Designing Multimodal Digital Media for Public Speaking in a Rural EFL Classroom: A 4D Development Study

Buhari*, Sam Hermansyah, Andi Sadapotto
Department of English Education, Universitas Muhammadiyah Sidenreng Rappang, Indonesia
*Corresponding Author Email: buharifakkah9@gmail.com

HOW TO CITE:

Buhari, Sam Hermansyah, Andi Sadapotto (2026). Designing Multimodal Digital Media for Public Speaking in a Rural EFL Classroom: A 4D Development Study *International Journal of Special Education, 41* (19s)975 - 985.

ABSTRACT

This study reports the design and classroom evaluation of a multimodal digital flipbook for public-speaking instruction in a rural Indonesian EFL setting. Guided by the 4D model (Define, Design, Develop, and Disseminate), the study involved 29 seventh-grade students, an English teacher, and two expert validators. Needs analysis indicated strong demand for learning resources combining text, images, video, and guided practice. The resulting Canva-based material was published through Heyzine and organized around speech structure, language support, audiovisual modelling, rehearsal, and performance tasks. Expert appraisal rated the media design at 86.7% and the language-content component at 95.6%. A five-student pilot and a 24-student classroom trial produced reported practicality ratings of 90.6% and 92.3%, while the teacher evaluation reached 87.5%. The findings show that context-responsive multimodal materials can make public-speaking pedagogy more accessible and engaging in a low-resource EFL classroom, although claims about proficiency gains require performance-based longitudinal evidence.

Keywords: EFL public speaking, multimodality, digital flipbook, rural education, materials development

1. Introduction

Speaking is a central but demanding component of English as a foreign language (EFL) learning. Learners must coordinate vocabulary, grammar, pronunciation, fluency, interactional awareness, and confidence in real time. Public speaking intensifies these demands because speakers are expected not only to produce intelligible English but also to organize an extended message, manage voice and pace, use gesture and eye contact, and maintain an audience's attention. It is therefore best understood as a multimodal communicative performance rather than as the oral production of sentences alone. For adolescents, structured public-speaking activities can create a meaningful bridge between classroom language and wider academic, civic, and professional communication.

The challenge is particularly acute in rural Indonesian schools. English may be encountered mainly through textbooks and teacher explanation, with few opportunities to observe varied models of spoken performance or to rehearse in supportive conditions. Limited connectivity, uneven access to devices, and teachers' restricted time for producing digital materials can further narrow the range of classroom activities. In such circumstances, learners may regard English as a school subject to be completed rather than as a resource for communicating ideas. Speaking anxiety then becomes both a linguistic and a participation problem: students avoid practice because they feel unprepared, and the absence of practice prevents them from becoming more prepared.

Digital materials can broaden the semiotic resources available to teachers and learners. A single screen-based resource may combine written explanations, photographs, icons, diagrams, audio, video demonstrations, hyperlinks, and interactive prompts. From a social-semiotic perspective, meaning is distributed across modes, and each mode offers particular affordances for representation and action (Kress 2010; Jewitt 2014). In public-speaking instruction, written language can make speech organization visible, video can model posture and delivery, audio can foreground stress and intonation, and task prompts can guide rehearsal. The pedagogical value lies not in the mere presence of multiple media but in the purposeful coordination of modes around a

communicative objective.

Multimedia learning research likewise suggests that appropriately integrated words and images can support comprehension when the design reduces unnecessary processing and directs attention to relevant information (Mayer 2009). In language education, digital storytelling, mobile-assisted learning, blended learning, and flipped instruction have been associated with greater learner participation and expanded opportunities for practice (Hung 2015; Ghazal and Singh 2016; Shadiev et al. 2017; Albiladi and Alshareef 2019). However, many interventions assume stable connectivity and relatively well-equipped classrooms. A technology-rich design may therefore widen rather than reduce inequality when it cannot be used under the conditions for which it was intended (Warschauer and Matuchniak 2010; Selwyn 2016). Rural EFL materials need to be lightweight, navigable, culturally recognizable, and usable with intermittent access.

The source study was conducted at SMP Satap 8 Maiwa in Enrekang Regency, South Sulawesi. Preliminary classroom information showed that learners were interested in visual and video-supported English activities, while the teacher required a practical resource that could be incorporated into existing lessons. The instructional focus was public speaking, an area that had received less attention than vocabulary, grammar, reading, or general conversation in the school's available materials. The development team therefore created a digital flipbook that integrated brief explanations, examples, video models, language support, and staged speaking tasks. Canva was used for page design and Heyzine Flipbook for screen-based navigation.

A further issue concerns the unit of analysis in technology-enhanced speaking research. Digital products are frequently described through lists of functions, whereas speaking outcomes are reported through broad scores or learner satisfaction. This separation makes it difficult to explain how a particular design feature might support a particular aspect of communication. The present study responds by documenting the link between design decisions and public-speaking actions. A video is treated as a model for noticing embodied and vocal choices; a visual outline represents discourse organization; a language frame supports formulation; and a checklist directs attention during rehearsal. This design-to-action account cannot establish causality, but it makes the pedagogical mechanism open to scrutiny and provides hypotheses for later performance-based research.

