Akademika 96(1), April 2026
https://doi.org/10.17576/akad-2026-9601-18

Digital Storytelling in English Language Teaching (2015–2025): A Narrative Review of Outcomes and Design Principles

Penceritaan Digital dalam Pengajaran Bahasa Inggeris (2015–2025): Ulasan Naratif tentang Hasil dan Prinsip Reka Bentuk

ROFIZA ABOO BAKAR, NOOR AZLI AFFENDY LEE*, RENI KUSUMANINGPUTRI & LATISHA ASMAAK SHAFIE

Received: 1-9-2025 /Accepted: 31-1-2026

ABSTRACT

Digital storytelling (DST) has matured from a novelty technology to a principled pedagogy within English language teaching (ELT). This narrative review synthesises twelve empirical evidences published between 2015 and 2025 across Web of Science, Scopus, and ScienceDirect to address two questions: what effects does DST have on core language outcomes and affect in ELT, and which implementation features are associated with stronger results? Drawing on classroom-based experimental, quasi-experimental, mixed-methods, design-based and analytical observational studies, we find convergent evidence that DST improves speaking performance while reducing speaking anxiety and strengthening learners' self-regulation. Writing gains are likewise robust and more fine-grained when assessed through CALF: complexity, accuracy, lexical diversity, and fluency; particularly in courses that stage DST as a multi-week composing cycle. Vocabulary growth is most consistent when stories are learner-generated, and target words are embedded in narration and peer review. Design clearly matters: scripted collaboration (with explicit roles and checkpoints) and, in some contexts, social-embodied facilitation (e.g., robots) amplify engagement and learning beyond "tool-first" DST. Persistent gaps include limited attention to listening and reading outcomes, infrequent delayed post-tests beyond one month, and uneven reporting of duration and assessment anchors, all of which constrain cross-study comparability. The researchers conclude with implications for course and assessment design, teacher education, and technology selection, and we outline priorities for longitudinal research and broader skill coverage to consolidate DST's value proposition in ELT.

Keywords: English language education; digital storytelling; teacher education; narrative review

ABSTRAK

Penceritaan digital (DST) telah berkembang daripada teknologi baharu kepada pedagogi yang berasaskan prinsip dalam pengajaran bahasa Inggeris (ELT). Ulasan naratif ini mensintesis dua belas bukti empirikal yang diterbitkan antara 2015 hingga 2025 merentas pangkalan data Web of Science, Scopus dan ScienceDirect untuk menjawab dua persoalan: apakah kesan DST terhadap hasil pembelajaran teras dan aspek afektif dalam ELT, dan ciri reka bentuk manakah yang dikaitkan dengan keputusan yang lebih kukuh? Bertitik tolak daripada kajian bilik darjah yang berbentuk eksperimen, kuasi-eksperimen, kaedah campuran, berasaskan reka bentuk serta pemerhatian analitik, dapatan menunjukkan DST meningkatkan prestasi pertuturan sambil mengurangkan kebimbangan bertutur dan mengukuhkan keupayaan kendiri pelajar. Peningkatan penulisan turut konsisten dan lebih terperinci apabila dinilai menggunakan kerangka CALF: kompleksiti, ketepatan, kepelbagaian leksikal dan kefasihan; terutamanya apabila DST dilaksanakan sebagai kitaran penggubahan berfasa. Perkembangan kosa kata paling ketara apabila cerita dihasilkan sendiri oleh pelajar dan perkataan sasaran diintegrasikan dalam naratif serta semakan rakan sebaya. Reka bentuk pelaksanaan amat penting: kolaborasi berskrip (peranan dan titik semak yang jelas) serta fasilitasi berperantara sosial (contohnya robot) dapat meningkatkan penglibatan dan pembelajaran melebihi pendekatan "teknologi dahulu". Kekangan yang dikenal pasti termasuk perhatian terhad terhadap kemahiran mendengar dan membaca, ketiadaan ujian tertunda melebihi sebulan, serta pelaporan yang tidak seragam. Implikasi dibincangkan untuk reka bentuk kursus dan pentaksiran, pendidikan guru, pemilihan teknologi, serta keutamaan penyelidikan longitudinal dan peluasan domain kemahiran.

Kata kunci: pengajaran bahasa Inggeris; penceritaan digital; pendidikan guru; ulasan naratif

INTRODUCTION

Digital storytelling (DST) integrates narrative composition with digital media production (audio, images, video, animation), positioning learners as authors and designers. In English language teaching (ELT), DST has evolved into a robust pedagogical tool that fosters multimodal literacies, learner autonomy, and communication skills. To evaluate the pedagogical potential of DST across diverse ELT contexts, a narrative review was chosen for its ability to accommodate diverse methodologies while allowing thematic synthesis and conceptual development (Baumeister & Leary, 1997; Snyder, 2019). Recent research emphasizes the relevance of digital storytelling not just as a student-centred practice, but as a sociocultural activity that fosters learner identity construction, emotional engagement, and multimodal meaning-making (Hung, 2019; Nair & Yunus, 2021). In multilingual classrooms, DST has been observed to bridge language barriers and enhance intercultural communication by allowing learners to share personal narratives using voice, visuals, and symbolic imagery (Yu & Wang, 2025).

As DST becomes more embedded in curriculum design, understanding how it affects not only linguistic competencies but also learner agency and digital fluency is critical (Adnan et al., 2024). Narrative reviews that centre the learner's perspective and pedagogical design can capture these broader educational implications. Unlike systematic reviews, which aim for exhaustive coverage and meta-analysis, narrative reviews are particularly suited to emerging or heterogeneous fields like technology-enhanced language learning, where study designs, interventions, and outcomes vary significantly (Ferrari, 2015; Xiao & Watson, 2019).

This paper synthesizes research published between 2015 and 2025 to assess DST's effects on core language outcomes (speaking, writing, vocabulary), affective variables (anxiety, self-regulation), and to identify design principles that moderate its effectiveness. By applying narrative review techniques, this study aims to consolidate findings, clarify pedagogical implications, and offer an agenda for future research.

