



Using an Electronic Student Portfolio as a Learning Tool for Enhancing the English Vocabulary Skills for Undergraduate Students

Jantira Suwanmanee^{1*} · Nunthikan Songsangrak² · Nuchanat Chaimut³
Palachai Khaonuan⁴

¹⁻⁴English for International Communication Program, Faculty of Liberal Arts,
Rajamangala University of Technology Srivijaya, Songkhla, THAILAND

⁵General Education Program, Faculty of Liberal Arts, Rajamangala University of
Technology Srivijaya, Songkhla, THAILAND

*Corresponding Author E-mail: nick.jantira41@gmail.com

(Article history) Received: 2025-04-22, Revised: 2026-04-23, Accepted: 2026-05-16

Abstract

Electronic portfolios can integrate multimedia resources, learner-controlled access, assessment, and evidence of progress, but the strength of an e-portfolio intervention depends on both the quality of its instructional design and the quality of its content. Evidence is especially limited for small, carefully validated vocabulary-learning tools developed for non-English-major engineering students. This study developed and evaluated an electronic student portfolio for English vocabulary learning, with vocabulary drawn from songs, films, and social media, and examined expert validation, instructional efficiency, pre-post vocabulary change, and learner attitudes. The completed study used a one-group pretest-posttest design with seven engineering students purposively selected from RUTS Test performance bands (two high, three medium, and two low). The Google Sites/Google Drive e-portfolio was reviewed independently by three media/design experts and three content experts. A 30-item vocabulary test and three 10-item embedded lesson tests were used. Reported analyses included descriptive statistics, Wilcoxon signed-rank testing, E1/E2 efficiency, an E-post-E-pre-effectiveness index, and attitude ratings in PSPP. Secondary reanalysis was limited to quantities derivable from the published participant-level scores. Design quality was rated $M = 4.53$ ($SD = 0.26$) and content quality $M = 4.83$ ($SD = 0.29$). Pre-test scores ($M = 19.71$, $SD = 2.69$) increased to $M = 26.43$ ($SD = 2.70$); the report gave Wilcoxon Asymp. Sig. (two-tailed) = .017. Efficiency was $E1/E2 = 80.0/88.1$ and $E\text{-post-E-pre} = 22.39$ percentage points. Overall learner attitude was $M = 4.66$ ($SD = 0.44$). A secondary calculation from the seven published score pairs yielded a mean gain of 6.71 points, $SD = 3.20$, 95% CI [3.76, 9.67], and paired Cohen's $d_z = 2.10$. The findings support the feasibility of a validated multimedia e-portfolio as a vocabulary-learning tool in this small engineering-student sample. Because the study had only seven purposively selected participants and no control group or delayed post-test, the results should be interpreted as pilot evidence rather than causal proof of superiority.

Keywords: Electronic Portfolio; E-portfolio; Vocabulary learning; Engineering students; Technology-enhanced language learning; Instructional design; Expert validation

1. INTRODUCTION

Vocabulary knowledge is a central resource for comprehension and communicative performance in a second or foreign language. Pedagogical research has repeatedly shown that vocabulary development benefits from deliberate attention, repeated encounters, meaningful contextualization, and opportunities to connect form and meaning (Nation, 2013; Schmitt, 2008). At the same time, a substantial portion of learners' English exposure now occurs through digitally mediated and audiovisual environments. Songs and television or film can create repeated, contextualized encounters with lexical items, and empirical studies demonstrate that vocabulary can be acquired incidentally from such sources, although learning depends on exposure frequency, prior knowledge, and characteristics of the input (Pavia et al., 2019; Peters & Webb, 2018; Webb & Rodgers, 2009). These observations provide a rationale for instructional designs that organize multimedia vocabulary resources rather than treating entertainment media as unstructured enrichment.

Electronic portfolios are well suited to this purpose because they can assemble texts, images, audio, video, learner tasks, and evidence of progress in one digital space. Earlier e-portfolio research has emphasized reflection, learner engagement, self-regulation, and the capacity to organize evidence of learning (Barrett, 2007; Beckers et al., 2016; Lin, 2008). In higher education, e-portfolios have also been associated with knowledge-management performance and opportunities for authentic or sustained assessment (Chang et al., 2013). However, reviews have cautioned that theoretical enthusiasm for e-portfolios has often exceeded the amount and quality of empirical evidence, particularly when studies do not report the design and validation of the digital instrument in sufficient detail (Bryant & Chittum, 2013). This concern is especially relevant for language-learning interventions: a gain in scores cannot be interpreted well unless readers also know whether the learning content, navigation, media, instructions, and assessment alignment were judged suitable.