The study is positioned at the intersection of applied linguistics, multimodal pedagogy, and materials development. It does not treat the digital product as an end in itself. Rather, it asks how modes, tasks, and contextual constraints can be assembled into a usable resource for spoken performance. The research addresses three questions: (1) What learner needs and classroom constraints should guide the design of multimodal public-speaking materials in a rural EFL context? (2) How can the 4D model be used to translate those needs into a coherent digital flipbook? (3) How do expert validators, students, and the classroom teacher evaluate the validity and practicality of the resulting material? By answering these questions, the study contributes a context-sensitive account of technology-enhanced speaking pedagogy while distinguishing evidence of usability from stronger claims about learning effectiveness.

2. Multimodality, public speaking, and rural EFL materials

2.1 Public speaking as multimodal performance

Public speaking draws on several modes that operate simultaneously. Verbal wording carries propositional content, but audiences also interpret rhythm, loudness, pause, gaze, facial expression, posture, gesture, spatial positioning, and visual support. These resources are not decorative additions to language; they shape emphasis, interpersonal stance, credibility, and audience engagement. Kress (2010) describes modes as socially shaped resources for meaning-making, while Jewitt (2014) emphasizes that multimodal analysis examines how meaning emerges from the orchestration of modes in particular contexts. For EFL pedagogy, this means that a speech model should not be reduced to a transcript. Learners need opportunities to notice how linguistic and embodied choices work together.

The multimodal view also changes how performance is scaffolded. A public-speaking lesson may begin with a visual outline of introduction, body, and conclusion; move to a short video in which learners observe delivery; provide a language bank for opening, sequencing, exemplifying, and closing; and culminate in rehearsal with peer or teacher feedback. The resource becomes a sequence of representational and action-oriented supports. Text makes the organization inspectable, visuals make relations salient, and audiovisual examples make temporal features such as pace and pause observable. These supports are especially important for younger learners who may struggle to hold the entire structure of a speech in working memory.

For analytical purposes, the public-speaking task can be divided into at least four interrelated dimensions. The discourse dimension concerns

the organization of ideas and the signalling of relations among them. The linguistic dimension concerns vocabulary, grammar, pronunciation, and formulaic expressions appropriate to the topic and audience. The delivery dimension includes voice quality, pace, pausing, gaze, gesture, and posture. The interpersonal dimension concerns how the speaker establishes a relationship with listeners, projects confidence, and responds to audience expectations. A pedagogical resource that attends to only one dimension risks producing fragmented performance. The flipbook therefore attempted to connect the dimensions: organizational templates were paired with language choices, and both were paired with audiovisual delivery models and audience-oriented performance prompts.

Public-speaking anxiety is also partly a design issue. Learners often fear being evaluated before they have had sufficient opportunity to plan and rehearse. A digital resource can create intermediate stages between explanation and whole-class performance. Students may first listen, repeat, annotate a model, record a brief segment, compare it with a checklist, and then deliver a longer speech. Such sequencing is consistent with sociocultural accounts of scaffolded learning, in which temporary support enables performance that learners cannot yet manage independently (Vygotsky 1978). It is also compatible with communicative language teaching when digital activities lead toward purposeful spoken use rather than isolated manipulation of forms (Richards and Rodgers 2014; Nunan 2015).

2.2 Multimedia design and language-learning affordances

Mayer's (2009) cognitive theory of multimedia learning proposes that learners process verbal and pictorial information through partly distinct channels with limited capacity. Effective design therefore requires selection, organization, and integration rather than the accumulation of attractive elements. In a public-speaking flipbook, this principle supports concise explanations, clear signalling, segmenting of complex procedures, and close alignment between a video model and the task that follows. Unrelated animation, dense text, or excessive links can compete with the communicative objective and increase cognitive load.

Multimodal resources can also support metalinguistic and metacommunicative awareness. When a learner compares a transcript with a video, the contrast reveals information that is not represented by words alone: where the speaker pauses, which expression receives stress, when gaze shifts, and how gesture marks emphasis. Guided comparison can help learners name and discuss these features. Such awareness is important in EFL contexts because students may otherwise interpret successful speaking mainly as grammatical correctness or memorization. The material in this study explicitly presented delivery as a set of choices that interact with message organization. This orientation encourages learners to evaluate how a speech works for an audience, not merely whether every sentence is error-free.

Technology-mediated learning is additionally shaped by interaction among tool features, pedagogical goals, learner characteristics, and institutional conditions (Bower 2019). The same video that functions as a useful model in one setting may become unusable in another because of bandwidth, audio quality, language level, or cultural distance. Materials development must therefore involve adaptation and evaluation. Tomlinson (2012) argues that language-learning materials should provide meaningful exposure, engage learners affectively and cognitively, and facilitate opportunities for communicative use. These principles are particularly relevant to public speaking, where learners need both models and occasions to produce personally meaningful messages.

