 Bukittinggi, was chosen through purposive case-based sampling, not through random or convenience sampling. Three reasons guided the decision. First, the school teaches a Grade VIII EFL group drawn mainly from a Minangkabau background, so it offers a rich case for asking whether and how local wisdom enters teacher-made materials in the very cultural setting where such integration would matter most for teaching. Second, the material available for study is a full teacher-made package built around Deep Learning and multimodal ideas, which satisfies the need for a bounded document that actually embodies both frameworks rather than only referring to them in passing. Third, full access was obtained to the complete four-meeting package, including planning notes, worksheets, and assessment tools, which made it possible to examine one integrated object instead of scattered fragments. Because critical document analysis values depth of reading over statistical generalisation, a single-site, single-document design fits its purpose; the results are therefore offered as a case-specific critical reading, not as claims that apply to all Indonesian EFL materials. Patton (2015) calls this purposeful sampling, where cases are chosen because they are information-rich and can shed light on the questions being studied.

Analysis proceeded through three stages. The coding work in this study followed the thematic procedure described by Braun and Clarke (2006), who present thematic analysis as a way of finding, examining, and reporting patterns across a data set**.** First, repeated close reading of the document was conducted to develop familiarity with its content, structure, and rhetorical claims. Second, systematic thematic coding was applied using two code families derived from the theoretical frameworks: a multimodal code family (presence of modes, semiotic function assigned to each mode, integration or supplementation, learner role in multimodal meaning-making) and a local wisdom code family (explicit cultural references, structural parallels with local genres, community-based content, cultural values). Third, findings were interpreted critically, with analytical attention given both to what the materials do and to what they do not do relative to the theoretical frameworks. Trustworthiness was maintained through thick description grounded in direct document reference (Lincoln & Guba, 1985).

To further strengthen the dependability of the interpretation, three additional steps were taken beyond thick description. First, the multimodal and local wisdom code sets were used separately in two readings carried out a week apart, and the outcomes were then compared; where the two readings differed, the original document was revisited and the code definitions were tightened, which served as an intra-coder consistency check. Second, a clear audit trail was kept so that each claim made in the Findings and Discussion can be traced back to a particular excerpt from the document with its page reference, making the path from evidence to interpretation open to inspection. Third, disconfirming evidence was searched for on purpose: instead of noting only examples that supported the presence of multimodal or local wisdom elements, the analysis also highlighted partial or contradictory cases, such as the Story Arrangement activity, which prevents a purely negative account of the multimodal design. These steps do not replace inter-rater agreement with a second independent coder, and that absence is noted here as a limitation and as a suggestion for future replications. Shenton (2004) suggests that such checks, including audit trails and careful attention to disconfirming cases, help readers judge whether qualitative findings can be trusted

4. Findings

The findings are organised to answer the three research questions in turn. Section 4.1 addresses RQ1 by identifying which multimodal elements are present in the materials and analysing the semiotic function each is assigned, rather than merely cataloguing their occurrence. Section 4.2 addresses RQ2 by examining whether and how Minangkabau local wisdom is integrated as a pedagogical resource. Section 4.3 then answers RQ3 directly by synthesising the specific gaps between the theoretical ideals reviewed in Section 2 and what the document analysis in 4.1 and 4.2 actually found, together with the recommendations that follow from each gap.

4.1 Multimodal Elements Identified in the Instructional Materials

Across the four instructional meetings, the materials identify three learner characteristic types: visual, auditory, and kinaesthetic and specify learning activities addressed to each. This triple acknowledgement of modality constitutes the explicit multimodal orientation of the materials. Meeting 1 uses a short recount text presented alongside contextual images of adolescent holiday experiences. Meeting 2 displays holiday photographs with an untitled recount text. Meeting 3 uses

a recount text with bolded verb forms as a visual foregrounding device. Meeting 4 presents two contrasting recount texts (well-structured vs. poorly structured) for comparative analysis. Digital tools referenced include YouTube, Canva Education, Quiziz, Interacty.me, and Liveworksheets.com. Tomlinson (2012) argues that materials should be judged by how far they help learners notice language, engage affectively, and use language for real communication, which is the standard used for this evaluation here.