The completed research underlying this article addressed that problem through a multimedia electronic student portfolio designed for engineering students at Rajamangala University of Technology Srivijaya (RUTS), Songkhla. The e-portfolio organized English vocabulary from three familiar sources—songs, films, and social media—and was implemented as a web-based learning environment using Google Sites with files and links managed through Google Drive. Rather than relying solely on a learner outcome test, the project incorporated two independent expert-validation tracks: one for media/design quality and one for vocabulary-content quality. The intervention then followed a pretest–learning–posttest sequence with embedded lesson tests and a post-use attitude questionnaire.

This article is deliberately positioned as a pilot development-and-evaluation study rather than a definitive efficacy trial. The sample contained only seven students and there was no comparison group. Accordingly, the principal scientific contribution is not a claim that the e-portfolio is superior to alternative teaching methods. Instead,

the paper integrates four forms of evidence already present in the completed report: (a) design-quality validation, (b) content-quality validation and test congruence, (c) process and pre-post learning indicators, and (d) learner attitudes. This emphasis blends the outcome-oriented original study with the second publication proposal, which foregrounded the development and validation of the multimedia instrument.

Research Gap and Novelty

Three gaps motivate the focused journal reconstruction. First, e-portfolio studies often report learner outcomes without a transparent audit trail showing how the instructional tool was reviewed and revised before use. Second, vocabulary-learning research frequently separates multimedia input research from e-portfolio design; the present project combines a portfolio architecture with vocabulary drawn from songs, films, and social media. Third, non-English-major engineering students are less prominent in the e-portfolio language-learning literature than language majors, teacher-education students, or writing-focused cohorts. The study therefore offers a small but unusually well-documented case of a dual-validated multimedia vocabulary tool used with a technically oriented student population.

Research Objectives and Research Questions

The article addresses four questions using only the completed study and secondary calculations that are directly reproducible from its published tables:

- RQ1. What levels of design quality and content quality did expert panels assign to the developed multimedia e-portfolio, overall and by reported design/content domains?*
- RQ2. Did students' vocabulary scores increase from pre-test to post-test, and what magnitude of within-participant change can be derived from the published score pairs?*
- RQ3. Did the e-portfolio meet the study's 80/80 instructional-efficiency criterion, and what did the lesson-level process scores indicate?*
- RQ4. How did students evaluate the e-portfolio after use, and which aspects received the strongest and weakest ratings?*

2. CRITICAL LITERATURE REVIEW

2.1 Electronic Portfolios as Learning and Assessment Environments

An e-portfolio is more than a digital storage container when it is intentionally designed to support learning. Barrett (2007) distinguished portfolio uses that merely document achievement from uses that support learning through reflection, learner ownership, and engagement. Lin (2008), in a year-long study of preservice teachers constructing e-portfolios online, similarly showed that the process of building a portfolio can become a learning experience in its own right. Lopez-Fernandez and Rodriguez-Illera (2009) examined university students' adaptation to a digital learner course portfolio and highlighted that successful use requires attention to learner

adaptation and the design of the digital environment, not simply the availability of technology.

Evidence from higher education is nevertheless mixed in strength and purpose. Chang et al. (2013) reported that university students learning with e-portfolios outperformed a non-portfolio group on several dimensions of knowledge management, demonstrating that e-portfolios can structure the acquisition, accumulation, application, innovation, and sharing of knowledge. By contrast, Bryant and Chittum's (2013) review concluded that the empirical case for e-portfolio effectiveness was less robust than the popularity of the technology suggested. This tension is important: an e-portfolio should be evaluated as an instructional system whose quality, use, and outcome measures are all documented.

