



Multimodal Digital Tasks to Foster Integrated Skills in College EFL Learners: A Conceptual Framework

Hari Prasad Sapkota

MPhil Scholar at the School of Education, Kathmandu University, Nepal

Email: sapkotah2@gmail.com

Abstract

English as a Foreign Language (EFL) instruction faces a persistent gap: no principled framework currently guides how multimodal digital tasks should be designed to develop college-level learners' integrated reading, writing, listening, speaking, and digital skills, even as digital technologies and multimodal communication continue to reshape the field. This conceptual paper addresses that gap by proposing a theoretically grounded framework for designing such tasks. Using a systematic narrative synthesis of empirical and conceptual research published between 2020 and 2025 (with foundational earlier sources retained where still influential), the framework integrates five theoretical traditions (Task-Based Language Teaching, multiliteracies pedagogy, social semiotic theory, sociocultural theory, and multimedia learning theory) into five interconnected dimensions: theoretical grounding, multimodal task design principles, digital tool integration criteria, skill integration mapping, and multimodal assessment design. The resulting framework offers a structured, evidence-informed basis for task design, illustrated through a sample multimodal documentary-project task. As a conceptual contribution, the framework has not yet been tested empirically, and its applicability across institutional and cultural contexts, particularly under-resourced ones, remains to be established. The paper concludes with implications for curriculum designers, language teachers, teacher educators, and policymakers, and outlines directions for future empirical validation.

Keywords: Digital multimodal composing, multiliteracies pedagogy, integrated skills, EFL college-level learners

Introduction

Language teaching has changed dramatically in the first decades of the twenty-first century, driven by digital technologies, multimodal communication practices, and new pedagogical paradigms. These shifts have redefined what it means for learners to communicate effectively, moving the goalposts from mastering isolated language skills

Copyright 2026 ©Author(s) This open access article is distributed under a *Creative Commons*



Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

to navigating meaning across multiple modes and platforms. EFL instruction has been no exception to this shift: how language courses are conceived, planned, and taught has changed accordingly (Hafner & Ho, 2020; Lim, 2020). The communication form of the modern world is a multimodal phenomenon, and the human being must learn to work and to orchestrate the meaning of texts, images, sounds, actions, and space (Kress, 2010; Yu et al., 2024). The traditional model of teaching reading, writing, listening, and speaking as separate, isolated skills is increasingly seen as insufficient for preparing learners to meet twenty-first-century, multimodal communication demands (Kress, 2010; Yu et al., 2024).

Digital transformation has been a wonderful transformation; it has displayed tremendous potential and challenges in the field of language teaching in the digital world. This period also exposed a need for principled guidelines on the effective use of digital technologies and multimodality in language teaching (Akhmetzadina et al., 2023), prompting renewed attention to how multimodal digital tasks should be designed for EFL classrooms.

EFL learners at the college level face increasingly complex communicative demands: they must interact with multiple genres of writing, produce multimodal academic and professional texts, and engage in digitally mediated, collaborative writing (Fan, 2025). Yet in many EFL contexts, instruction remains largely grammar-based and text-centered, with limited attention to multimodal literacies or integrated skill development (Lim & Kessler, 2023). This mismatch between how students are taught and how they are expected to communicate is central to the problem this paper addresses. Research on multimodal pedagogy and technology-enhanced language instruction is well developed, but far less attention has been paid to how multimodal digital tasks specifically should be designed to cultivate integrated language skills. Existing work tends to emphasize either multimodality or technology-integrated task design without theorizing how the two relate in terms of task design, tool integration, skill integration, and assessment (Lim & Kessler, 2023; Jiang & Hafner, 2024). This gap is especially pronounced for digital multimodal composing (DMC) and integrated skill development, both considered essential twenty-first-century competencies (Cope & Kalantzis, 2023; Kang & Kim, 2023).

To bridge this gap, this conceptual article is designed to present a theoretical framework to design multimodal digital tasks for the development of integrated skills for EFL college-level learners. It brings together ideas and concepts from five theoretical backgrounds (Task-Based Language Teaching (TBLT), multiliteracies pedagogy, social semiotic theory, sociocultural theory, and multimedia learning theory) and offers five dimensions of multimodal digital task design that are interrelated. Peer-reviewed research on multimodal pedagogy and technology-mediated task design published between 2020 and 2025, identified through a systematic search of Google Scholar, ERIC, Scopus, and Web of Science (detailed in Methods and Procedures), is reviewed and synthesized, and thus, the article presents the latest research in the field

of multimodal language teaching. At the end of the article, implications for practice, policy, and future research are discussed.

Literature Review

This review synthesizes recent research across the five theoretical traditions underpinning the proposed framework: multimodality and language education, digital multimodal composing, task-based language teaching and technology, sociocultural theory, and multimedia learning theory. Each strand is examined in turn, and Methods and Procedures then describes how this review was conducted.