A design-sensitive view of DST also aligns with contemporary theories of multimedia learning and cognitive load: when narration, imagery, and pacing are sequenced to support coherence and signalling, learners can allocate resources to language form and discourse rather than to interface management. Recent classroom trials report that well-scaffolded DST tasks neither overload novices nor dilute language focus, provided that rehearsal, role clarity, and feedback checkpoints are built into the cycle (Tatlı, Saylan, & Kokoç, 2022). In parallel, the newer literature foregrounds identity and agency: learners' authorial control such as choosing topics, framing audiences, and shaping performance. This appears to mediate anxiety and self-regulation effects, now demonstrated with randomized evidence in high-stakes speaking contexts (Bai & Xian, 2024). Equally notable is the shift from "tool-first" enthusiasm to principled design: studies that embed collaboration scripts, plan-to-prose scaffolds, and explicit CALF targets report stronger and more transferable gains in writing and vocabulary (Ajabshir, 2024; Chen & Yeh, 2025; Nami & Asadnia, 2024). Finally, recent work connects DST to digital citizenship and multimodal literacy: curated image–text alignment and citation of media assets become teachable moments for ethics and academic discourse, with Malaysian evidence showing that visual-verbal tasks can cultivate deliberate lexical choice, an affordance that ports directly into DST production (Adnan, Husin, Hashim, & Setia, 2024).

AIM AND RESEARCH QUESTONS

This narrative review aims to synthesise empirical research on digital storytelling (DST) in English language teaching (ELT) published between 2015 and 2025 to evaluate its pedagogical effectiveness and design implications. The research questions are the following:

1) How does DST influence core ELT outcomes, particularly speaking, writing, vocabulary, and affective factors such as anxiety and self-regulation?
2) Which implementation features (e.g., collaboration scripts, storyboard scaffolds, social-embodied facilitation) contribute most to DST effectiveness?
3) What research gaps persist in the DST-ELT literature, and how might these inform future studies and pedagogical design?

LITERATURE REVIEW

Literature reviews in applied fields like ELT play a critical role in consolidating fragmented empirical findings, identifying design principles, and clarifying gaps for future study (Paré et al., 2015). Narrative reviews, in particular, allow for qualitative synthesis across diverse study designs, including quasi-experimental, mixed-methods, and observational research, making them ideal for investigating complex educational interventions like digital storytelling (Snyder, 2019). Within second-language theory, a review of conceptual transfer in SLA consolidates how L1–L2 conceptual mappings shape production, context that helps explain why DST's story-grammar and audience design can trigger form–meaning reorganization (Xu, Mamat, & Zin, 2024).

Compared to systematic reviews, which emphasize replicable protocols, comprehensive database coverage, and formal assessment of bias (Page et al., 2021; Xiao & Watson, 2019), the narrative review approach is more interpretive and flexible (Baumeister & Leary, 1997; Ferrari, 2015; Snyder, 2019). This flexibility is necessary in an emerging field like DST, where studies vary widely in design, intervention duration, assessment tools, and skill focus, making meta-analysis or rigid protocol-driven synthesis less feasible (Paré et al., 2015; Snyder, 2019). However, narrative reviews must also be approached with caution: their strength in interpretation can create vulnerability to selection bias, uneven weighting of evidence, and limited replicability (Ferrari, 2015; Xiao & Watson, 2019). Scoping reviews, on the other hand, map the breadth of a field but seldom offer the conceptual integration needed to understand pedagogical mechanisms (Paré et al, 2015). By adopting a narrative review, this study prioritizes depth of interpretation and cross-study conceptual cohesion, while acknowledging that its conclusions are contextually grounded rather than generalizable. This critical positioning clarifies why narrative synthesis is appropriate here and also highlights the methodological boundaries within which this review should be interpreted.

Several studies have documented significant improvements in speaking fluency, coherence, and anxiety reduction when DST is integrated into classroom practice. For example, in a randomized controlled trial with IELTS learners, DST-based instruction led to sustained speaking gains and reduced anxiety over a one-month follow-up period (Bai & Xian, 2024). Pre-service teachers also showed improved academic speaking skills through DST-enhanced public presentation tasks (Roza et al., 2023). Online implementations yielded similar benefits in fluency and vocabulary with minimal cognitive overload (Tatlı et al., 2022).

Where synthesis articles previously emphasised broad motivational benefits (e.g., pre-2020 reviews), the post-2020 corpus is substantively different in rigour and granularity. A PRISMA-guided review of DST for speaking, for instance, tallied consistent positive effects on oral performance alongside anxiety reduction, but also called for better reporting of treatment fidelity and delayed measures (Nair & Yunus, 2021). Subsequent primary studies have begun to answer that call by reporting subskill-level outcomes (fluency, lexical range, pronunciation) and by testing maintenance at one month or beyond, which strengthens causal interpretations (Bai & Xian, 2024). These design upgrades matter for curriculum transfer: when DST is mapped onto recognised descriptors (e.g., IELTS-style bands) and assessed with transparent rubrics, teachers can plan for skill growth rather than hope for incidental gains.

Digital storytelling improves writing quality, especially when measured through CALF (complexity, accuracy, lexical diversity, and fluency). In a quasi-experimental study, Ajabshir (2024) found that DST significantly outperformed traditional storytelling in improving all CALF dimensions and learner engagement. Scripted collaboration, where learners are assigned specific peer roles, also led to more creative and accurate writing compared to standard DST approaches (Chen & Yeh, 2025).

Vocabulary acquisition is most effective when target words are embedded in student-generated stories and peer review tasks. In a large university study, Nami and Asadnia (2024) showed that collaborative DST significantly improved vocabulary knowledge over a control group. These findings support the use of multimodal DST to create meaningful contexts for lexical elaboration.

Robotic or embodied facilitation in DST has been linked to heightened learner engagement and narrative coherence. Liang and Hwang (2023) demonstrated that students using robot-assisted DST showed higher narrative engagement and multimodal fluency than those using standard animation tools.

In teacher education, DST has been shown to improve digital competence, collaboration, and pedagogical design. Yu and Wang (2025) found that pre-service teachers developed stronger digital and collaborative skills after producing teachable DST packages.

Several systematic and scoping reviews have examined technology-enhanced language learning, but few focus specifically on DST in ELT. For example, Nair and Yunus (2021) identified general trends in DST use but called for deeper synthesis on the quality of design principles. Hung (2019), using a process-tracing study, noted that the sequencing of pre-writing, digital editing, and post-reflection influenced how learners develop multiliteracies and voice. These studies suggest that DST's efficacy depends not only on the tools used but also on the coherence of pedagogical design.