Systematic work has also linked e-portfolios to self-directed learning when the design balances autonomy with adequate support. Beckers et al. (2016) identified multiple factors influencing self-directed learning through e-portfolios, suggesting that design choices, guidance, feedback, and learner agency interact. Poole et al. (2018), examining e-portfolio integration in education, likewise emphasized implementation challenges and supports. Together, this literature suggests that a strong e-portfolio study should report not only outcome scores but also design validation, usability, instructions, and the fit between assessment and learning objectives.

2.2 E-portfolios in Second Language Learning

Second-language e-portfolio research has often concentrated on writing and assessment. Ngui et al. (2020), for example, used an ADDIE-oriented development process to design an e-portfolio for second-language learners in higher education and reported generally positive student responses to the tool. Pourdana and Tavassoli (2022) found that electronic portfolio assessment could support engagement and writing development among university EFL learners, while also documenting the complexity of how feedback and task design affect different components of performance. More recent work continues to report generally favorable learner perceptions of e-portfolios while also highlighting implementation conditions and the need for context-sensitive design (Rezadoust Siah Khaleh Sar et al., 2024).

The present study differs from much of this literature in two respects. It treats the e-portfolio primarily as a vocabulary-learning tool rather than as a writing dossier or speaking-assessment record, and it builds the vocabulary around popular-media domains that are familiar to students. This shifts the portfolio from a retrospective collection of completed work toward a curated multimedia learning space that combines exposure, practice, progress checks, and post-use evaluation.

2.3 Vocabulary Learning through Multimedia and Popular Culture

Vocabulary learning requires repeated and meaningful encounters with lexical items. Nation (2013) argues that effective vocabulary development combines deliberate learning with rich input and opportunities for use. Schmitt (2008) likewise emphasizes that no single method is sufficient; vocabulary growth requires repeated attention across contexts. The task-induced involvement framework of Laufer and Hulstijn

(2001) further predicts deeper lexical processing when tasks create need, search, and evaluation demands.

Popular audiovisual media can provide such contexts. Webb and Rodgers (2009) quantified the lexical demands of television, showing why lexical coverage is central to comprehension. Peters and Webb (2018) demonstrated incidental vocabulary acquisition through viewing L2 television and identified frequency, prior vocabulary knowledge, and cognateness as influential factors. Pavia et al. (2019) showed that listening to songs can also produce vocabulary learning and that repeated listening and exposure frequency improve gains. Captioning research similarly indicates that audiovisual support can facilitate language processing and vocabulary-related learning (Winke et al., 2010). These findings do not prove that any specific multimedia e-portfolio will be effective; rather, they justify the design principle of embedding vocabulary within contextualized, repeated, multimodal materials.

2.4 Need for Development and Validation Evidence

An instructional tool can produce attractive screens and still be pedagogically weak. For that reason, the most distinctive feature of the present report is its separation of media/design validation from content validation. Three design specialists evaluated presentation, visual/interface design, learning media, instructions, and the student manual; three content specialists evaluated vocabulary appropriateness, alignment, and interest. The pre/post assessment was additionally reviewed for item-objective congruence. This layered validation strategy provides evidence about whether the digital environment is usable, whether the vocabulary content is suitable, and whether the test items align with intended objectives.

The current article therefore treats validation results as substantive evidence rather than as a procedural footnote. It also re-aggregates the appendix item means by domain as a secondary analysis. These calculations introduce no new observations; they simply make the validation profile more interpretable for instructional designers.

3. MATERIALS AND METHODS

3.1 Study Design

The completed study used a one-group pretest–posttest design, represented in the report as $E \rightarrow T1 \rightarrow X \rightarrow T2$, where E denotes the study group, T1 the pre-test, X the e-portfolio vocabulary-learning intervention, and T2 the post-test. Although one sentence in the methodology chapter labels the study 'qualitative research,' the implemented design, instruments, statistical procedures, and results are quantitative and pre-experimental. For this manuscript, the design is therefore described according to the procedures actually reported: a quantitative, pre-experimental one-group pretest–posttest pilot study.

3.2 Setting and Participants

The study was conducted at Rajamangala University of Technology Srivijaya, Songkhla, in Academic Year 2019 (B.E. 2562). Seven engineering students were purposively selected using RUTS Test performance information to represent three

proficiency bands: two students from a high-scoring band, three from a middle-scoring band, and two from a low-scoring band. The report does not publish the participant-level mapping between the seven IDs and these proficiency bands; therefore, no subgroup outcome analysis by RUTS proficiency is attempted.