Research on mobile and digital language learning has identified benefits such as increased access, flexibility, repetition, and learner control (Stockwell 2012; Kukulska-Hulme 2020). Digital storytelling studies have also reported positive effects on participation and speaking confidence because learners combine narrative planning, multimodal composition, and rehearsal (Ghazal and Singh 2016; Kim 2020; Yunus and Ang 2021). Yet the transfer of these benefits cannot be assumed. Learners may enjoy a resource without developing measurable oral proficiency, and high satisfaction scores may reflect novelty. Consequently, product validation and user response should be reported as evidence of quality and practicality, while proficiency claims should rely on independently scored performances, comparison conditions, and longer-term observation.

2.3 Rural access, digital equity, and teacher mediation

Digital equity concerns more than the distribution of devices. It includes the quality of access, the relevance of content, the capacity to participate, and the outcomes enabled by technology (Warschauer and Matuchniak 2010). In resource-constrained environments, mobile or lightweight materials can be useful when they minimize data demands, permit repeated access, and fit existing classroom routines (Pimmer et al. 2016). Conversely, a sophisticated platform may be

impractical if it requires continuous streaming, individual accounts, or complex technical support. Context-responsive design therefore begins with constraints rather than treating them as problems to be addressed after the product is complete.

Teachers remain central to this process. The technological pedagogical content knowledge framework highlights the need to connect knowledge of subject matter, pedagogy, and technology rather than treating them as separate domains (Mishra and Koehler 2006). Teacher beliefs, confidence, and institutional culture strongly influence whether a tool is adopted and sustained (Ertmer and Ottenbreit-Leftwich 2010). In a rural public-speaking classroom, the teacher selects when the flipbook is projected, which videos are replayed, how tasks are grouped, what feedback is provided, and how technical interruptions are managed. The product does not replace instruction; it structures and extends the teacher's mediation.

The present study therefore evaluates the digital flipbook through multiple perspectives. Expert validation addresses linguistic, pedagogical, visual, and technical quality. Student response addresses clarity, attractiveness, navigation, and perceived usefulness. Teacher evaluation addresses classroom fit and efficiency. Together, these sources provide a more credible account of practicality than any single rating. They do not, however, establish causal improvement in public-speaking proficiency. This distinction informs the interpretation of the results below.

3. Method

3.1 Research design

The study used a Research and Development design organized through the 4D model: Define, Design, Develop, and Disseminate (Thiagarajan et al. 1974). The model was selected because it offers a transparent sequence from needs analysis to prototype construction, expert review, revision, and classroom use. In this study, dissemination was limited to implementation and evaluation within the participating school rather than regional or commercial distribution. The design combined quantitative questionnaire summaries with qualitative observations and open-ended feedback. This combination enabled the researchers to describe both numerical evaluations and the classroom experiences that informed revision.

Table 1 summarizes the relationship among the four phases, participants, activities, and evidence. The sequence was iterative in practice. Feedback obtained during expert validation and the small-scale trial led to revisions before the larger classroom trial. Thus, while the model is presented as four phases, decisions made at later stages prompted movement back to earlier design choices.

Table 1. The 4D development process and evidence sources

Phase	Main activity	Participants/evidence	Output
Define	Needs and context analysis	29 students; English teacher; questionnaire and discussion	Design requirements and public-speaking objectives
Design	Content mapping, storyboarding, modal selection, task sequencing	Development team; curriculum and classroom information	Initial prototype
Develop	Canva and Heyzine production, expert validation, revision, five-student pilot	Media validator; content validator; five students	Validated and revised flipbook
Disseminate	Teacher orientation, classroom implementation, response evaluation	24 students; English teacher; questionnaire and observation	Initial classroom practicality evidence

3.2 Context and participants

The research was conducted at SMP Satap 8 Maiwa, a rural junior secondary school in Enrekang Regency, South Sulawesi, Indonesia. The setting was selected because it represented the instructional problem motivating the study: limited access to varied English speaking resources and restricted opportunities for students to observe and practise public-speaking performance. The school's available technology supported teacher-led and shared access more readily than continuous individual online use. The material was therefore designed to work as a whole-class resource, a small-group resource, and a guide for individual rehearsal when access permitted.

Participants were 29 seventh-grade students, their English teacher, one media-design validator, and

one English-language content validator. The students represented varied levels of English proficiency and familiarity with digital learning. Five students took part in the initial pilot, and the remaining 24 participated in the larger classroom trial. The teacher contributed to the needs analysis, reviewed classroom suitability, and used the product during implementation. The validators were selected for complementary expertise so that technical design and language-pedagogical quality were considered separately.