A complementary strand examines mechanisms rather than outcomes alone. Evidence synthesized across recent classroom studies indicates that DST supports metacognitive and motivational regulation—learners plan, monitor, and evaluate their performances over iterative drafts—while social presence (peer audiences, and in some cases robots) scaffolds engagement and persistence (Liang & Hwang, 2023). Emerging analyses argue that these regulatory benefits are not epiphenomenal; they are central to why DST "sticks" beyond a single assignment and why well-structured projects can sustain improvement across weeks (see also broader educational syntheses linking DST to (meta)cognitive, emotional, and behavioural regulation).

Our review complements this by foregrounding implementation features that have demonstrable effects on language learning outcomes and by critically mapping under-explored areas such as listening, reading, and teacher transfer. Furthermore, we align with methodological

frameworks in narrative synthesis that call for transparency in inclusion criteria, saturation of thematic coding, and integration of both outcomes and processes (Paré et al., 2015; Page et al., 2021).

Despite the growing body of work on DST in ELT, the existing literature remains fragmented across skills, contexts, and methodological approaches. Much of the research is concentrated on speaking, writing, and vocabulary, with far fewer studies examining listening, reading, or long-term learning trajectories. In addition, inconsistencies in reporting intervention duration, assessment procedures, and treatment fidelity limit the comparability of findings across studies. Recent reviews also note that many DST interventions prioritise product quality over language-focused processes, leaving the mechanisms that drive improvement under-specified. These gaps underscore the need for a consolidated narrative synthesis capable of drawing together dispersed evidence, clarifying patterns, and identifying design principles that warrant further exploration.

METHODOLOGY

A narrative review methodology was adopted for this paper to allow for interpretive depth and thematic synthesis across a diverse body of literature. This approach aligns with best practices for narrative reviews in education and applied linguistics, where the diversity of interventions and assessment strategies often precludes quantitative meta-analysis (Baumeister & Leary, 1997; Xiao & Watson, 2019). Narrative reviews are appropriate when the aim is to interpret and integrate findings conceptually, rather than aggregate effect sizes (Ferrari, 2015).

Following the process outlined by Paré et al. (2015), the review followed six structured steps: (1) problem formulation; (2) literature search (via Scopus, Web of Science, and ScienceDirect); (3) inclusion/exclusion screening; (4) data extraction on participants, interventions, outcomes, and design features; (5) thematic synthesis of findings; and (6) critical reflection on gaps and research directions. The decision to include empirical studies across mixed methods, quasi-experimental, and design-based paradigms reflects the methodological diversity of the DST-ELT field.

Searches were carried out in Web of Science, Scopus, and ScienceDirect for publications between 2015 and 2025. For this review, digital storytelling (DST) is defined as a learner-centred, multimodal narrative process in which students create short stories by integrating written or spoken text with digital media elements such as images, audio, video, or animation. Eligible studies had to include (a) learner-generated or co-constructed narratives, (b) a digital production phase using technological tools, and (c) a clear ELT-related outcome that could be traced to the storytelling task. Studies were excluded if they used teacher-created stories only, used video without a narrative component, or implemented digital tools without requiring narrative construction. This working definition guided all inclusion and exclusion decisions during screening. Search strings combined "digital storytelling" and "DST" with terms related to ELT, EFL, ESL, and specific skills such as speaking, writing, vocabulary, listening, and reading. Studies were included if they involved ELT learners or teachers, applied empirical or mixed-methods designs, and reported measurable outcomes related to language or affect. Exclusions were made for non-ELT contexts, conceptual essays, theses, and anecdotal classroom reports. After deduplication and screening, eligible studies were examined in full. Data were extracted on context, participants, design, intervention features,

and outcomes. The results were then synthesised thematically to trace patterns and identify influential design elements.

Two coders independently conducted the title and abstract screening, full-text assessment, and thematic grouping of studies. To ensure consistency, both coders first piloted the inclusion and exclusion criteria on a subset of records and refined the guidelines through discussion. Inter-coder reliability was assessed using Cohen's kappa, yielding values of .82 for screening decisions and .86 for thematic grouping, indicating strong agreement. Any discrepancies were resolved through discussion until full consensus was achieved. To improve interpretability, we documented treatment fidelity (presence of scripts, storyboard stages, and revision passes), time-on-task parity with comparison groups, and the specificity of outcome measures (e.g., reporting CALF computation details, inter-rater reliability for writing ratings, and named scales for anxiety/self-regulation). We also noted whether studies included delayed post-tests, given their rarity in earlier work. Following PRISMA 2020 reporting logic, we logged reasons for exclusion at full text (e.g., DST not central to the intervention; ELT context ambiguous; outcomes not language-linked), and we coded delivery mode (in-person/online), education level, and DST modality to enable contextual subgroup reads (Page et al., 2021).