Multimodality and Language Education

Theoretically and pedagogically, multimodality has become a prominent phenomenon in language education over the last two decades. From a theoretical point of view, multimodality is the presence of more than one semiotic mode in the process of meaning-making and the communication of meaning, such as written language, spoken language, images, sound, gesture, and spatial design (Kress, 2010). The theory of multimodality is rooted in Halliday's systemic-functional linguistics and has been further developed in Kress and van Leeuwen's (2021) social semiotic approach to multimodality, which argues that meanings can be conveyed across all modes and that communication is inherently multimodal. Multimodality has significant meaning in the field of language education, especially in understanding and teaching literacy. The New London Group (1996) first proposed that literacy education should not and cannot only be taught in the context of print literacy, but should be taught to recognize the multiplicity of communication devices and the greater diversity of society in terms of language and culture.

Multiliteracies pedagogy has continued to expand well beyond its original scope, adapting to new technologies as they reshape how meaning is made and communicated. This expansion reflects a broader shift in literacy education: from treating print as the default mode of communication to treating it as just one option among many. This call has since been renewed and updated by Cope and Kalantzis (2023), who reconceptualize multiliteracies in light of how digital technologies and artificial intelligence have transformed meaning-making practices.

They are now focusing on "multisituational and multiform" meaning-making, meaning that literate people need to be able to produce and comprehend texts that integrate one or more modes in different situations. There has been extensive research on multimodality in the EFL classroom over the last few years. Lim et al. (2022) conducted a systematic literature review on multimodality in the English language classroom was conducted on articles published between 2010 and 2021. 98 articles were reviewed, and it was found that teachers use different multimodal texts, like digital videos, ads, and websites, and engage students in different multimodal receptive and productive activities. The review pointed out important gaps, especially regarding

explicit multimodal instruction, assessment of multimodal competence, and research on multimodality in EFL contexts. A more recent systematic review on multimodal immersion in EL education in higher education by Rahmanu and Molnár (2024) reviewed 34 studies published between 2013 and 2024. They concluded that multimodal pedagogies are advantageous for language learning but stressed the necessity to further explore their impact in higher education contexts and to design models that offer a principled approach to integrating multimodality into curricula. This recurring call echoed across the multimodality literature but is rarely answered with a concrete, theory-grounded design model is precisely the gap the present framework is intended to fill.

Digital Multimodal Composing in Second Language Contexts

Digital multimodal composing (DMC): writing that draws on multiple modes (image, sound, video, animation) together with digital technology to produce texts was developed as a research focus primarily through work such as Hafner and Ho's (2020) process-based model, and has since become a fruitful area of research in multimodal and language education. Digital Multimodal Composition (DMC) is a writing process that makes use of multiple modes of language (images, sound, video, animation, and more) and digital technology to produce texts. DMC is increasingly used in second language (L2) writing research and is an effective tool for L2 writing skills, engagement, and competence. Zhang and Yu (2022) conducted a detailed research study in the journal 'Language Teaching' and came up with some salient findings. The study predicts a greater effectiveness of students' learning in the four domains of language when dealing with the four language areas as opposed to the traditional monomodal writing task, in which students' learning is focused on four language areas (Cho & Kim, 2024). Secondly, DMC has been proven to be an effective tool to improve the engagement, motivation, and creativity of the learners (Kang & Kim, 2023; Usmani et al., 2025).

Finally, the DMC activities co-created were found to have a positive influence on the ability to learn from peers (negotiate) and the digital literacy skills (Cheung, 2022). Nevertheless, there are some questions and problems that are not covered in the literature. From the results of this study, it can be concluded that there are some gaps in the field of DMC that need to be explored in more detail, e.g., pedagogical design of the tasks in DMC, teacher training in DMC, assessment of multimodal texts, and collaboration and translanguaging in DMC processes. From the 7-month longitudinal study, Kang and Kim (2023) positively concluded that the overall attitude of EFL adolescent learners toward DMC use was positive with a different degree of perception of the benefits of using DMC. The results confirm the importance of task design, including elements of catering for the diversity of learners and assisting them to access the multimodal digital tasks.

Task-based Language Teaching and Technology

Task-Based Language Teaching (TBLT) is a language teaching methodology

that is based on sound pedagogical principles to create meaningful and communicative language teaching tasks. TBLT was originally developed in the 1980s to enable teachers to introduce a more learner-centred approach to language teaching, to use the language to be taught for real tasks, and to put the language into use in real communicative situations (Ellis et al., 2020; Jackson, 2022). Based on the above, technology is an innovative way of TBLT because it would offer students contexts for their L2 tasks and would increase the opportunities for L2 meaningful communication (González-Lloret & Ziegler, 2022).

The relationship between TBLT and technology has grown more complex over the past decade, as the sheer number of available tools keeps expanding the questions worth asking about how technology shapes task performance and language development. This growing complexity is not simply a matter of more tools becoming available, but of technology increasingly reshaping what a “task” itself can look like and how learners engage with it. This trajectory has been mapped by East (2024), who reviewed a decade of TBLT research and practice and organized its technology-related developments into four categories. González-Lloret and Ziegler (2022) suggested that technology and TBLT can complement each other; TBLT can be used as a pedagogical method to help implement the use of technology in tasks.