3.3 The developed material

The product was a digital flipbook designed in Canva and published through Heyzine Flipbook. Its instructional sequence centred on a familiar topic, 'My School Activities', and connected topic vocabulary with public-speaking functions. The material included short explanations of speech purpose and structure, vocabulary and expression banks, model openings and closings, images related to school contexts, embedded or linked video examples, guided comprehension prompts, rehearsal instructions, and performance checklists. Navigation elements allowed learners and teachers to move between sections without reading the material as a linear textbook.

The design sought to coordinate modes around specific learning actions. For example, a page introducing speech openings combined a concise written explanation, visually separated language frames, and a video model. The next page asked learners to identify the greeting, topic statement, and preview in the model before drafting their own opening. Subsequent pages addressed sequencing, voice, gesture, audience orientation, rehearsal, and reflection. The aim was not to maximize multimedia density but to provide a progression from noticing to supported production.

3.4 Instruments and procedures

A student needs-analysis questionnaire examined preferences for printed and digital resources, images, video, e-books, and perceived ease of learning with digital media. Teacher needs were collected through discussion and a short response form focusing on classroom constraints, desired content, and ease of implementation. The media validator used a Likert-type instrument covering layout, colour, typography, image quality, video clarity, navigation, consistency, and usability. The content validator assessed language accuracy, relevance to curriculum, clarity of instructions, sequencing, appropriateness for seventh-grade learners, and support for public-speaking practice.

Student practicality questionnaires were administered after the pilot and larger trial. Items addressed attractiveness, clarity, relevance of pictures and videos, navigation, motivation, and perceived support for learning. The teacher evaluation focused on creativity, communicativeness, efficiency, classroom fit, and the extent to which the material supported explanation and participation. Classroom observations and field notes documented how learners navigated the resource, where they requested help, which elements attracted attention, and what technical or pedagogical adjustments were needed.

During the Define phase, questionnaire and teacher information were used to specify learner needs, topic level, technological constraints, and public-speaking objectives. During Design, the team produced a content map and storyboard, selected modes for each instructional purpose, drafted the language, and created an initial visual system. During Develop, the prototype was constructed, reviewed by the two validators, and revised. Major revisions included adding source acknowledgements for images, improving video clarity, strengthening navigation cues, and adding a back cover. The revised version was piloted with five students and adjusted again before the 24-student classroom trial. During Disseminate, the teacher received a short orientation and used the resource in the participating class.

3.5 Data analysis and ethical procedures

Questionnaire and validation responses were summarized as percentages and interpreted using the categories employed in the source study. Qualitative comments and field notes were read repeatedly and grouped into themes concerning engagement, clarity, navigation, public-speaking support, and access. Because the original dataset reported aggregated percentages rather than full item-level response distributions, the analysis in this article remains descriptive. No inferential claim is made about speaking improvement, and the terms validity and practicality refer to expert and user evaluation of the material.

Several procedures were used to strengthen descriptive trustworthiness. The needs analysis included both student and teacher perspectives so that learner preferences were not treated as the only design criterion. The two validators examined different dimensions of the product, and their written suggestions were retained as a revision trail. The small pilot preceded the larger trial, allowing usability problems to be addressed before the main classroom implementation. Quantitative summaries were interpreted alongside observations and open-ended comments rather than as self-sufficient evidence. Nevertheless, the absence of raw item-level response files and multiple validators in each domain limits reliability, and this limitation is

reported rather than obscured.

Permission was obtained from the school administration. The teacher, students, and parents or guardians were informed about the purpose of the study, voluntary participation, and confidentiality. Student identities were not included in reporting. Multimedia components were either produced by the development team, obtained from sources permitting educational use, or revised to include attribution following the media validator's recommendation.

4. Results

4.1 Define: learner needs and classroom constraints

The needs analysis indicated that students responded positively to visual and digital forms of support. The strongest reported responses concerned the attractiveness of digital media (86.2%) and the perception that digital media made English easier to understand (85.3%). Pictures, videos, and an e-book format were also positively evaluated. The overall need index reported in the source dataset was 88.3%. These responses did not show that students rejected print; rather, they suggested that print-only instruction did not provide the range of models and supports learners wanted for speaking.

Teacher feedback clarified the pedagogical meaning of these preferences. The teacher did not simply request more technology. The practical need was for a resource that could make speech organization visible, model delivery, reduce repeated explanation, and create a sequence of manageable tasks. The resource also had to function within limited lesson time and uncertain connectivity. These constraints led to four design requirements: concise pages, familiar topics, audiovisual modelling, and activities that could continue through teacher projection or printed/locally available alternatives when individual access was not possible.

Table 2 presents the student needs-analysis summary. The percentages are reproduced from the source study's questionnaire report. They are interpreted as indicators of perceived need rather than as measures of proficiency.