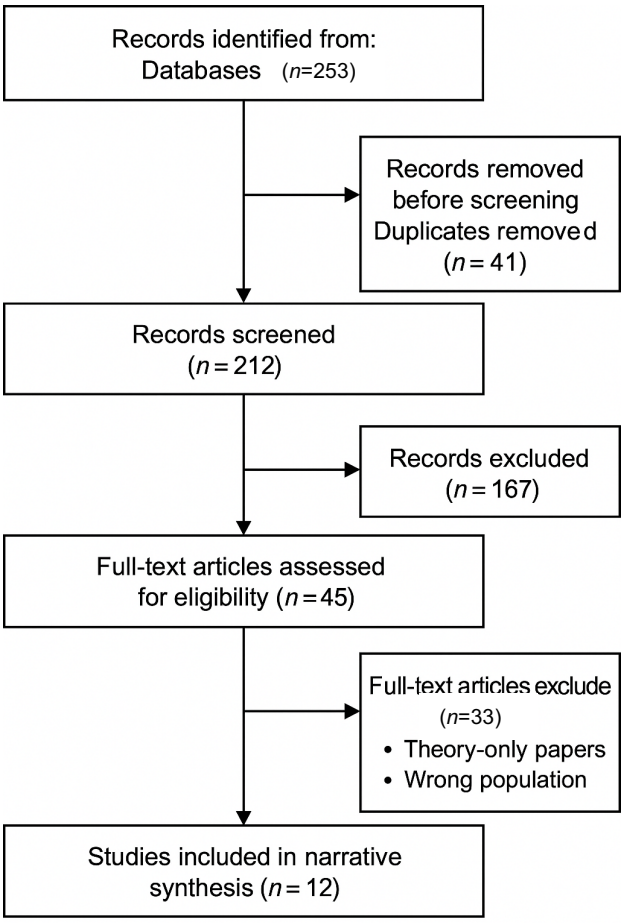

FIGURE 1. PRISMA Flow Diagram of Study Selection Process

The study selection process followed PRISMA 2020 guidelines for transparent reporting of literature reviews (Page et al., 2021). As shown in FIGURE 1, a total of 253 records were identified through database searching (Scopus, Web of Science, and ScienceDirect). After removing 41 duplicates, 212 records were screened by title and abstract. Of these, 167 were excluded for not meeting relevance criteria (for example, not empirical, not ELT or DST-focused). Full-text assessment was conducted on 45 articles, with 33 excluded for being theory-only, unrelated to DST, or not situated in English language teaching. A total of 12 studies were included in the final narrative synthesis. This process ensured that only empirical, classroom-based research related to DST in ELT contexts from 2015–2025 was retained for thematic analysis.

FINDINGS

A summary of the empirical studies included in this review, including details on sample, design, skills addressed, DST modality, and main findings, is presented in Table 1 (see Appendix A). It summarizes key attributes such as research context, participant characteristics, study design, DST modalities, and main findings. The synthesis that follows is organised thematically by outcome domain, highlighting trends and noteworthy variations across speaking, writing, vocabulary, teacher education, and under-represented skills.

SPEAKING PERFORMANCE, ANXIETY, AND SELF-REGULATION

Across the 2015–2025 corpus, digital storytelling produces consistent, practically meaningful gains in oral proficiency. The strongest evidence comes from designs that combine repeated rehearsal with public sharing: learners plan a script, record versions iteratively, solicit feedback, and present to an audience beyond the teacher. This cycle appears to drive improvements in fluency, vocabulary deployment, and pronunciation, while at the same time reducing speaking anxiety. Studies by Roza et al. (2023), Tatlı et al. (2022), and Eissa (2019) provide supporting evidence of improved fluency, coherence, and engagement under DST-enhanced speaking tasks. In an IELTS preparation course, Bai and Xian (2024) found that DST-based instruction significantly improved speaking and self-regulation and reduced anxiety, with effects persisting at one-month follow-up. Importantly, reductions in anxiety are not merely incidental; they are linked to clearer task structure, a sense of authorship over content, and opportunities to rehearse before summative assessment.

When follow-up measures are included, gains in speaking performance and self-regulation can persist for at least a month, suggesting that DST supports not only immediate performance but also the strategic routines that underwrite sustained oral development. Effect sizes vary by context. Larger benefits are reported where DST tasks are tightly integrated with course outcomes (e.g., IELTS-style descriptors) and where feedback cycles are explicit. Smaller or mixed effects occur when DST is added as a one-off media assignment without clear performance criteria or when the final product overshadows oral practice time. Liang and Hwang (2023) report that robot-assisted prompting heightened social presence and rehearsal quality, which helped reduce learner apprehension, a benefit that is pedagogically replicable with peer audiences when robotics are unavailable.

Across tertiary and exam-prep settings, the rehearse–record–revise–rehearse loop appears to be the hinge: it gives learners time to stabilise formulaic language for organization (e.g., discourse markers), upgrade lexis under peer pressure, and calibrate prosody for audience. In the RCT with IELTS candidates, these micro-processes likely underpin the observed gains in self-regulation, which, in turn, track reductions in speaking anxiety, a virtuous cycle that sustained improvements at the one-month check (Bai & Xian, 2024). Similar patterns show up in mixed-methods classroom projects, where qualitative accounts attribute fluency gains to iterative tries in low-stakes drafts, followed by public sharing that "locks in" performance (Roza et al., 2023; Tatlı et al., 2022). Robot-assisted facilitation makes the audience tangible and prompts more turns, but the pedagogical analogue is a well-briefed peer panel; both mechanisms increase accountability and rehearsal density (Liang & Hwang, 2023).

WRITING DEVELOPMENT AND CALF METRICS

Evidence for writing improvement is strongest when outcomes are reported at the level of CALF: complexity, accuracy, lexical diversity, and, to a lesser extent, fluency. Ajabshir (2024) found significant CALF improvements in DST groups compared to traditional storytelling. Scripted collaboration led to more structured revisions and greater lexical diversity, as shown by Chen and Yeh (2025). DST appears to raise syntactic complexity and accuracy by shifting attention from sentence-level correctness during drafting to discourse-level coherence during planning and revision. Learners who storyboard their narratives first, then map segments to paragraphs, are better able to maintain thematic progression and coherent referencing across the text.

Lexical diversity benefits from the need to match language to scene, image, or soundtrack, which pushes students to select and recycle domain-appropriate vocabulary. Fluency results vary across studies, partly because fluency is sensitive to drafting time and because some courses prioritise multiple revision passes that trade speed for quality. A recurring pattern is that scripted collaboration, with named roles (e.g., language lead, editor, narrator) and checkpoints for peer CALF checks, amplifies writing gains relative to free-form group work. Where rubrics make CALF expectations explicit and peer feedback is anchored to those criteria, students demonstrate clearer uptake of form-focused advice without sacrificing narrative quality. Null or small writing effects are usually associated with minimal scaffolding, brief project durations, or assessments that privilege visual polish over textual development.

Two design decisions consistently magnify writing effects. First, specifying how CALF will be measured, example using clause-level T-units for complexity, error-free T-units for accuracy, moving-average type–token ratio for lexical diversity, and words per minute for fluency as it helps learners align goals with revision behaviour. Second, coupling peer review with automated corrective feedback (ACF) at key milestones increases accuracy and speeds up uptake: recent Malaysian findings show that Grammarly/QuillBot pipelines yield more consistent, error-specific corrections than instructor-only cycles, making them credible adjuncts within DST drafting rounds (Radin & Adam, 2024). When these choices are in place, studies report larger gains in syntactic complexity and accuracy without depressing creativity scores, a balance echoed in scripted-collaboration designs that distribute roles like language lead, editor, and narrator (Chen & Yeh, 2025).