A terminology inconsistency remains unresolved. The English title and English abstract identify the participants as undergraduate engineering students, while several Thai sections use wording that translates as graduate-level students. The underlying report does not provide independent enrollment records that allow this discrepancy to be resolved. Because the confirmed manuscript title specifies undergraduate students, that wording is retained in the title, while the Methods and Discussion use the neutral term 'engineering students' and explicitly disclose the inconsistency.

Table 1: Participant Sampling Structure Reported in the Full Study

Sampling characteristic	Reported	Use in this manuscript
High RMUTSV Test band	2	Describes purposive sampling only; participant IDs not linked to bands
Middle RMUTSV Test band	3	Describes purposive sampling only; no subgroup inference
Low RMUTSV Test band	2	Describes purposive sampling only; no subgroup inference
Total	7	All seven participants contributed pre-test, post-test, embedded-test, and attitude data

Remark: The study used purposive rather than random sampling. No class-distribution concept applies because this was not a classification or machine-learning study.

3.3 E-portfolio Platform and Learning Content

The electronic student portfolio was constructed as a website using Google Sites. The report documents the use of Google Drive for uploading, storing, and sharing files and links, with shared links incorporated into the e-portfolio website. The learning environment used text, images, sound, and video as core media. Vocabulary was organized into three thematic lessons based on songs, films, and social media. The retained process-assessment data show one 10-item embedded test for each lesson, giving 30 possible process points across the three lessons.

The report does not specify computing hardware, browser versions, operating-system versions, Google Sites release numbers, or network specifications. Those details are therefore not invented. Reproduction of the original platform can be conceptual rather than software-version identical: a web-based portfolio capable of displaying multimedia content, linking to shared resources, and presenting vocabulary practice and assessment.

3.4 Instrument Development and Expert Validation

Development proceeded through document/content analysis, selection of vocabulary judged interesting and appropriate for the target group, multimedia design, expert review, and revision. The e-portfolio was reviewed by separate panels of three experts for media/design quality and three experts for content quality. The report's

appendix contains item-level ratings, permitting a more detailed validation profile than the overall means alone.

The media/design panel evaluated five domains: presentation, visual/interface design, learning media, instructions, and the student manual. Inspection of the appendix establishes 20 design-quality items, not 15 as stated in the subsequent proposal document. The content panel evaluated three domains—appropriateness, alignment, and interest—across 10 items, not nine. The appendix counts are used because they are directly auditable from the full report.

The pre/post vocabulary test contained 30 multiple-choice items, corresponding to three 10-item parts. The methods narrative states that the test achieved IOC = 1.00, but the appendix item-level table shows that item 12 received two +1 ratings and one 0 rating, producing IOC = 0.67; the remaining 29 items have IOC = 1.00. All 30 items therefore exceeded the report's stated acceptance threshold of 0.50, but the manuscript does not repeat the unsupported claim that every item achieved 1.00.

3.5 Data Collection Procedure

Data collection followed four steps. First, participants completed the pre-test. Second, they learned English vocabulary through the developed e-portfolio and completed a 10-item embedded test after each of the three lessons. Third, participants completed the post-test. Fourth, they responded to a 10-item attitude questionnaire on a five-point Likert scale. The report states that the e-portfolio, tests, and evaluation forms were accessible through shared links after expert review and revision.