Table 2. Reported student needs-analysis responses (N = 29)

Item	Reported agreement	Interpretation
Enjoy learning English with printed books	67.2%	Moderately positive
Prefer learning English with pictures	82.7%	Positive
Prefer learning English with videos	76.5%	Positive
English is easier to understand with digital media	85.3%	Strongly positive
An e-book should be used for English learning	82.7%	Positive
An e-book should contain pictures	82.7%	Positive
An e-book should contain videos	80.0%	Positive
Digital media for English is attractive	86.2%	Strongly positive

4.2 Design: multimodal architecture of the flipbook

The storyboard transformed the needs analysis into a sequence of learning actions. Each unit began with a communicative purpose and a familiar context. Explanatory text was kept brief and placed near examples. Images supported topic activation and vocabulary recall. Video models were paired with focused observation questions so that learners watched for speech features rather than passively viewing. Interactive or written prompts asked learners to identify structural components, complete language frames, plan content, rehearse short segments, and reflect on delivery.

The design used recurring visual cues for four activity types: notice, plan, practise, and perform. This consistency reduced the need for lengthy instructions and helped students anticipate what to do on each page. Public-speaking content was organized around opening a speech, introducing a topic, sequencing ideas, using examples, controlling voice and body language, closing, and responding to a simple self-evaluation checklist. Localized school-related content allowed students to speak from experience instead of searching for unfamiliar information while simultaneously managing English.

The product's multimodality was functional rather than ornamental. Written language supported accuracy and planning; images activated context and provided memory cues;

audio and video represented temporal and embodied features; layout signalled relationships among stages; and task prompts converted observation into production. The resource therefore operated as a scaffold for performance. It did not provide automatic speech recognition or automated feedback, and it did not remove the need for teacher modelling and response.

4.3 Develop: expert validation and revision

The media-design validator awarded 52 of 60 points, reported as 86.7% and categorized as very good. Positive features included the overall layout, visual appeal, consistency, and navigability. Recommended revisions concerned image attribution, video resolution, and the addition of a back cover. These recommendations were implemented before the classroom trials. The content validator awarded 45 of 50 points, reported as 95.6% and categorized as excellent. The validator judged the language and content to be clear, relevant, and aligned with the intended level and instructional purpose.

The difference between the two validation scores was informative. The language-content component was closer to the intended standard, whereas technical and documentary features required more revision. This pattern illustrates why materials should be reviewed by specialists with different expertise. A linguistically accurate resource may still be difficult to navigate or ethically incomplete if images are not attributed. Conversely, an attractive interface may not provide appropriate language progression. The revision process therefore treated validity as multidimensional.

4.4 Disseminate: pilot, classroom trial, and teacher response

The five-student pilot produced a reported practicality rating of 90.6%. Students described the material as attractive and easy to navigate, and they responded particularly well to video examples and staged exercises. Observation notes suggested that the resource encouraged students to attempt short speaking tasks before whole-class performance. The pilot also helped identify points at which navigation instructions needed to be more explicit.

After revision, the larger trial involved 24 students and produced a reported overall practicality rating of 92.3%. Item-level responses were positive across interest, comprehension, systematic presentation, clarity, relevance of images and videos, visual design, ease of operation, and motivation. Attractive visual design received the highest reported item response (94.7%), followed by increased motivation (91.5%) and greater interest in English with digital media (91.2%). The responses indicate that students found the resource usable and engaging in the classroom context.

The teacher evaluation reached 87.5% and was categorized as very good. The teacher emphasized that the flipbook reduced lesson monotony, supported explanation, and prompted greater participation in public-speaking activities. The resource was also seen as efficient because examples, language support, and tasks were consolidated in one sequence. Table 3 summarizes validation and trial results, while Table 4 presents the reported item responses from the larger trial.

Table 3. Summary of validation and practicality results

Evaluation stage	Participants	Reported result	Interpretation
Media validation	One media-design expert	86.7%	Very good
Content validation	One English-language content expert	95.6%	Excellent
Small-scale pilot	Five students	90.6%	Highly practical
Large-scale classroom trial	24 students	92.3%	Highly practical
Teacher evaluation	English teacher	87.5%	Very good

Table 4. Reported large-scale trial responses (N = 24)

Aspect	Reported agreement	Interpretation
English is more interesting with digital media	91.2%	Very positive
The topic is easier to understand with the e-book	87.7%	Very positive
Content is presented systematically	86.0%	Positive
Text is easy to understand	88.6%	Positive
Pictures and videos are relevant	89.5%	Very positive

to the content		
Visual design is attractive	94.7%	Very positive
The material is easy to operate	90.0%	Very positive
The material increases motivation to learn English	91.5%	Very positive

Across the data sources, three qualitative themes recurred. First, audiovisual modelling made abstract features of delivery more observable. Students could connect terms such as eye contact, pace, and gesture with a visible performance. Second, staged tasks lowered the entry threshold for speaking. Learners were asked to produce manageable parts before delivering a complete speech. Third, teacher mediation remained necessary. Students still needed clarification, encouragement, language correction, and decisions about when to replay or skip media. The resource supported participation, but classroom interaction gave that support pedagogical direction.