Complementing these CALF-aligned gains, a recent study comparing Grammarly and QuillBot with Malaysian ESL writers reported measurable accuracy improvements and more consistent error-specific feedback uptake than instructor-only feedback, underscoring the value of automated corrective feedback when DST drafts cycle through multiple revisions (Radin & Adam, 2024).

VOCABULARY LEARNING AND MULTIMODAL LITERACIES

DST can be a productive context for vocabulary growth, especially when stories are genuinely learner-generated and target lexis is operationalised in the task. In a university EFL setting, Nami and Asadnia (2024) demonstrated that learner-generated DST, coupled with collaborative scaffolds, resulted in significantly greater vocabulary retention than control conditions. Tatlı et al. (2022) likewise reported that vocabulary improved in EFL contexts when learners engaged with online DST tasks. The most reliable designs require explicit integration of new words in the narration, provide glossing/definition checks at storyboard stage, and include peer review focused on appropriateness and variety of lexical choices.

This combination encourages elaborative processing (linking form, meaning, and use) and repeated retrieval across planning, recording, and revision. Gains are stronger when vocabulary is tied to communicative intentions within the story rather than appended as a list to be memorised. Multimodal storytelling also builds broader literacies: learners learn to align image, pace, soundtrack, and narration, which in turn shapes prosody and discourse markers in the spoken track. However, there are pitfalls. Without constraints, students may prioritise visual effects over depth of language use, or recycle a narrow set of familiar words. Requiring a pre-production lexical plan, limiting stock phrases, and using post-production transcripts for lexical analysis can mitigate these risks. Short delayed quizzes two to four weeks after submission help establish whether lexical gains endure beyond the production window; too few studies implement this check, which limits claims about retention. Related multimodal work shows that digital infographics tasks promote academic discourse moves and intentional lexical selection in language classrooms, reinforcing our claim that visual-verbal alignment can deepen vocabulary elaboration within DST projects (Adnan et al., 2024).

Vocabulary growth is strongest when tasks force retrieval at multiple points: targeted lexis appears first in storyboards (with glossing), resurfaces in narration scripts, and is then audited in transcripts against a checklist (coverage, variation, collocation). Under these conditions, collaborative, learner-generated DS has produced significantly larger vocabulary gains than comparison tasks in university EFL (Nami & Asadnia, 2024), a result echoed in online DST implementations that reported manageable cognitive load alongside improved speaking/vocabulary (Tatlı et al., 2022). To validate durability, brief delayed quizzes at 2–4 weeks should be standard; too few studies run them, which flattens claims about retention. Visual-verbal alignment tasks such as digital infographics also train lexical selection for academic moves and can be staged upstream of DST to prime higher-yield word choices during production (Adnan et al., 2024).

TEACHER-EDUCATION OUTCOMES AND CLASSROOM TRANSFER

Within preservice teacher education, DST projects function as design-practice tasks that integrate language pedagogy with digital and collaborative competence. In their study, Yu and Wang (2025) showed that DST-based instructional design projects for pre-service teachers led to measurable gains in digital competence, collaboration, and reflective practice. When teacher candidates must produce a teachable storytelling package, complete with learning objectives, story grammar, language targets, and an assessment plan, they report sharper awareness of how multimodal choices support language goals. Collaborative production obliges candidates to negotiate roles, manage assets, and justify sequencing decisions; these are the same project-management skills they will need in schools.

Importantly, DST in teacher education appears to surface tacit assumptions about audience, authenticity, and assessment: candidates become more explicit about the difference between an attractive video and a language-learning artefact. Reported gains include self-efficacy with technology, clearer pedagogical rationales for integrating media, and improved collaboration practices. Yet, transfer is not automatic. Candidates placed in schools without resources or institutional buy-in struggle to apply DST, and some revert to traditional tasks when classroom time is tight. Programmes that pair DST production with school-based micro-teaching and provide templates for low-tech implementation report higher rates of classroom uptake. Echoing our emphasis on practicum-ready design, a literature review by Xu, Mamat and Zin (2024) links teachers' digital competence, targeted training, and effective remote-teaching practices, evidence that the competencies rehearsed in DST capstones map onto broader professional standards.

Programs reporting the smoothest classroom transfer pair campus-based DST production with school-based micro-teaching, require candidates to prepare low-tech variants of their projects, and include explicit mapping to digital-competence frameworks used in local systems. In such designs, candidates articulate why a given multimodal choice serves a language objective, how to run the cycle under time and device constraints, and how to assess both product and process. Recent teacher-education research corroborates that these experiences lift digital competence, collaborative planning, and confidence with remote or blended delivery, competencies increasingly codified in professional standards (Xu, Mamat, & Zin, 2024; Yu & Wang, 2025).

UNDER-REPRESENTED OUTCOMES, EQUITY, AND REPORTING QUALITY

Listening and reading remain the least studied language outcomes in DST research. Where listening is assessed, it is often as a by-product of editing and rehearsal rather than as a targeted skill with validated measures. Rahimi and Yadollahi (2017) represent one of the few studies that directly compared online and offline DST formats on reading and writing, reporting that online DST resulted in greater literacy gains in lower-intermediate EFL learners. However, reading development is rarely isolated from writing in DST tasks. Future investigations should align receptive-skill assessments with recognised descriptors and ensure time-on-task parity with control conditions.

Equity and access also require more attention. High-engagement variants such as robot-assisted facilitation or professional editing suites are attractive but not essential; the evidence suggests that carefully scripted collaboration, clear language targets, and iterative feedback deliver most of the benefits even in low-tech settings. Finally, uneven reporting of duration, sampling

frames, and rubric anchors continues to constrain synthesis. Studies that state CALF computation details, inter-rater reliability, and delayed testing intervals contribute disproportionately to cumulative knowledge and should be treated as publication standards for the field.

Receptive skills remain a priority for the next wave of work. The few studies that target reading within DST report literacy benefits when production is preceded by guided reading of model texts and followed by transcript-based analysis; a 2017 benchmark comparison also suggested that online DST may outperform offline variants for reading–writing integration (Rahimi & Yadollahi, 2017). Listening, too, is typically incidental (through editing/rehearsal) rather than intentional; future designs should adopt validated listening measures and ensure time-on-task parity with controls. On equity, robot-assisted DST showcases what is possible, but most of DST's learning yield appears to come from scripts, scaffolds, and feedback, which are implementable with basic devices, a critical point for low-resource schools (Liang & Hwang, 2023).