Further, there are a few recent systematic reviews that have been conducted, indicating an increasing synergy between TBLT and technology. The literature review of Xie and Lan (2025) on TBLT and speaking competence in different settings of EFL showed that between 2014 and 2023, 38 research papers were published, indicating that TBLT in the context of EFL has a positive influence on the students’ oral competence. Pankeaw and Satayaban (2025) examined the potential of digital tools in supporting interaction and collaboration in the TBLT pedagogical practices and presented some issues and possibilities of digital literacy and access to digital tools in post-pandemic education. It is worth mentioning a systematic review of the methodology of ‘TBLT’ research using educational technology for the Annual Review of Applied Linguistics by Kim and Namkung (2024), which talks about methodological characteristics of TBLT research with educational technology and that methodological gaps need to be tackled in future studies.

Addressing these methodological gaps, particularly the need for research designs that link technology-mediated tasks to specific, measurable integrated-skill outcomes, is one of the motivations for the design-oriented framework proposed in this article.

Sociocultural Theory and Mediated Learning

The Sociocultural theory (SCT) is one of the critical theories of learning founded on the work of Vygotsky (1978), and it is a social-institutional theory of learning. A key principle in SCT is that of Zone of Proximal Development (ZPD), which refers to the difference between what the learner can do alone and what he or she can do

with support from an “other who is more knowledgeable” (Lantolf & Poehner, 2014). A context in which scaffolding is seen to be widely applied in language education is the context of temporary or structured scaffolding that is gradually phased out as the learner becomes competent (Walqui, 2006).

SCT, however, can serve as a foundation for exploring how Digital technology can be applied to help students move forward in their ZPD related to Digital language learning. Alkhudiry (2022) highlighted the theories of Vygotsky and their significance in the establishment and growth of a learning environment where technology can aid children’s development and learning in the realm of technology-enhanced education. Sarmiento-Campos (2022) in his research on effective scaffolding technique of ZPD to improve the oral skill development of EFL students revealed that applying the scaffolding technique of EFL students could be effective to improve the oral skills of students significantly by using the scaffolding technique of ZPD technique. Implementing SCT using a multimodal digital task design implies that tasks should be well sequenced, multiple modes should be used to access the task, and opportunities should be provided for making meaning together using digital media.

Multimedia Learning Theory

Multimedia learning theory offers a cognitive account of how learners process information presented in both verbal and pictorial form. This perspective bears directly on the cognitive design of multimodal digital tasks (Mayer, 2021). This theory assumes that meaningful learning is fostered when information is presented in both verbal and pictorial form and that the optimal design of the multimedia will facilitate the processing of information by decreasing the unnecessary cognitive load and increasing the generative processing. The use of well-designed multimodal tasks that integrate the modalities of text, image, sound, and video in a principled manner will likely improve the language comprehension and retention of the EFL teaching and learning context (Pellicer-Sánchez et al., 2020).

The principles of multimedia learning have also been verified in recent years through empirical research in the context of EFL. In his study entitled *The Impact of Digital Storytelling as a Multimodal Composing Approach on EFL Learners’ Writing Development*, Ajabshir wanted to see the effect of a multimodal composing approach, which combines text, images, narration in audio format, and video, on the writing development of EFL learners. The results of the quasi-experimental study showed that there was a significant difference between the students who performed digital story tasks and the control group in the improvement of syntactic complexity, accuracy, and lexical diversity. The researchers hypothesized that this enhancement was because of the multimodal character of the tasks, which incorporated multiple sensory modalities and boosted cognitive handling of the linguistic input. This finding lends empirical support to Mayer’s (2021) claim that well-designed multimodal input reduces cognitive load and increases generative processing. It is this mechanism, rather than novelty alone,

that the present framework treats as the operative one in Dimension 5's assessment design principles.

Methods and Procedures

Conceptual Methods

This article uses systematic narrative synthesis as its conceptual research method, developing a theory-based framework for multimodal digital task design at the EFL college level. Rather than drawing on empirical (experimental or observational) data, this approach relies on critical analysis and synthesis of existing scholarship to build theoretical frameworks that connect theory to practice (Jaakkola, 2020): a method with established precedent in applied linguistics for developing pedagogical tools and informing subsequent empirical study (Ellis et al., 2020; Jiang & Hafner, 2024).

The review draws on a systematic search of three databases (ERIC, Scopus, Web of Science) covering January 2020 to early 2025, a period chosen to ensure the framework reflects the most active era of multimodal digital pedagogy and language-teaching research; foundational studies published before 2020 are included only where they remain influential. Search terms were combined using Boolean operators across three concept clusters: multimodality/multiliteracies; EFL and task-based language teaching; and integrated skills/digital tool integration. Sources were included if they were peer-reviewed, directly relevant to one or more dimensions of the proposed framework, and situated in language learning and/or multimodal pedagogy; sources addressing only primary or secondary school contexts were excluded.

The five theoretical pillars: Task-Based Language Teaching, multiliteracies pedagogy, social semiotic theory, sociocultural theory, and multimedia learning theory were selected using two criteria: each is supported by a substantial evidence base, and each explains an aspect of multimodal digital task design not covered by the others. Other candidate traditions (e.g., complexity theory, genre-based pedagogy) were considered but not adopted as standalone pillars; where relevant, their contributions are absorbed into the five traditions retained.