5. Discussion

5.1 Multimodal scaffolding for spoken performance

The findings show how multimodal design can operationalize public-speaking pedagogy. Public speaking combines verbal, vocal, visual, and spatial resources; consequently, a text-only explanation provides an incomplete model. The flipbook coordinated modes so that students could observe, analyse, plan, and perform. This design is consistent with Kress's (2010) view that modes offer different meaning potentials and with Jewitt's (2014) emphasis on modal orchestration. It also extends the idea of multimedia learning beyond content comprehension. Here, words and images supported the development of a performance in which language, voice, and body had to be coordinated over time.

The positive student responses to pictures, videos, clarity, and visual design suggest that modal diversity was meaningful to learners. However, the result should not be interpreted as evidence that more media automatically produces better learning. The design's value lay in alignment. Videos were linked to observation questions; language frames were connected to drafting; and visual cues marked the progression from noticing to performance. This reflects Mayer's (2009) argument that multimedia supports learning when information is selected and organized rather than merely added. The product's recurring activity cues may also have reduced procedural uncertainty, allowing attention to shift toward the speech itself.

The study also suggests that public speaking is a productive object for applied-linguistic inquiry because it brings discourse, pronunciation, embodiment, and audience design into a single classroom activity. Much speaking research isolates fluency or accuracy, while many technology studies emphasize platform acceptance. A multimodal public-speaking task requires attention to how these domains converge in situated performance. The design process described here offers one way of making that convergence pedagogically visible. Future analysis could examine whether learners' use of signposting expressions co-develops with pausing, gaze, and gesture, or whether improvement in one mode compensates for persistent difficulty in another. Such questions would move evaluation beyond global impressions and clarify the communicative consequences of multimodal instruction.

The staged sequence addressed the affective dimension of public speaking. Students were not asked to move directly from explanation to a complete performance. Short planning and rehearsal tasks created lower-risk opportunities to participate. This finding parallels reports that technology-supported storytelling and online interaction can increase confidence and motivation by expanding rehearsal and learner control (Kim 2020; Wu et al. 2011; Yunus and Ang 2021). Yet confidence is not equivalent to proficiency. The present data capture willingness, perceived clarity, and practicality, while a rigorous assessment of oral development would require pre- and post-performance recordings rated for fluency, accuracy, organization, delivery, and audience engagement.

5.2 Context responsiveness and digital equity

A major contribution of the study is its attention to rural implementation. Educational technology is often evaluated in terms of innovation without sufficient attention to the conditions required for use. The 4D process began with local constraints: limited resources, uneven connectivity, teacher workload, and learners' need for familiar contexts. As Warschauer and Matuchniak (2010) argue, meaningful digital access depends on use and outcomes as well as equipment. The flipbook's relevance therefore rested on its ability to fit the school's actual ecology of shared screens, teacher guidance, and intermittent access.

The familiar topic 'My School Activities' also served an equity function. When learners speak

about known experiences, cognitive and linguistic resources can be directed toward organization and delivery rather than toward acquiring unfamiliar content. Contextualization does not mean restricting learners to local topics indefinitely. It means beginning from accessible knowledge so that new communicative demands are manageable. Tomlinson's (2012) principles of materials development similarly emphasize relevance, engagement, and meaningful use. The source study shows that global technologies such as Canva and Heyzine can be appropriated for local pedagogical purposes when content and task design are not imported uncritically.

The results further indicate that learner motivation may arise from several sources that should be distinguished. Visual novelty can attract initial attention, but sustained engagement is more likely when learners understand the task, see a feasible path toward completion, and can recognize progress. The flipbook supported these conditions through familiar content, stable visual cues, short stages, and repeated opportunities to prepare. The high practicality ratings may therefore reflect not only aesthetic preference but also reduced uncertainty about what public speaking requires. Future studies should test this interpretation through interviews, usage logs, and observation across repeated lessons after the novelty of the product has diminished.

Nevertheless, a digital flipbook cannot by itself resolve structural inequality. Devices, projection equipment, electricity, connectivity, teacher professional learning, and technical maintenance remain important. Selwyn (2016) cautions against technological solutionism, the belief that social and educational problems can be solved primarily through tools. The present product is better understood as a feasible classroom intervention within existing constraints. Its sustainability depends on teacher ownership, the possibility of updating links and videos, institutional support, and alternatives when technology fails.

5.3 The 4D model as a materials-development framework

The 4D model provided a practical structure for linking diagnosis, design, validation, and use. Its strength in this study was not linearity but traceability. Needs identified during Define could be connected to design features; expert comments could be connected to revisions; and classroom responses could be connected to further improvement. This traceability is valuable in language-materials research because product descriptions often list features without explaining why they were selected.