The kinaesthetic mode is operationalized most substantively through the Story Arrangement activity in Meeting 2, where students physically rearrange scrambled paragraphs to reconstruct the generic sequence of a recount text. The auditory mode is realised through teacher-led questioning, group discussion, oral presentation, and a YouTube video reference. The worksheet tasks include translation exercises, reading comprehension questions, generic structure labelling tasks, social function identification, and a personal recount writing task.

Read analytically rather than as a simple inventory, this distribution of modes reveals a consistent pattern: no meeting assigns two or more modes a shared, non-redundant semiotic task in relation to the same piece of content. The image-plus-text pairings in Meetings 1 and 2 place photographs and written text side by side without a worksheet prompt that requires students to draw information from the image that is not already stated in the text, so the image's semiotic contribution remains supplementary rather than load-bearing. The bolded verb forms in Meeting 3 typographically foreground grammatical form, but the accompanying tasks test recognition of that form rather than asking students to reason about why the writer chose to foreground it, so the visual device operates as a highlighting cue rather than as an object of analysis in its own right. Story Arrangement is the partial exception: because reconstructing paragraph order requires students to read for cohesion and genre structure simultaneously with physical manipulation, the kinaesthetic and cognitive tasks are functionally coupled rather than merely co-present. This distinction matters for answering RQ1, because it locates the materials' multimodal design on a specific point of a continuum between additive multimodality, where each mode duplicates or decorates the same content, and integrative multimodality in Kress's (2010) sense, where each mode is assigned a distinct and necessary semiotic function; the evidence indicates the materials sit closer to the additive end, with Story Arrangement as a partial counter-example rather than the norm.

The distribution across the four meetings also shows an escalation in cognitive but not multimodal complexity: Meeting 4's comparative task (well-structured versus poorly structured recount texts) requires evaluative reasoning absent from Meetings 1 to 3, yet it is delivered entirely through linguistic text, with no corresponding image, audio, or kinaesthetic component. This suggests that as the materials move toward higher-order tasks, they default back to a monomodal, text-only format rather than sustaining multimodal support for increasingly demanding content, a pattern that qualifies the materials' claim to systematic multimodal design across the sequence as a whole.

4.2 Local Wisdom Elements Identified in the Instructional Materials

The document does not contain explicit references to Minangkabau cultural values, oral narrative traditions, or community practices as named pedagogical resources. The term local wisdom does not appear in the document. The thematic content of the model texts and worksheet

reading passages like a cinema visit, a zoo trip, a holiday in Bali, a camping trip, a weekend at home, and a cooking experience, are universally relatable adolescent experiences, but none is drawn from or anchored in a specifically Minangkabau cultural context. The only potential point of cultural contact is the mention, in the learning partnership section, of "environmental learning experiences (mutual community) as a possible topic for student writing. However, this mention is brief, contextually peripheral, and not followed through in the actual tasks, model texts, or assessment instruments.

4.3 Summary of Gaps between Theoretical Ideals and Actualisation

Answering RQ3 directly, three specific gaps emerge from the findings in 4.1 and 4.2, each with a distinct implication for revision. First, an integration gap: Section 4.1 found that visual and kinaesthetic modes are present but function mostly as motivational or illustrative supplements rather than as epistemologically load-bearing resources in Kress's (2010) sense, with Story Arrangement as the partial exception; the implication is that task design, not the inclusion of additional modes, is the lever most likely to close this gap, since the materials already contain enough modal variety and instead need tasks that require students to draw meaning from more than one mode simultaneously. Second, a complexity-regression gap: the analysis in 4.1 found that the most cognitively demanding task, the Meeting 4 comparative activity, reverts to a monomodal, text-only format precisely where multimodal scaffolding would be most useful for lower-proficiency learners; the implication is that multimodal support should be extended into, rather than withdrawn from, the higher-order tasks in the sequence. Third, a cultural relevance gap: Section 4.2 found no explicit reference to Minangkabau cultural knowledge, oral narrative tradition, or community-based content despite the materials being designed for a Minangkabau student cohort and despite the structural overlap between the Kaba oral narrative tradition and the recount genre described in Section 2.2; the implication is that local wisdom integration requires a specific textual or task-level intervention, such as substituting or supplementing the existing holiday-recount model texts with a Kaba-based or community-sourced recount, rather than a general appeal to cultural relevance. Taken together, these three gaps indicate that the materials' limitations are not failures of resource provision but of task design and content selection, which is a more tractable problem for revision than a lack of multimodal or culturally relevant material would be.