The framework itself was built through four iterative steps: (1) thematic mapping of the literature to identify recurring design principles and theoretical constructs; (2) theoretical integration, examining convergences and complementarities across the five traditions; (3) elaboration of each dimension, translating abstract principles into concrete design guidance; and (4) a coherence review, checking that the five dimensions function as an interconnected whole rather than independent components. The illustrative task presented later in this article demonstrates this framework applied to one instructional design.

Results and Discussion

The Proposed Conceptual Framework

This section, based on the theories and empirical studies mentioned above, provides a comprehensive conceptual framework for designing multimodal digital tasks to develop integrated skills during EFL college-level learners' language learning. The framework consists of five dimensions that are interconnected: (a) theoretical grounding; (b) multimodal task design principles; (c) digital tool integration criteria; (d) integration of skills in the map; and (e) multimodal assessment design.

Dimension 1: Theoretical Grounding

The first dimension is the theoretical basis for all the following design decisions. The framework adopts an integrative approach that combines theories from various complementary fields, as no single theory can cover the multimodal digital task design that includes language learning, technology use, multimodal meaning-making, and social interaction.

It is based on five theoretical pillars. The first part of TBLT includes the general picture of the pedagogical approach, meaning of tasks, goal orientation in tasks, and communicative authenticity in tasks (Ellis et al., 2020; Jackson, 2022). Secondly, multiliteracies pedagogy (New London Group, 1996), more recently extended by Kalantzis and Cope (2025), moves beyond a print-based definition of literacy to encompass multiple modes of meaning-making. Thirdly, the application of social semiotic theory (Kress, 2010) is used to analyze the meaning-making process in different modes. Lastly, the sociocultural theory of Vygotsky (1978) is used, and concepts of ZPD, scaffolding, and mediated learning are applied to sequence and scaffold activities.

These angles complement each other. The pedagogical architecture is set forth by TBLT, it is engagement with situated practice, overt instruction, critical framing and transformed practice with the help of multiliteracies pedagogy; there is the use of social semiotic theory to guide the decision making process of modes; there is the use of social semiotic theory to inform the scaffolding and collaborative arrangements, and there is the use of multimedia learning theory to reduce the cognitive overload and effective multimodal processing.

Dimension 2: Multimodal Task Design Principles

The second dimension is grounded in seven key principles derived from theoretical foundations and recent empirical studies on DMC and technology-mediated TBLT. The first principle is that communication must be truly multimodal, i.e., communication must be multimodal. Tasks should involve students' interaction with and creation of multimodal texts that are authentic in terms of academic, professional, and social communicative practices (González-Lloret & Ziegler, 2022). The task will be for the learner to produce a video presentation in which he or she takes a stand

on a social issue; it involves all three modes (visual, auditory, and linguistic) in a way similar to that expected in authentic academic and professional communication. The second principle is “deliberate skill integration.” Tasks should also focus on developing multiple languages simultaneously, since natural language use involves multiple language skills (Rahmanu & Molnár, 2024). A good task could start with reading and viewing tasks, involve joint discussions, and end with the production of a multimodal text.

The third principle is “Progressive Scaffolding. Tasks should offer a progressive and structured level of support to learners as they become more competent and confident, including teacher modelling, graphic organizers, cooperative learning, and support for automated feedback. Structured supports should be incorporated that will be reduced as learners reach a more competent and confident level in completing tasks, including teacher modelling, graphic organizers, peer collaborations, and automatic feedback. Structured supports should be incorporated that will be reduced as learners reach a more competent and confident level in completing tasks, including teacher modelling, graphic organizers, cooperative learning, and support for automated feedback. Generative AI (GenAI) tools, such as writing assistants and conversational chatbots, can extend this scaffolding by offering learners on-demand, adaptive feedback and modelling at the point of need; however, their use should be paired with explicit instruction in critical AI literacy, so that learners evaluate and revise AI-generated suggestions rather than adopt them uncritically. The fourth one is called “modal awareness”. Tasks should be made explicit to make learners more aware of the modes; have them think about and say which one they have selected. Fan (2025) showed that explicit engagement with multimodal design was an effective means to support students in constructing and conveying academic discourse in their first year of EFL study.

The fifth rule is called ‘Engagement together’. Because DMC tasks work best in group settings to improve linguistic and digital literacy outcomes (Cheung, 2023), there should be opportunities to provide for interaction and negotiation of meaning among learners as well as for co-construction of multimodal texts (Cheung, 2022).

The 6th principle of critical digital literacy is at the core. Tasks should encourage students to consider the meaning, power, and ideology within multimodal texts, and to remember that reading, writing, and interacting with multimodal texts is a learning capacity – not a technical skill (Kalantzis & Cope, 2025; Yap & Gurney, 2023). The seventh principle is ‘Iterative process orientation’. Tasks need to stress planning, drafting, revising, and reflecting, not only the product, and allow for repeated versions of the text to help the student make semiotic choices and refine the text.