The expert validation also demonstrated the importance of separating technical and pedagogical quality. The media score of 86.7% and content score of 95.6% indicate strong overall appraisal, but the qualitative comments were more useful than the percentages alone. Image attribution, video clarity, navigation, language level, and curriculum alignment concern different dimensions of quality. A future validation process should include more than one expert in each area, report inter-rater agreement, and provide the complete rubric. Such procedures would strengthen the reliability of the validity claim.

Dissemination in this project was deliberately limited. The product was implemented in one school after a small pilot. In development research, dissemination is sometimes described too broadly, suggesting large-scale adoption when the evidence comes from a local trial. A more precise description is classroom implementation and initial user evaluation. Wider dissemination should follow replication, technical testing across devices, teacher training, and evidence from diverse rural and urban contexts.

5.4 Teacher mediation and classroom sustainability

The teacher's positive evaluation is significant because teachers decide whether educational technologies become routine or remain temporary projects. Ertmer and Ottenbreit-Leftwich (2010) show that technology integration depends on knowledge, confidence, beliefs, and culture. The participating teacher valued the resource because it consolidated explanations, models, and tasks and because it encouraged participation. These are pedagogical advantages rather than purely technical ones.

At the same time, the observations confirm that the teacher remained essential. Video models required framing; learners needed help noticing relevant features; speaking plans required language feedback; and performance anxiety required encouragement. The teacher also adapted pacing and grouping to classroom conditions. This pattern aligns with the technological pedagogical content knowledge framework, in which effective integration emerges from relations among technology, pedagogy, and subject matter (Mishra and Koehler 2006). Professional development for similar resources should therefore focus not only on operating the platform but also on modelling, scaffolding, feedback, and assessment of multimodal speaking.

Sustainability also depends on editability. A static product may become outdated when links fail, curricula change, or teachers need different topics. Canva-based source files can support adaptation when teachers have access and permission to modify them. Future versions

should include a teacher guide, downloadable alternatives for videos, accessibility features such as captions and transcripts, and clear licensing information. These additions would make the resource more robust across connectivity conditions and more inclusive for learners with different needs.

5.5 Methodological limitations and future research

Several limitations constrain interpretation. First, the study involved one class in one school, so the findings are context-specific. Second, the expert validation involved only two validators, one for media and one for content. Third, the trial data relied heavily on self-report and teacher evaluation, which may be influenced by novelty and social desirability. Fourth, the source dataset provided aggregated percentages but not complete response distributions, limiting recalculation and statistical analysis. Fifth, implementation was short, and no independently rated longitudinal speaking performances were reported.

Future research should combine development evaluation with a stronger learning-outcome design. Students' speeches could be recorded before and after a teaching cycle and rated by multiple assessors using an analytic rubric covering organization, linguistic accuracy, fluency, pronunciation, vocal delivery, gesture, and audience engagement. Inter-rater reliability should be reported. A comparison group or staggered implementation could help distinguish the effects of the material from ordinary classroom development. Interviews and stimulated recall could also examine which modes learners use during planning and how they transfer video observations into their own performance.

Further design research should investigate accessibility and connectivity more directly. Researchers could compare streaming and locally stored versions, track technical interruptions, and document the amount of teacher support required. Studies across several rural schools could identify which design principles are transferable and which require local adaptation. Finally, multimodal analysis of student speeches could move beyond overall scores to examine how verbal organization, gaze, gesture, and voice develop together. Such work would bring materials development into closer dialogue with linguistic and semiotic research on spoken performance.

6. Conclusion

This study developed and evaluated a multimodal digital flipbook for public-speaking instruction in a rural Indonesian EFL classroom. The 4D process connected learner and teacher needs with a design that combined concise text, familiar visual contexts, audiovisual modelling, language support, rehearsal, and performance tasks. Expert validators rated the media and content positively, and both student trials and teacher evaluation indicated high practicality. The results support the use of context-responsive multimodal materials as a scaffold for participation and public-speaking preparation in low-resource settings.

The study's most important implication is that multimodality should be treated as pedagogical orchestration rather than media accumulation. Each mode needs a clear role in the movement from noticing to planning and performance. The flipbook was useful because it made speech structure and delivery observable, divided a demanding task into manageable stages, and gave the teacher a coherent sequence for classroom mediation. Its value was also contextual: familiar content and relatively lightweight access responded to the realities of the participating school.

At the same time, the evidence establishes validity and perceived practicality rather than causal improvement in English proficiency. Stronger claims require longitudinal, performance-based assessment with multiple raters and broader samples. The product should therefore be regarded as a promising and revisable instructional resource, not as a completed solution. By making this distinction explicit, the study offers a more credible foundation for further research on multimodal speaking pedagogy, teacher-mediated technology use, and digital equity in rural EFL education.

Acknowledgements

The authors thank the administration, English teacher, students, and parents of SMP Satap 8 Maiwa, Enrekang Regency, for supporting the study. The authors also acknowledge the media and language-content validators and Universitas Muhammadiyah Sidenreng Rappang for academic support.