5. Discussion

5.1 Multimodal Design: Partial Application and Theoretical Gaps

This discussion is organised around the three gaps identified in Section 4.3, which directly answer RQ3: Section 5.1 interprets the integration and complexity-regression gaps in relation to RQ1, Section 5.2 interprets the cultural relevance gap in relation to RQ2, and Section 5.3 considers assessment-related implications that follow from both. Section 5.4 then addresses the practical feasibility of the recommendations that follow.

The findings reveal that the instructional materials demonstrate awareness of multimodal principles but fall short of their full theoretical operationalization as theorised by Kress (2010) and Jewitt (2008). The central theoretical gap concerns the semiotic function assigned to non-linguistic modes. In multimodal social semiotic theory, the integration of image, digital media, and embodied

activity is pedagogically meaningful only when these modes are assigned distinct semiotic work that is, when they carry meaning that the written linguistic mode cannot carry alone, and when learners are positioned as active interpreters of those non-linguistic meanings. The question is not whether multiple modes are present in the materials, but whether each mode is deployed to do specific, non-redundant epistemic work in the construction of genre knowledge. Jewitt and Kress (2003) argue that such a collection is not yet multimodal literacy, because literacy in this sense means learning how modes combine to shape what can be known.

When examined against this theoretical standard, the visual resources in the materials function primarily as motivational activators rather than as independent semiotic carriers of recount text knowledge. The holiday photographs in Meeting 2 and the images in Meeting 1 are used to stimulate prior knowledge and create affective engagement before reading - a legitimate pedagogical function, but one that treats images as pre-textual scaffolding rather than as co-equal meaning resources. In Kress (2010)'s terms, the materials treat the visual as a supplement to the linguistic, rather than recognising that "image and writing make different contributions to the production of meaning". A theoretically complete multimodal design would assign images specific analytical tasks: students might be asked to compare what the image communicates about the experience with what the written recount communicates, identifying what each mode can and cannot represent, thereby developing explicit metalinguistic and meta-semiotic awareness.

The bolded verb forms in Meeting 3 represent a more linguistically sophisticated deployment of visual mode, using typographic emphasis to direct learner attention to a specific language feature within a running text. This technique, which functions as what Kress and van Leeuwen (2004) call visual salience, does represent a genuine multimodal design choice: the visual contrast of the bold type carries semiotic information - these words matter, these words are different - that the surrounding unbolded text does not. However, the analytical task that follows - categorising verbs as regular or irregular - remains at the taxonomic-cognitive level and does not invite students to reflect on why the past tense morphology looks the way it does or what it communicates about temporal experience, a level of metalinguistic reflection that would deepen the multimodal meaning-making encounter.

The Story arrangement activity in Meeting 2 is the most theoretically robust multimodal element in the materials. By requiring students to physically handle and reorder text segments, the activity engages the kinaesthetic mode in a task that has direct cognitive implications for genre understanding: the act of sequencing is epistemologically identical with the act of recognising how generic staging works. This is a case where the modality chosen, physical manipulation is not interchangeable with a written substitution; the embodied act of ordering physically enacts the chronological logic of the recount genre in a way that reading a written description of that logic does not. This design choice is consistent with what Jewitt (2008) describes as the epistemic potential of the kinaesthetic mode in classroom contexts.