DISCUSSION

Taken together, the studies confirm DST's potential as a design-sensitive pedagogy in ELT. Improvements in speaking, writing, and vocabulary are consistent across contexts, but the degree of impact is contingent on how tasks are structured. Scripted collaboration has emerged as a particularly powerful design feature, echoing socio-constructivist theories that emphasise interdependence and role clarity in group learning (Chen & Yeh, 2025). Similarly, the integration of rehearsal and authentic audience engagement explains why DST is effective in lowering anxiety and improving self-regulation (Bai & Xian, 2024; Roza et al., 2023). Robot-assisted DST may further enhance narrative coherence and engagement in settings where such tools are available (Liang & Hwang, 2023).

In response to research question 1, the synthesis confirms that digital storytelling (DST) consistently enhances core English language teaching (ELT) outcomes, notably speaking, writing, and vocabulary, while also reducing anxiety and supporting learner self-regulation. These benefits are observed across diverse contexts and designs, suggesting that DST is a robust design-sensitive pedagogy in ELT. However, the degree of impact depends strongly on task structure. Scripted collaboration has emerged as a particularly powerful design feature, echoing socio-constructivist principles that emphasise interdependence and role clarity in group learning (Chen & Yeh, 2025). Likewise, structured rehearsal and authentic audience engagement explain why DST effectively reduces speaking anxiety and improves self-regulation (Bai & Xian, 2024; Roza et al., 2023). Socially embodied facilitation, such as robot-assisted storytelling, may further amplify engagement and narrative coherence where feasible (Liang & Hwang, 2023).

Regarding research question 2, the review identifies several implementation features that moderate DST effectiveness, including iterative rehearsal cycles, storyboard-to-draft scaffolds, explicit CALF-based rubrics, and the integration of automated corrective feedback tools. Collectively, these findings reinforce that DST's benefits derive from repeatable, well-structured practice architectures rather than from media novelty. Iterative drafting under authentic audience conditions cultivates automatized discourse routines for speaking; storyboard pipelines externalise planning for writing; and collaborative scripts distribute cognitive load, enabling learners to focus on language form and meaning (Ajabshir, 2024; Radin & Adam, 2024).

Addressing research question 3, the synthesis highlights several persistent research gaps that constrain cross-study comparison and cumulative insight. Listening and reading outcomes remain underexplored; delayed post-tests beyond one month are rare; and inconsistent reporting of treatment duration, assessment anchors, and fidelity continues to limit synthesis. Additionally, equity concerns, such as access to high-tech tools, require further attention, as evidence suggests that well-scripted, low-tech implementations can achieve comparable outcomes (Rahimi & Yadollahi, 2017; Liang & Hwang, 2023). These observations inform the research agenda outlined in the Future Research Directions section, where priorities include methodological transparency, longitudinal testing, and equitable design replication.

The converging picture is that DST's benefits derive from repeatable practice architectures rather than from media novelty. Iterative drafting under audience expectations cultivates automatized discourse routines for speaking; storyboard-to-draft pipelines externalize cohesion planning for writing; and collaboration scripts distribute cognitive load so that learners can attend to form and meaning. At the same time, the field now has clearer boundary conditions. "One-off" DST assignments with vague criteria or minimal revision rarely move the needle; short projects with more editing than language use risk overvaluing polish; and misaligned assessments (e.g., grading aesthetics over linguistic outcomes) can mask or even suppress language gains. These constraints explain some mixed results and should guide implementation choices going forward. Recent syntheses beyond ELT similarly argue that DST strengthens (meta)cognitive, motivational, emotional, and behavioural regulation; mechanisms that plausibly mediate the durable effects observed in higher-rigour ELT studies.

While much of the evidence is promising, the heterogeneity of study designs, measures, and reporting practices constrains cross-study synthesis. More standardised reporting and shared methodological frameworks are required to strengthen the field and to enable robust comparison across contexts (Snyder, 2019; Paré et al., 2015).

An emergent consideration is the intersection between DST and digital citizenship. As students engage in creating and sharing stories, they develop competencies in online safety, intellectual property, and respectful discourse, making DST a vehicle for media literacy (Adnan et al., 2024). This dimension has implications for curriculum integration, especially in secondary and tertiary education, where digital responsibility is a core competency. In addition, findings suggest that the collaborative structure of DST projects may influence classroom culture by promoting peer accountability and inclusive participation. Students reported higher motivation and a sense of ownership when their contributions were valued in team settings (Yu & Wang, 2025).

A further area of interest is the integration of AI tools for DST production. Grammar checkers, text-to-speech engines, and AI-assisted image generation are beginning to reshape how learners engage with digital storytelling tasks. Preliminary research indicates that these tools can scaffold weaker writers and accelerate revision cycles (Radin & Adam, 2024). However, their pedagogical value depends on critical integration rather than substitution. Teachers need training to harness these affordances while preserving learner creativity and autonomy.

IMPLICATIONS

For curriculum design, DST should be scheduled as a multi-week composing cycle that moves from topic pitch and story grammar to storyboarding, narration drafting, peer review, media assembly, rehearsal, and a public sharing event. Each stage should carry a language target, for example, cohesion devices and audience design in storyboarding; lexis, clause combining, and prosody in narration; and pragmatic moves and pronunciation during rehearsal. Where courses aim at high-stakes outcomes (e.g., IELTS speaking), descriptors can be built into rubrics and self-checklists so that DST rehearsals map transparently to assessment. For writing, specifying CALF computation in rubrics makes expectations visible and enables students to revise strategically. Automated corrective feedback can be woven into revision passes to accelerate accuracy gains without replacing teacher feedback; care should be taken to frame ACF as advisory to protect learner voice (Radin & Adam, 2024).

Given the interpretive and non-replicable nature of narrative reviews, the implications outlined here should be understood as tentative and context-sensitive rather than generalisable claims. The recommendations emerging from this synthesis are intended to guide reflective practice and inform pedagogical design, not to prescribe universal interventions. Readers are therefore encouraged to adapt these insights with consideration of their own institutional, technological, and learner-specific contexts.