References

Albiladi, Waheeb S. and Khlood K. Alshareef (2019) Blended learning in English teaching and learning: A review of the current literature. Journal of Language Teaching and Research 10(2), 232-238. https://doi.org/10.17507/jltr.1002.03.

Bower, Matt (2019) Technology-mediated learning theory. British Journal of Educational Technology 50(3), 1035-1048. https://doi.org/10.1111/bjet.12771.

Ertmer, Peggy A. and Anne T. Ottenbreit-Leftwich (2010) Teacher technology change: How knowledge, confidence,

beliefs, and culture intersect. Journal of Research on Technology in Education 42(3), 255-284. https://doi.org/10.1080/15391523.2010.10782551.

Ghazal, Saman and Manjet Kaur Mehar Singh (2016) Digital storytelling in language classrooms: Student engagement and learning outcomes. Procedia - Social and Behavioral Sciences 232, 347-352. https://doi.org/10.1016/j.sbspro.2016.10.034.

Hung, Hsiu-Ting (2015) Flipping the classroom for English language learners to foster active learning. Computer Assisted Language Learning 28(1), 81-96. https://doi.org/10.1080/09588221.2014.967701.

Jewitt, Carey (ed.) (2014) The Routledge Handbook of Multimodal Analysis. 2nd ed. London: Routledge.

Kim, Hye-Soon (2020) Effects of digital storytelling on EFL learners' speaking confidence and motivation. Language Learning & Technology 24(2), 88-102.

Kress, Gunther (2010) Multimodality: A Social Semiotic Approach to Contemporary Communication. London: Routledge.

Kukulska-Hulme, Agnes (2020) Mobile-assisted language learning. In: Chapelle, Carol A. (ed.) The Cambridge Handbook of Language Learning Technology. Cambridge: Cambridge University Press, 217-233.

Mudinillah, A., Kuswandi, D. ., Erwin, E., Sugiarni, S., Winarno, W., Annajmi, A., & Hermansah, S. (2024). Optimizing Project-Based Learning in Developing 21st Century Skills: A Future Education Perspective. *Qubahan Academic Journal*, *4*(2), 86–101.

M U, Sari H, Hermansyah S, Maming K, Kahar A, Hasan, Elfahmi FK (2025), "Understanding Indonesian students' reading knowledge in digital literacy within socio-cultural of rural middle schools". *International Journal of Information and Learning Technology*, Vol. 42 No. 5 pp. 432–448, doi:

Mayer, Richard E. (2009) Multimedia Learning. 2nd ed. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511811678.

Mishra, Punya and Matthew J. Koehler (2006) Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record 108(6), 1017-1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x.

Nunan, David (2015) Teaching English to Speakers of Other Languages: An Introduction. New York: Routledge.

Pimmer, Christoph, Magdalena Mateescu and Urs Gröhbiel (2016) Mobile and ubiquitous learning in higher education settings: A systematic review of empirical studies. Computers in Human Behavior 63, 490-501.

Richards, Jack C. and Theodore S. Rodgers (2014) Approaches and Methods in Language Teaching. 3rd ed. Cambridge: Cambridge University Press.

Sam Hermansyah, Usman M., Syamsu T. (2026). Developing Digital Learning Media for Teaching Public Speaking to Improve English Skills of Junior High School Students. *International Journal of Special Education*, *41*(16s), 502–509. Retrieved from

Selwyn, Neil (2016) Education and Technology: Key Issues and Debates. 2nd ed. London: Bloomsbury.

Shadiev, Rustam, Wu-Yuin Hwang and Yueh-Min Huang (2017) Review of research on mobile language learning in authentic environments. Computer Assisted Language Learning 30(3-4), 207-227. https://doi.org/10.1080/09588221.2017.1308383.

Stockwell, Glenn (2012) Mobile-assisted language learning. In: Chapelle, Carol A. (ed.) The Encyclopedia of Applied Linguistics. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0768.

Thiagarajan, Sivasailam, Dorothy S. Semmel and Melvyn I. Semmel (1974) Instructional Development for Training Teachers of Exceptional Children: A Sourcebook. Bloomington, IN: Indiana University, Center for Innovation in Teaching the Handicapped.

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

Vygotsky, Lev S. (1978) Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Warschauer, Mark and Tina Matuchniak (2010) New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. Review of Research in Education 34(1), 179-225. https://doi.org/10.3102/0091732X09349791.

Wu, Wen-Chi Vivian, Ling-Ling Yen and Michael Marek (2011) Using online EFL interaction to increase confidence, motivation, and ability. Educational Technology & Society 14(3), 118-129.

Yunus, Melor Md. and Wei Siong Ang (2021) The use of digital storytelling in enhancing ESL learners' speaking skills. Arab World English Journal 12(1), 251-266. https://doi.org/10.24093/awej/vol12no1.17.