Dimension 3: Digital Tool Integration Criteria

The third dimension offers the criteria for selecting and using digital tools, and based on the principle of ‘pedagogy before technology’ - the selection of digital tools is based on pedagogical objectives rather than their novelty (González-Lloret, 2023)

The five criteria proposed in the framework are as follows: (1) Modal Affordance Alignment tools that will align with the specific modes required for the task; (2) Accessibility and Equity tools that will not create new inequalities, (3) Collaborative Functionality – that will support real-time co-editing, commenting, and peer feedback; (4) Learner Agency – tools that will provide learners with creative control over multimodal compositions; and (5) Institutional Sustainability: cost, technical support, and compatibility with existing systems. There is still plenty of flexibility and context when choosing tools, such as collaborative writing platforms, video creation and editing tools, podcasting software, presentation tools, and social media, to reach authentic audiences (Hasumi & Chiu, 2024). Generative AI tools warrant particular attention within this dimension, since they can support several criteria simultaneously: as drafting or brainstorming aids (Learner Agency), as scaffolds that model multimodal composition (Modal Affordance Alignment), and as low-cost alternatives to specialized software in under-resourced settings (Institutional Sustainability). Selecting GenAI tools should follow the same pedagogy-before-technology principle applied to other tools in this framework, with explicit attention to transparency (disclosing AI use), academic integrity, and the risk of homogenizing student voice if AI output is adopted without critical revision.

Dimension 4: Skill Integration Mapping

The fourth dimension adds to the systematic representation of multimodal digital tasks, which involve multiple language skills throughout the TBLT task cycle and thus provide intentional and systematic development of language skills. Pre-task activities involve learners in receptive multimodal input (including viewing documentary videos or reading infographics) and guided group discussions, which will prepare them to perform the main task in the target language, using the target language and modes. Scaffolding can be done using pre-teaching of vocabulary, graphic Organizers, and discussion guides.

In the task phase, learners plan, write a text, narrate, choose the visual elements, and coordinate the multimodal design. Group scripting, peer recording, and collaborative editing are all collaborative activities that provide rich opportunities for discussion and negotiation of meaning through speech. Post-task: In the post-task phase, learners share their multimodal texts, provide peer feedback, discuss the class about an effective multimodal design, and write reflective journals to help them remember what they learned and to develop metacognitive understanding of the composing process.

Dimension 5: Multimodal Assessment Design

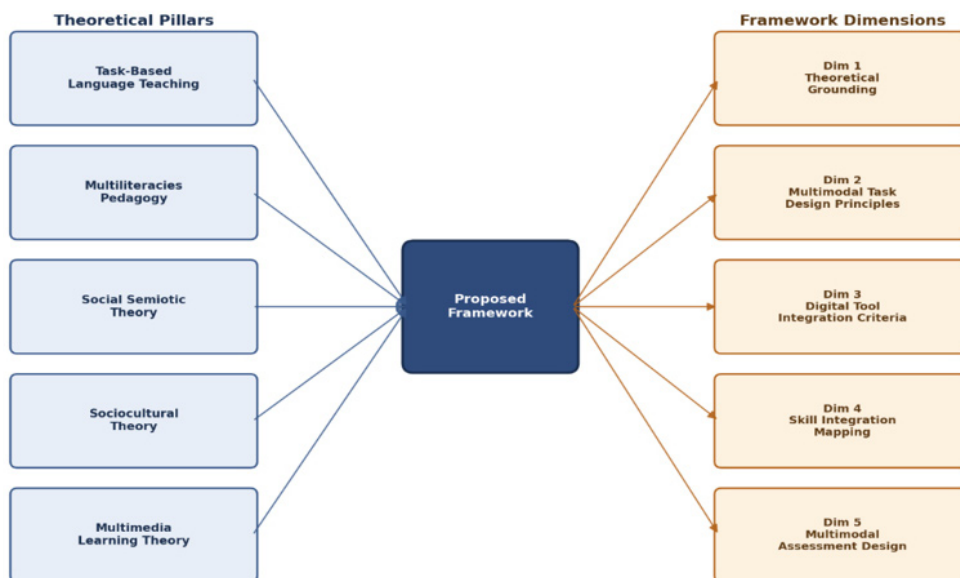
The 5th dimension is related to evaluating the results and operations of a multimodal digital task. A difficulty that arises in running the multimodal approach is the conflict between multimodal instruction and the traditional monomodal assessment; it is suggested here to use both product- and process-oriented criteria in the framework.

The assessment adopts product-oriented assessment as per Jiang et al. (2022), which focuses on five assessment criteria, namely linguistic quality, multimodal coherence, visual quality, design quality, communicative effectiveness, and critical awareness of the multimodal texts. Based on the work of Zhang and Yu (2022) and Yu et al. (2024), process-oriented assessment is made up of four categories: planning, collaborative interaction, revision practices, and reflective engagement, which are measured through digital portfolios, process journals, screen recordings, and rubrics. Further, there are suggestions to develop students' learning through self- and peer-assessment. It further recommends that the self-regulated learning (SLL) skills of the learners could be assessed by self- and peer-assessment (Fan,2025).

Figure 1 summarizes how the five theoretical pillars converge into the proposed framework and radiate outward into its five design dimensions.

Figure 1

Hub-and-spoke Conceptual Framework Linking Theoretical Pillars to Task-design Dimensions



Illustrative Task Design: A Multimodal Documentary Project

To show how the framework can be implemented in practice, this section includes a mini case study of a multimodal documentary project, which involves a team of four to five college-level EFL students researching, writing, filming, editing, and presenting a very short documentary video on a socially relevant topic within the space of about four weeks.