At the program and policy level, DST sits naturally within digital-competence agendas: it integrates media ethics (attribution, citation), collaboration, and reflective practice into one artefact. Teacher-education programs can require low-tech fallbacks for every story plan so that graduates can implement DST in constrained settings, an equity move supported by evidence that scripts and feedback, not expensive hardware, drive most gains (Liang & Hwang, 2023; Xu et al., 2024). Institutions seeking to document impact can adopt portfolio assessment, archiving storyboards, scripts, drafts, and reflections; this not only clarifies growth for accreditation but also gives learners durable exemplars. In online and hybrid courses, studies show that DST can lift speaking and vocabulary with acceptable cognitive load, making it a strategic anchor for blended ELT (Tatlı et al., 2022).

LIMITATIONS

This review is subject to several limitations. As a narrative review, it cannot provide pooled effect sizes or statistical generalisation. Publication bias may mean that positive results are overrepresented, while variations in contexts and reporting practices make it difficult to compare across studies. The period between 2015 and 2016 yielded relatively few rigorous DST-ELT studies, leaving the evidence base clustered in later years. In addition, most research focused on speaking, writing, and vocabulary, with listening and reading still underrepresented despite promising results in cases like Rahimi and Yadollahi (2017).

FUTURE RESEARCH DIRECTIONS

Future trials should prioritise durability and transfer, with 3–6-month delayed post-tests and cross-task evaluations (e.g., new speaking prompts, unseen writing genres). The RCT evidence for maintained speaking gains at one month is encouraging; extending that logic to longer intervals and varied tasks would clarify whether DST cultivates reusable routines or task-bound performance (Bai & Xian, 2024). Researchers should also report time-on-task and fidelity to enable fair comparisons with active controls, and they should pre-register analysis plans where feasible.

Receptive-skill outcomes are under-measured. A clear path forward is to pair DST with listening-for-production modules (e.g., mining model stories for discourse moves and prosody) and reading-to-write pipelines that feed storyboarding; both can be evaluated with validated CEFR-aligned instruments. The 2017 online/offline comparison for literacy suggests that delivery mode can matter; updated replications with modern platforms would be valuable (Rahimi & Yadollahi, 2017).

Mechanism-focused studies should integrate self-regulated learning measures (planning, monitoring, reflection) and test mediators such as social presence and authorship. Designs that compare scripted vs. unscripted collaboration already show sizable writing/creativity differences (Chen & Yeh, 2025); analogous manipulations in speaking (e.g., role-based rehearsal scripts) could clarify how structure translates to oral outcomes. As AI-assisted tools become common in drafting, researchers should evaluate pedagogical guardrails that preserve learner voice while leveraging accuracy support (Radin & Adam, 2024).

Finally, equity-oriented studies should test low-resource implementations head-to-head against enhanced-tech versions to estimate cost-effectiveness. Multi-site collaborations across regions would reveal how sociocultural factors (audience norms, media access) mediate effects. Teacher-education research should track classroom transfer longitudinally, from campus projects through practicum into early career teaching, using artefact-based portfolios and classroom observations (Xu et al., 2024; Yu & Wang, 2025).

CONCLUSION

The decade of research from 2015 to 2025 confirms that DST is a valuable pedagogical approach for ELT, especially for enhancing speaking, writing, and vocabulary while reducing anxiety and supporting self-regulation. Its effectiveness, however, depends on thoughtful design, with scripted collaboration, scaffolded planning, and social presence proving critical. Beyond learner outcomes, DST also strengthens teacher education by building digital competence and collaborative capacity. To fully realise its potential, future work must extend the skill domains under study, apply more consistent methodological standards, and examine long-term impacts. DST is thus best understood not as a mere technological tool but as a pedagogical framework that, when well designed, can transform ELT practice.

In summary, the last decade shows a field moving from promise to practice. DST's strongest results appear when educators design for iteration, clarity, and audience; iteration through rehearsal and revision, clarity through explicit language targets and measurement (CALF; speaking subskills), and audience through social presence that raises stakes without raising anxiety. The practical message for ELT is straightforward: a modest toolkit, rigorously structured,

is enough to unlock the bulk of DST's effects. The research message is equally clear: with stronger reporting, longer follow-ups, and attention to listening/reading, DST can transition from a compelling classroom craft to a well-specified pedagogy with predictable returns.

ACKNOWLEDGEMENT

This research was made possible through the collaborative effort of lecturers from Universiti Teknologi MARA, Malaysia and University of Gadjah Mada, Indonesia**.**

REFERENCES

Adnan, W. N. W. M, Husin, N., Hashim, N., & Setia, R. (2024). Digital infographics as a dynamic information-transfer model for academic discourse in language classroom. *Akademika, 94*(2), 438–453. https://doi.org/10.17576/akad-2024-9402-25

Ajabshir, Z. F. (2024). Empowering EFL writing through digital storytelling: A quasi-experimental assessment of CALF and engagement. *Acta Psychologica, 250*, 104564. https://doi.org/10.1016/j.actpsy.2024.104564

Bai, Y., & Xian, H. (2024). Exploring the interplay of digital storytelling, L2 speaking skills, self-regulation, and anxiety in an IELTS preparation course. *Humanities & Social Sciences Communications, 11*, 121. https://doi.org/10.1057/s41599-024-04109-8

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology, 1*(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311

Chen, C.-H., & Yeh, H.-C. (2025). Scripted synergy: Elevating EFL writing and creativity through collaborative digital storytelling. *Technology, Pedagogy and Education*. https://doi.org/10.1080/1475939X.2024.2382970

Eissa, A. (2019). Pedagogic effectiveness of digital storytelling in improving speaking skills of Saudi EFL learners. *Arab World English Journal, 10*(3), 409–421.