The digital tools referenced in the document like YouTube, **Canva** Education, Quiziz, Interacty.me, and Liveworksheets.com, represent a potentially significant multimodal resource base. However, the document specifies their use at the level of platform names rather than pedagogical tasks. It is not specified, for instance, what the YouTube video is expected to contribute semiotically that the written model text does not, or what specific design tasks Canva is used for, or how Quiziz formatively assesses genre knowledge. Without task specification, these platforms function rhetorically in the document as signifiers of technological modernity rather than

as theoretically motivated semiotic resources. This gap between platform citation and pedagogical specification is a recurring limitation in the multimodal design of the materials. Warschauer and Healey (1998) similarly warn that technology in language teaching only adds value when it is tied to a clear pedagogical purpose, rather than used for its own sake.

The auditory mode is the most theoretically underdeveloped in the materials. Oral communication is represented primarily through teacher questioning and student group discussion - activities that are pedagogically valuable but not specifically multimodal in design. Genuinely multimodal auditory tasks would, for instance, ask students to compare the spoken narration of an experience (as in a video) with its written recount representation, analysing what prosodic features such as stress, intonation, and pause communicate about evaluation and emphasis that written punctuation and syntax can only approximately render. No such cross-modal comparative tasks appear in the worksheet or activity sequences. The reading aloud rubric in the assessment instruments, which evaluates fluency, accuracy, pronunciation, and intonation, suggests that oral performance is valued, but its integration into the instructional tasks themselves is not elaborated.

In summary, the multimodal design of the materials is best characterised as additive rather than integrative in the sense theorised by Kress and van Leeuwen (2004). An additive multimodal design places multiple modes alongside each other in the same instructional space but assigns each a separate, non-intersecting function - text for content, image for motivation, activity for engagement. An integrative multimodal design, by contrast, requires learners to work across modes, comparing, translating, and analysing what different modes do differently with the same experiential content. The materials predominantly exemplify the additive model, with the Story Arrangement activity as the most notable exception.

5.2 Local Wisdom: Absence as a Pedagogical Problem

The absence of Minangkabau local wisdom as a deliberate pedagogical resource in the instructional materials constitutes a significant theoretical and practical gap that warrants direct critical examination. This absence is not incidental; it reflects a broader pattern in Indonesian EFL materials development in which the target language's cultural apparatus - or a decontextualised, universalised adolescent experience - is treated as the natural content basis for instruction, while the learners' own cultural knowledge is rendered invisible. Byram (1997) argues that intercultural competence requires learners to draw on their own cultural frames as a basis for understanding others, which is why this absence matters for learning.

From the perspective of Gay and Helsel's (2020) and Patag's (2024) culturally responsive teaching frameworks, effective instruction uses "the cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse students to make learning encounters more relevant to and effective for them". The model recount texts in the materials - a cinema visit, a zoo trip, a Bali holiday, a weekend cooking at home - are thematically accessible and linguistically appropriate, but they are culturally generic in a way that potentially positions Minangkabau students as culturally neutral learners rather than as members of a specific, knowledge-rich cultural community. The Bali holiday text, in particular, positions a place that is culturally and geographically distant from West Sumatra as the default site of memorable experience, which may

inadvertently communicate that experiences rooted in the students' own cultural environment are less narratable or less valuable as recount text material.

The most significant missed opportunity in the materials concerns the genre parallel between recount text and the Minangkabau oral narrative tradition of the Kaba. The Kaba is a structured oral genre that recounts past events, for instance ancestral histories, heroic deeds, or community experiences, in a sequential, time-ordered narrative format that typically opens with an orientation establishing participants and context, develops through a sequence of events, and concludes with a reflective or evaluative statement (Fidawati, et al. 2026; Kibtiyah, et al. 2025; Pratama & Yandra, 2025). This structural profile is functionally identical to the recount text generic stages of Orientation, Events, and Reorientation as specified in the instructional materials. Had the materials drawn students' attention to this parallel - presenting a Kaba excerpt alongside a recount text, or asking students to retell a Kaba narrative using the recount genre structure - the genre itself would have become culturally meaningful rather than culturally inert. Students would have been positioned as cultural insiders whose community knowledge illuminates the target genre, rather than as cultural outsiders approaching a foreign textual form.