In the pre-task phase (Week 1), learners will watch model documentaries and discuss how these use narration, interviews, background music, text overlaid, and visual imagery to create meaning. They read academic texts, discuss and/or negotiate selection of topics, and are scaffolded by analytical frameworks, vocabulary lists, and discussion protocols. During the task phase (Weeks 2–3), the students engage in intensive work in composing in multimodal forms. Learners research information, develop narration and set up questions for an interview, record video, and edit with easy-to-use tools. The instructor provides continuous support through checkpoint conferences, model scripts, technical support, and co-peer support. There is a post-task phase (Week 4) in which students analyze the videos in written reflection papers and participate in discussions of the videos in small groups with their peers. In Assessment, the multimodal rubric outlined in Dimension 5 is used to assess the documentary product and the composing process through portfolio documentation. This task design exemplifies a task that integrates all five dimensions of the task design framework into a coherent instructional design.

Implications

Theoretical Contribution

The proposed model makes several contributions to the theory of multimodal digital task design in EFL education. First, it provides a theoretical framework that integrates the five theoretical traditions mentioned above, stepping in to fill the gap in multimodal language pedagogy regarding the need for a more comprehensive theoretical model (Lim & Kessler, 2023; Jiang & Hafner, 2024). The previous frameworks have tended to emphasize either task-based approaches or multimodal literacy approaches, with less attention given to articulating the relationship between these approaches. The current model illustrates how these views can be fruitfully combined to give a more comprehensive view of multimodal digital task design.

Second, it also addresses the identified gap in principled approaches to integrating skills into multimodal tasks. Although it has been well recognized in the literature that integrated skill development is the most desirable approach, few frameworks have offered a systematic way to develop and/or map out skill integration throughout a multimodal task cycle. The mapping dimension of skill integration, part of the present skill integration framework, provides a structured approach that can help educators go beyond intuitive, ad hoc approaches.

Thirdly, the framework builds on the extensive literature on multimodal assessment by suggesting an assessment model in line with the multimodal and integrated nature of the tasks. This study tackled the assessment challenges pointed out in the previous studies (Cheung, 2023; Fan, 2025) through the use of dual attention for process and product assessment, genre-based multimodal assessment rubrics, and self- and peer-assessment strategies.

Practical Implications

The framework has several implications in practice, all relating to different stakeholders in EFL education. The framework provides the groundwork for curriculum designers in developing curriculum-informed, pedagogically valid, and practically feasible multimodal digital tasks, and for language teachers in developing language-informed digital tasks, respectively. The five dimensions of the framework can be used as a checklist to make sure that the educators' tasks address the following: theoretical foundations, multimodal design principles, digital tool selection, skill integration, and assessment.

The multimodal digital task design is emphasized in teacher education programs with regard to in-service teachers' and pre-service teachers' comprehensive professional development. Teachers need to be mindful of the technical problems involved in multimodal composing and of relevant theories to create suitable pedagogy for teaching L2 multimodal composing. As Yu et al. (2024) argue, although there has been much research and practice in multimodal literacy, the majority of educational contexts tend to consider literacy mainly as text-based reading and writing, which limits teachers' ability to implement innovative solutions through these multimodal approaches. The framework is very much focused on the institutional and policy support needed so that educational policymakers can develop some pedagogical modes. Despite the numerous studies and practices of multimodal literacy, there is still a widespread misunderstanding that literacy is primarily a text-oriented process of reading and writing in most educational contexts, making it challenging for teachers to integrate new, innovative approaches to literacy in these multimodal ways.

Challenges and Contextual Considerations

The proposed framework presents some implementation challenges, and the contextual factors that can enable and/or limit its use should be considered. The digital divide – in terms of students' access to the technologies and internet connections needed for multimodal digital composing (Pankeaw & Satayaban, 2025)- is a major problem. Access to digital resources is not equal in many EFL contexts, particularly in developing countries, and continues to be an issue in TMP in the classroom. This concern is addressed within the framework via the digital tool integration criterion of accessibility and equity; however, it is important to note that this will have to be accompanied by institutional commitments to tackle resource inequality.

The other challenge is the preparedness of teachers. Multimodal assessment (Akhmetzadina et al., 2023). Most EFL teachers do not have training in multimodal pedagogy and digital tools. To be effective, there must be continuous professional development on the theoretical foundations of and practical skills for the multimodal design of digital tasks. The framework may serve as a tool to support professional development and as a blueprint for the new teacher(s) to this multimodal pedagogy.

Assessment is one of the areas that poses a special difficulty. The tasks in this framework will be used to evaluate multimodal and integrated language skills, while traditional assessment in many EFL contexts is still geared towards discrete language skills and standardized testing formats, which are not suited for the purpose of evaluating the integrated and multimodal language skills developed from the tasks in this framework (Cheung, 2023). Assessment policies should be flexible to allow the use of multimodal approaches, such as portfolios, peer assessment, and performance-based assessment, within institutional assessment policies. The assessment dimension of the framework is the beginning of the adaptation, and more institutional and policy support will be required for a system-level change.