Ferrari, R. (2015). Writing narrative style literature reviews. *Nurse Education Today, 35*(5), 1–5. https://doi.org/10.1016/j.nedt.2015.02.006

Hung, H.-T. (2019). Design-based digital storytelling for EFL multiliteracies: A process-tracing study. *Education and Information Technologies, 24*(4), 4017–4035. (ERIC record)

Liang, J.-C., & Hwang, G.-J. (2023). A robot-based digital storytelling approach to enhancing EFL learners' multimodal storytelling ability and narrative engagement. *Computers & Education, 197*, 104827. https://doi.org/10.1016/j.compedu.2023.104827

Nair, V., & Yunus, M. M. (2021). Digital storytelling to improve speaking skills in EFL/ESL: A systematic review. *Sustainability, 13*(17), 9829. https://doi.org/10.3390/su13179829

Nami, N., & Asadnia, F. (2024). Collaborative learner-generated digital storytelling for vocabulary knowledge in EFL. *System, 120*, 103811. https://doi.org/10.1016/j.system.2023.103811

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management, 52*(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

Radin, N. N. M., & Adam, A. F. M. (2024). Evaluating automated grammar corrective feedback tools: A comparative study of Grammarly and QuillBot in ESL expository essays. *Akademika, 94*(2), 474–491. https://doi.org/10.17576/akad-2024-9402-27

Rahimi, M., & Yadollahi, S. (2017). Effects of online vs. offline digital storytelling on English literacy development. *Cogent Education, 4*(1), 1285531. https://doi.org/10.1080/2331186X.2017.1285531

Roza, A., et al. (2023). Digital storytelling to foster academic public speaking for TEFL pre-service students. *Education Research International, 2023*, 1–11. https://pmc.ncbi.nlm.nih.gov/articles/PMC9936128/

Snyder, H. (2019). Literature reviews as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Tatlı, Z. E., Saylan, S., & Kokoç, M. (2022). Online digital storytelling in an EFL course: Effects on speaking, vocabulary and cognitive load. *Participatory Educational Research, 9*(6), 41–60. https://doi.org/10.17275/per.22.123.9.6

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Information Management, 59*(1), 93–103. https://doi.org/10.1016/j.im.2018.10.003

Xu, Z., Mamat, R., & Zin, Z. M. (2024). Exploring the interconnections: Digital competence, teacher training and remote teaching. *Akademika, 94*(3), 108–125. https://doi.org/10.17576/akad-2024-9403-07

Yu, B., & Wang, W. (2025). Using digital storytelling to promote language learning, digital skills and digital collaboration among English pre-service teachers. *System, 129*, 103577. https://doi.org/10.1016/j.system.2025.103577

BIODATA OF AUTHORS:

Rofiza Aboo Bakar
Akademi Pengajian Bahasa,
Universiti Teknologi MARA Cawangan Pulau Pinang,
Malaysia
Email: rofiza@uitm.edu.my

Noor Azli Affendy Lee (Corresponding Author)
Akademi Pengajian Bahasa,
Universiti Teknologi MARA Cawangan Pulau Pinang,
13500 Permatang Pauh, Pulau Pinang, Malaysia
Email: noor.azli@uitm.edu.my

Reni Kusumaningputri
Doctoral Program of Humanities,
Department of English,
Faculty of Cultural Science,
University of Gadjah Mada, Indonesia
Email: renikusumaningputri@mail.ugm.ac.id

Latisha Asmaak Shafie
Universiti Teknologi MARA Cawangan Perlis, 02600
Arau, Perlis, Malaysia
Email: ciklatisha@uitm.edu.my

APPENDIX A

Table 1: Summary of Reviewed Studies (2015–2025)

Year	Authors	Country/ Level	N	Design	Skill(s)	DST Modality	Comparison	Duration	Main finding
2017	Rahimi & Yadollahi	Iran; lower-intermediate EFL (secondary/ tertiary)	42	Quasi-experimental (online vs offline DST)	Reading & Writing (literacy)	Online platform vs offline content-producing program	Online DST vs offline DST (process-writing for both)	≈5 months	Online DST > offline for literacy
2019	Hung	Taiwan; university EFL	NR	Observational/ process-tracing	Multiliteracies; engagement; cognitive/ metacognitive	Learner-created multimodal stories	None (process focus)	Course project	High engagement & strategy use
2019	Eissa	Saudi Arabia; adult EFL	NR	Quasi-experimental	Speaking	DST project work	DST vs traditional instruction	NR	Improved speaking
2021	Nair & Yunus (systematic review)	—	45	Systematic review (PRISMA)	Speaking	Various	—	—	Positive effects on speaking & affect
2022	Tatlı, Saylan & Kokoç	Turkey; university EFL (online)	33	Mixed-methods (pre/post + surveys)	Speaking, Vocabulary, Cognitive load	Online DST tasks in EFL course	Pre–post within-group	Semester	↑ speaking & vocab; manageable load
2023	Liang & Hwang	Taiwan; 11th grade EFL	NR	Experimental comparison	Multimodal storytelling; engagement	Robot-assisted vs animation-DST	Robot-DST vs animation-DST	NR	Robot-DST > animation-DST
2023	Roza et al.	Kazakhstan; TEFL pre-service	60	Mixed-methods; quasi-experimental	Academic public speaking	Learner DS + public presentations	DST vs traditional	7 weeks	↑ fluency, coherence, cohesion; + attitudes
2024	Ajabshir	Iran; university EFL	NR	Mixed-methods (quasi-experimental)	Writing (CALF) & engagement	Digital vs traditional storytelling	DST vs TST	NR	DST > TST on CALF; ↑ engagement
2024	Bai & Xian	China; IELTS prep EFL	89	RCT; pre/post & 1-month follow-up	Speaking, self-regulation, anxiety	DST-integrated IELTS speaking tasks	DST vs control	NR	↑ speaking & self-regulation; ↓ anxiety; sustained
2024	Nami & Asadnia	Iran; university EFL	NR	Sequential explanatory mixed-methods	Vocabulary knowledge	Collaborative learner-generated DS	DST vs control	NR	Higher vocabulary gains
2025	Chen & Yeh	Taiwan; university EFL	NR	Quasi-experimental	Writing, creativity, vocabulary	Scripted collaborative DST	Scripted-DST vs standard DST	NR	Scripted > standard DST

Year	Authors	Country/ Level	N	Design	Skill(s)	DST Modality	Comparison	Duration	Main finding
2025	Yu & Wang	China; pre-service English teachers	NR	Intervention (design-based/empirical)	Language, digital skills, collaboration	Teacher-education DST project	Pre–post; qualitative	NR	↑ digital competence, collaboration, language

Note: N = sample size; DST = Digital Storytelling; TST = Traditional Storytelling; CALF = Complexity, Accuracy, Lexical Diversity, Fluency