Furthermore, the Minangkabau value of *basamo mako manjadi,* the cultural principle that collective effort produces achievement, embedded in the communal social structure of the village which is directly relevant to the collaborative learning activities specified in the materials: group discussion, peer review, and joint text analysis. However, the materials invoke these collaborative structures without naming or acknowledging their resonance with Minangkabau social epistemology. Viewed from the perspective of Harahap et al. (2025) and Wagiran (2012), this represents a lost opportunity for students to see their cultural values reflected in and legitimised by formal schooling, which is itself a significant dimension of culturally responsive pedagogy.. The naming of cultural values in instructional materials is not merely decorative; it is an act of cultural recognition that positions learners as culturally constituted subjects rather than as generic students.

The one reference in the document to culturally grounded content that the mention of mutual collaboration and village events as potential topics for student recount writing in thef learning partnership section which is never operationalised in the worksheet tasks, model texts, or assessment instruments. It remains an aspiration at the planning level that does not penetrate the actual instructional content. This gap between rhetorical intention at the design framework level and absence at the material implementation level illustrates the systemic difficulty of local wisdom integration in practice: it requires not merely a declarative commitment at the planning stage but substantive embedding in the content, tasks, and assessment that constitute the actual learning encounter.

5.3 Alignment Between Learning Objectives, Tasks, and Assessment

Beyond the two primary theoretical frameworks, the analysis identifies a third area of concern: the alignment between stated learning objectives, instructional tasks, and assessment instruments. The four-meeting learning objectives specify a progressive sequence from identification of social function and generic structure (Meeting 1), through main idea and specific detail comprehension (Meeting 2), to language feature analysis (Meeting 3), and finally recount text writing production (Meeting 4). This sequence is consistent with the genre cycle model (Martin & Rose, 2008) and reflects sound genre-based pedagogical logic. Cope and Kalantzis (2015) argue that assessment for

multiliteracies must capture how learners design meaning across modes, not only whether they reproduce correct language forms.

However, a close examination of the worksheet tasks reveals that the reading comprehension tasks are concentrated primarily at the level of explicit information retrieval and structural identification. The three comprehension questions for the zoo text ask about the text's purpose, its entertaining or informing function, and why people share stories, questions that target literal comprehension and social function awareness but do not press students toward inferential reading, critical evaluation, or cross-textual comparison. The formative assessment rubric for the critical reasoning dimension specifies three performance levels from developing to proficient, with proficiency defined as presenting arguments supported by relevant data and evidence and responding to others' arguments. However, no worksheet task explicitly asks students to construct arguments about a text or evaluate competing interpretations, which means the aspiration to develop critical reasoning, as stated in both the Graduate Profile framework and the deep learning principles, is not sufficiently operationalised in the reading tasks.

Similarly, the metacognitive dimension of the learning objectives, students reflecting on and revising their writing, is addressed in the final meeting through a peer review activity and three reflective writing prompts. This is a theoretically appropriate design choice: metacognitive reflection is positioned at the production stage, where students have something to reflect on. However, the peer review process is not supported by explicit criteria or a structured feedback protocol in the worksheet, which risks producing superficial peer feedback rather than the substantive formative dialogue that would genuinely develop metacognitive competence.

5.4 Pedagogical Feasibility and Practical Limitations of Multimodal Design in an Indonesian Middle-School EFL Context

The critique offered above should not be taken to mean that a fully integrative multimodal design, as described by Kress (2010), can be put into practice without difficulty in the everyday setting of an Indonesian public middle school like SMP Negeri 4 Bukittinggi. A complete version of such a design demands real resources, time, and training, so the ideal discussed earlier needs to be weighed against the conditions in which teachers actually work.