Limitations

The framework is theoretically sound and is relevant to the subject in practice, but has some inherent limitations that need to be recognized. It is a conceptual framework and has not yet been tested empirically; the effectiveness of this framework in actual EFL college-level classrooms needs to be tested through experimental, quasi-experimental, or ethnographic studies. Despite the literature being current and extensive, it was largely drawn from Western and East Asian EFL contexts, which may affect the immediate transferability of the framework to under-resourced and/or linguistically diverse contexts such as those in South and Southeast Asia. Furthermore, it assumes that there is a minimum level of digital infrastructure and technological skills among teachers, which may not be the case in many EFL institutions worldwide. While based on theory, the choice of the five theoretical pillars also inevitably emphasizes some aspects of the task design of multimodal tasks and does not give equal weight to other aspects, such as genre-based pedagogy and complexity theory. Empirical research on the framework is required in the future to improve and validate it across different institutional and cultural contexts.

Conclusion

The present article proposes a comprehensive conceptual framework for designing multimodal digital tasks to develop EFL college students' integrated language skills. It combines five theoretical traditions, namely Task-Based Language Teaching (TBLT), multiliteracies pedagogy, social semiotic theory, sociocultural theory, and multimedia learning theory, into five interconnected dimensions of task design: theoretical grounding, multimodal task design principles, digital tool integration criteria, skill integration mapping, and multimodal assessment design.

It fills an existing gap in the literature by offering a systematic, principled approach to designing digital tasks focused on integrated skill development, where no detailed design framework has previously guided practice. The proposed framework gives educators, curriculum designers, teacher educators, and policymakers a basis for designing tasks for twenty-first-century college-level EFL learners. It also addresses key challenges of multimodal digital pedagogy, including the digital divide, teacher preparedness, assessment limitations, and cognitive load, through its focus on accessibility, progressive scaffolding, process-oriented design, and aligned assessment.

The framework will ultimately only be put into practice with institutional support, policy reform, and continued investment in teacher professional development. As communication continues to evolve through new technologies, norms, and modalities, well-defined frameworks like this one will become increasingly necessary for EFL didactics, offering both a theoretical and practical foundation for the next generation of multimodal digital language education.

Several avenues for future research follow from this framework. Empirical validation is the most immediate priority: quasi-experimental and mixed-methods studies could test the framework's tasks against traditional approaches, and longitudinal designs could examine how sustained engagement shapes learners' language development and multimodal competence over time. A second priority is exploring how generative AI might support task design, including as a scaffolding or co-composing partner, and the ethical questions this raises. Third, research on teacher cognition and professional learning, including participatory action research letting teachers adapt the framework to their own contexts, would clarify obstacles to implementation. Finally, the framework's assessment dimension remains under-tested empirically, and studies validating tools such as digital portfolios and technology-mediated peer assessment are needed across diverse EFL contexts.

References

- Ajabshir, Z. F. (2024). Empowering EFL writing through digital storytelling: A quasi-experimental assessment of CALF measures and multidimensional engagement. *Acta Psychologica*, 250, Article 104564. <https://doi.org/10.1016/j.actpsy.2024.104564>
- Akhmetzadina, Z. R., Mukhtarullina, A. R., Starodubtseva, E. A., Kozlova, M. N., & Pluzhnikova, Y. A. (2023). Review of effective methods of teaching a foreign language to university students in the framework of online distance learning: International experience. *Frontiers in Education*, 8, Article 1125458. <https://doi.org/10.3389/feduc.2023.1125458>
- Al-Jarf, R. (2024). Multimodal teaching and learning in the EFL college classroom. *Journal of English Language Teaching and Applied Linguistics*, 6(4), 55–72. <https://doi.org/10.32996/jeltal.2024.6.4.7>

- Alkhudiry, R. (2022). The contribution of Vygotsky's sociocultural theory in mediating L2 knowledge co-construction. *Theory and Practice in Language Studies*, 12(10), 2117–2123. <https://doi.org/10.17507/tpls.1210.19>
- Cheung, A. (2022). Verbal and on-screen peer interactions of EFL learners during multimodal collaborative writing: A multiple case-study. *Journal of Second Language Writing*, 58, Article 100931. <https://doi.org/10.1016/j.jslw.2022.100931>
- Cheung, A. (2023). Developing and evaluating a set of process and product-oriented classroom assessment rubrics for assessing digital multimodal collaborative writing in L2 classes. *Assessing Writing*, 56, Article 100723. <https://doi.org/10.1016/j.asw.2023.100723>
- Cho, H., & Kim, Y. (2024). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*, 28(6), 2487–2512. <https://doi.org/10.1177/13621688211046740>
- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6, Article 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.
- Fan, Y.-S. (2025). From text-centric to multimodal: Implementing multiliteracies and learning-oriented assessment in a first-year EFL curriculum. *RELC Journal*. <https://doi.org/10.1177/00336882251410729>
- González-Lloret, M. (2023). The road *System* travelled: Five decades of technology in language education. *System*, 118, Article 103124. <https://doi.org/10.1016/j.system.2023.103124>
- González-Lloret, M., & Ziegler, N. (2022). Technology-mediated task-based language teaching. In M. Ahmadian & M. Long (Eds.), *The Cambridge handbook of task-based language teaching* (pp. 326–345). Cambridge University Press.
- Hafner, C. A., & Ho, W. Y. J. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing*, 47, Article 100710. <https://doi.org/10.1016/j.jslw.2020.100710>
- Hasumi, T., & Chiu, M.S. (2024). Technology enhanced language learning in English language education: Performance analysis, core publications, and emerging trends. *Cogent Education*, 11(1), Article 2346044. <https://doi.org/10.1080/2331186X.2024.2346044>
- Jackson, D. O. (2022). *Task-based language teaching*. Cambridge University Press.
- Jiang, L., & Hafner, C. A. (2024). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching*, 58(4), 528–546.
- Jiang, L., Yu, S., & Lee, I. (2022). Developing a genre-based model for assessing digital multimodal composing in second language writing: Integrating theory