Four practical limits need to be kept in mind. First, public junior secondary classes in Indonesia often hold more than thirty learners, which makes individualised cross-modal analytical work of the kind suggested in Section 5.1 hard to run; whole-class teacher-led formats, though less ideal in theory, may serve as a realistic first step from which more distributed multimodal work can be built up little by little. Second, the platforms named in the materials (YouTube, Canva Education, Quizizz, Interacty.me, Liveworksheets.com) assume stable internet and device access that cannot be taken for granted in all Indonesian schools, including in West Sumatra; advice about task design only helps as far as the basic infrastructure holds. Third, English teachers in Indonesian middle schools usually receive little pre-service or in-service preparation in multimodal pedagogy as such, as opposed to general ICT training, so asking them to build epistemologically integrated multimodal tasks without matching professional development risks shallow compliance rather than real change in teaching. Fourth, the assessment culture that still shapes much of Indonesian secondary English teaching, with its focus on coverage and standardised measures, rewards quick, easily marked tasks over the open-ended, comparative, cross-modal tasks that integrative

multimodal design calls for, which may hold teachers back from trying such tasks even when they know about them.

These limits imply that the suggestions put forward here are best followed step by step rather than through a full redesign. A workable starting point is to extend in small, low-cost ways what the materials already contain, for instance by adding one comparative prompt to the existing holiday-photograph activities, instead of bringing in new digital platforms or rebuilding the lesson sequences from scratch. Seen in this light, multimodal and local-wisdom integration can be pursued within the time and resources teachers already have, even if the fully integrative model described in the literature remains a goal to work toward rather than a standard that most Indonesian middle-school EFL classrooms can reach at once

6. Conclusion

This critical document analysis has examined the instructional materials for recount text at SMP Negeri 4 Bukittinggi through the dual lenses of multimodal social semiotic theory and local wisdom pedagogy. The analysis finds that the materials demonstrate genuine awareness of multimodal learning principles and deploy a variety of modes across the four instructional meetings, but that this deployment remains primarily additive rather than integrative in the theoretically complete sense articulated by Kress (2010). Non-linguistic modes function predominantly as motivational scaffolding rather than as epistemologically autonomous semiotic resources that learners are asked to analyse and compare with linguistic representations. The Story Arrangement kinaesthetic activity is the most theoretically robust multimodal element, but it stands as an isolated instance rather than representing a systematic multimodal design principle.

Regarding local wisdom, the analysis finds a substantive absence: the Minangkabau cultural context of the learners, including the Kaba oral narrative tradition, communal values, and community-based experiential knowledge, is not intentionally integrated into the instructional materials. This absence is a pedagogical problem not merely because it fails to satisfy theoretical criteria, but because it represents a missed opportunity to position Minangkabau students as cultural insiders whose community knowledge enriches rather than impedes their engagement with the recount genre.

Based on these findings, three recommendations are advanced. First, future iterations of the materials should assign non-linguistic modes specific analytical tasks rather than motivational roles: for instance, asking students to compare what a photograph communicates about an experience with what a written recount communicates, thereby developing explicit meta-semiotic awareness. Second, the materials should deliberately incorporate Minangkabau cultural content, most productively by introducing the Kaba as a culturally familiar genre parallel to recount text, and by using community-based events such as mutual cooperation, communal village festival, ancestral narratives as the primary thematic basis for model texts and student writing tasks. Third, the peer review activity in Meeting 4 should be supported by an explicit criteria-based feedback protocol that guides students toward substantive, criteria-referenced evaluation of their peers' texts, thereby operationalising the metacognitive and critical reasoning dimensions of the Graduate Profile assessment framework.

Future research should examine how teachers actually implement these materials in the classroom, whether the multimodal affordances designed into the document are realised or

modified in practice, and how students from Minangkabau cultural backgrounds respond to and engage with the absence or presence of local cultural content in their EFL reading materials.

References

Atkinson, P., & Coffey, A. (2004). Analysing documentary realities. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (2nd ed., pp. 56–75). SAGE.

Barton, D., & Hamilton, M. (1998). *Local literacies: Reading and writing in one community*. Routledge.

Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–38. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: An International Journal, 4*(3), 164–195. https://doi.org/10.1080/15544800903076044

Cope, B., & Kalantzis, M. (2015). The things you do to know: An introduction to the pedagogy of multiliteracies. In B. Cope & M. Kalantzis (Eds.), *A pedagogy of multiliteracies: Learning by design* (pp. 1–36). Palgrave Macmillan.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Routledge.

Fidawati, F., Bina, U., & Getsempena, B. (2026). Development of a local wisdom-based educational e-book to enhance social participation of students with special needs at SD Negeri 31 Banda Aceh. *Jurnal Pendidikan Humaniora, 14*(1).

Gay, G., & Helsel, C. B. (2020). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.

Harahap, R., Hasanah, U., & Soraya, S. (2025). Local Wisdom-Based Education as Character Development Through Social Studies Education (IPS). *International Journal of Research and Review*, *12*(April), 198–202.

Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.

Jewitt, C. (2008). *Multimodality and literacy in school classrooms*. Routledge.

Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Jewitt, C., & Kress, G. (2003). *Multimodal literacy*. Peter Lang.

Kalantzis, M., & Cope, B. (2012). *Literacies*. Cambridge University Press.

Kibtiyah, A., Idawati, K., & Muaz, Y. A. (2025). Collaboration on local wisdom-based character education between schools and parents of students in Islamic religious education units. *Al-Hayat: Journal of Islamic Education, 9*(2), 458–475.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. USA: Routledge.

Kress, G., & van Leeuwen, T. (2004). *Multimodal discourse: The modes and media of contemporary communication. Language in Society, 33, 109–156.*https://doi.org/10.1017/S0047404504221054

Lankshear, C., & Knobel, M. (2011). *New literacies: Everyday practices and social learning* (3rd ed.). Open University Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.

Martin, J. R. (1992). *English text: System and structure*. John Benjamins.

Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture.* Equinox.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–92.

Patag, F. A. (2024). Development of culturally based standardized scale in visible learning for instructional leaders. *American Journal of Educational Research, 12*(4), 128–140. https://doi.org/10.12691/education-12-4-2

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). SAGE.

Pratama, R. H., & Yandra, A. (2025). Teaching local wisdom through oral literature: Insights from teachers' beliefs and practices on Minangkabau Petatah-Petitih. *Jurnal Tahuri, 22*(1), 34–51.

Prior, P. (2003). Are communities of practice really an alternative to discourse communities? *American Association of Applied Linguistics (AAAL) Conference*, 1–29.

Rose, D., & Martin, J. R. (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney school.* Equinox

Sadri, N. W., Bagus, I. G., & Bayu, W. (2025). Local wisdom-based education in Indonesian school. *The Eastasouth Journal of Learning and Educations, 3*(03), 226–236. https://doi.org/10.58812/esle.v3i03

Sahiruddin, Sudarwati, E., Lestari, I. P., & Abdiel, R. (2025). *Digital multimodal literacy model of English language education in Indonesian higher education.*

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*(2), 63–75. https://doi.org/10.3233/EFI-2004-22201

Street, B. V. (1995). *Social literacies: Critical approaches to literacy in development, ethnography and education.* Longman.

Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching, 45*(2), 143–179. https://doi.org/10.1017/S0261444811000528

Tomlinson, B. (2014). *Developing materials for language teaching.* Bloomsbury.

Unsworth, L. (2008). Multimodal semiosis and the construction of a new English curriculum. In L. Unsworth (Ed.), *Multimodal semiotics: Functional analysis in contexts of education* (pp. 1–12). Continuum.

Vaish, V., & Towndrow, P. A. (2006). Multimodal literacy in language classrooms. In *Language and literacy* (pp. 317–345).

Wagiran. (2012). Pengembangan karakter berbasis kearifan lokal. *Jurnal Pendidikan Karakter, 2*(3), 329–339.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2), 57–71. https://doi.org/10.1017/S0261444800012970