- with practice. *Journal of Second Language Writing*, 57, Article 100869. <https://doi.org/10.1016/j.jslw.2022.100869>
- Kalantzis, M., & Cope, B. (2025). Literacy in the time of artificial intelligence. *Reading Research Quarterly*, 60(1), Article e591. <https://doi.org/10.1002/rrq.591>
- Kang, S., & Kim, Y. (2023). EFL adolescent learners' perceptions of digital multimodal composing tasks and task outcomes: A longitudinal study. *International Journal of Applied Linguistics*, 33(2), 340–361. <https://doi.org/10.1111/ijal.12473>
- Kim, Y., & Namkung, Y. (2024). Methodological characteristics in technology-mediated task-based language teaching research: Current practices and future directions. *Annual Review of Applied Linguistics*, 44, 56–78. <https://doi.org/10.1017/S0267190524000096>
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge.
- Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education*. Routledge.
- Lim, F. V. (2020). *Designing learning with embodied teaching: Perspectives from multimodality*. Routledge. <https://doi.org/10.4324/9780429353178>
- Lim, F. V., & Nguyen, T. T. H. (2022). 'If you have the freedom, you don't need to even think hard' – Considerations in designing for student agency through digital multimodal composing in the language classroom. *Language and Education*, 37(1), 1–19. <https://doi.org/10.1080/09500782.2022.2107875>
- Lim, J., & Kessler, M. (2023). Multimodal composing and second language acquisition. *Language Teaching*, 57(2). <https://doi.org/10.1017/S0261444823000125>
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60–92. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Pankeaw, P., & Satayaban, N. (2025). Task-based language teaching (TBLT) in the digital age. *Interdisciplinary Academic and Research Journal*, 5(3), 153–172. <https://doi.org/10.60027/iarj.2025.281796>
- Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Llanes, A., & Serrano, R. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, 42(3), 577–598. <https://doi.org/10.1017/S0272263120000091>
- Rahmanu, I. W. E. D., & Molnár, G. (2024). Multimodal immersion in English language learning in higher education: A systematic review. *Heliyon*, 10(19), Article e38357. <https://doi.org/10.1016/j.heliyon.2024.e38357>
- Sarmiento-Campos, N.-V., Lázaro-Guillermo, J. C., Silvera-Alarcón, E.-N., Cuellar-Quispe, S., Huamán-Romaní, Y.-L., Apaza, O. A., & Sorkheh, A. (2022). A

- look at Vygotsky's sociocultural theory (SCT): The effectiveness of scaffolding method on EFL learners' speaking achievement. *Education Research International*, 2022, Article 3514892. <https://doi.org/10.1155/2022/3514892>
- Usmani, S., Ali, E., & Kottaparamban, M. (2025). The impact of digital storytelling on EFL learners' speaking and writing skills. *Forum for Linguistic Studies*, 7(4), Article 816. <https://doi.org/10.59400/fls.v7i4.816>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Walqui, A. (2006). Scaffolding instruction for English language learners: A conceptual framework. *International Journal of Bilingual Education and Bilingualism*, 9(2), 159–180. <https://doi.org/10.1080/13670050608668639>
- Xie, C., & Lan, G. (2025). The effectiveness of task-based language teaching in improving communicative competence. *International Journal of Advanced and Applied Sciences*, 12(1), 263–273. <https://doi.org/10.21833/ijaas.2025.01.025>
- Yap, J. R., & Gurney, L. (2023). Exploring practices of multiliteracies pedagogy through digital technologies: A narrative inquiry. *Literacy*, 57(3), 292–305.
- Yu, S., Zhang, E. D., & Liu, C. (2024). Research into practice: Digital multimodal composition in second language writing. *Language Teaching*, 59(2), 197–213.
- Zapata, G. C., Kalantzis, M., & Cope, B. (2023). *Multiliteracies in international educational contexts: Towards education justice*. Routledge. <https://doi.org/10.4324/9781003349662>
- Zhang, E. D., & Yu, S. (2022). Implementing digital multimodal composing in L2 writing instruction: A focus on developing L2 student writers. *Innovation in Language Learning and Teaching*, 17(4), 769–777. <https://doi.org/10.1080/17501229.2022.2135712>
- Zhao, Y., & Watterston, J. (2021). The changes we need: Education post COVID-19. *Journal of Educational Change*, 22(1), 3–12. <https://doi.org/10.1007/s10833-021-09417-3>