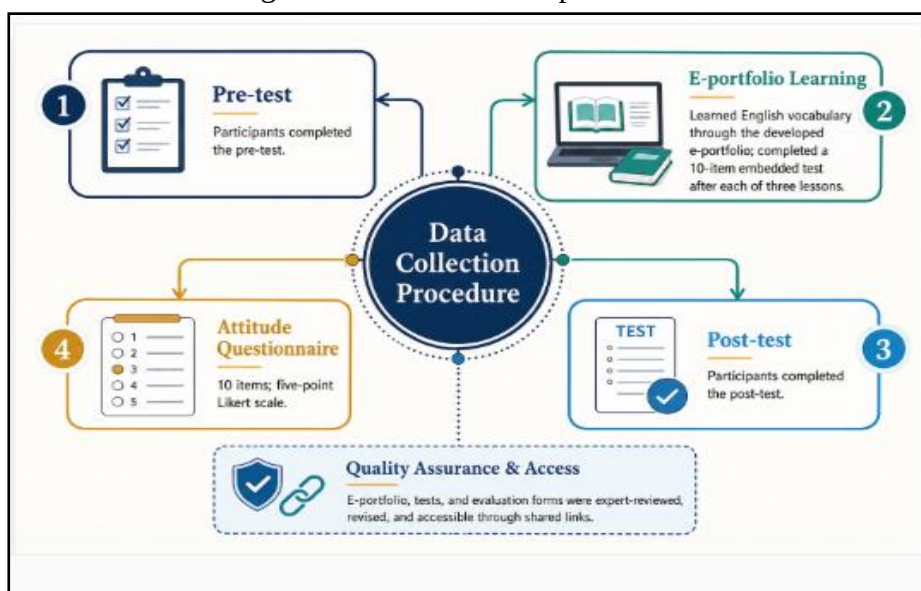


Figure 1: Study and Instrument Development Workflow

3.6 Outcome Measures and Original Statistical Analysis

The primary learning outcome was the 30-point pre/post vocabulary score. The report analyzed the paired scores using the Wilcoxon signed-rank procedure in PSPP and reported the asymptotic two-tailed significance value. Instructional efficiency was quantified using E1/E2. E1 was the percentage of available points obtained across the

three embedded lesson tests; E2 was the percentage of available points obtained on the post-test. The prespecified benchmark was 80/80. The report also calculated an effectiveness index as $E\text{-post} - E\text{-pre}$, where E-pre and E-post were pre-test and post-test percentages. Learner attitudes and expert ratings were summarized with means, standard deviations, and percentage distributions.

3.7 Secondary Analyses Performance

Secondary analyses were restricted to transparent calculations from values already published in the full report. No new participant, test administration, treatment, or outcome was introduced. From the seven published pre/post score pairs, the manuscript recalculated the mean gain and its sample standard deviation, a 95% confidence interval for the mean gain, paired Cohen's d_z (mean of paired differences divided by the standard deviation of paired differences), and an exact two-sided Wilcoxon signed-rank p value as a consistency check on the report's asymptotic PSPP result. Because $N = 7$, these statistics are descriptive/sensitivity supplements rather than grounds for strong generalization.

For instrument validation, item means printed in the appendix were arithmetically re-aggregated into the five design domains and three content domains. These domain means are explicitly labeled secondary reaggregations because the original report emphasized the overall panel means rather than presenting all domain averages in one table.

3.8 Computational Terminology, Comparison Methods, Hardware, and Reproducibility

The user-requested reporting template includes data preprocessing, model architecture, class distributions, cross-validation, ablation, calibration, unseen-dataset evaluation, and hardware. None of these concepts formed part of the completed educational study. There was no machine-learning or statistical-prediction model, no training/validation/test partition, no class label, no feature preprocessing, no hyperparameter optimization, no cross-validation, no ablation study, and no unseen external dataset. The only comparison was within-participant pre-test versus post-test, supplemented by the study's E1/E2 criterion and expert/attitude ratings. Hardware specifications were not reported. PSPP was the stated analysis software, but the version was not given.

Reproducibility is strongest for the reported numerical analyses because participant-level pre/post scores and lesson-level process scores are printed in the report. The original Google Sites/Drive implementation details are partly documented through screenshots, but no archived code repository, permanent data repository, software version, or persistent public URL is supplied. Accordingly, this manuscript supports numerical reanalysis from published outputs and conceptual replication of the instructional design, not byte-for-byte reconstruction of the original website.

Table 2: Source-data Verification and Resolution Decisions

Issue	Evidence in full report	Decision for this manuscript
Student level	English title/abstract: undergraduate; several Thai sections: graduate-level	Unresolved; confirmed title retained, body uses 'engineering students' and discloses conflict
Research-design label	One sentence says qualitative; design and all main analyses are quantitative one-group pre/post	Describe as quantitative pre-experimental pilot
Design-panel item count	Appendix contains 20 items; proposal document summarized 15	Use 20 appendix items
Content-panel item count	Appendix contains 10 items; proposal document summarized 9	Use 10 appendix items
Test IOC	Narrative says 1.00; appendix: item 12 = 0.67 and 29 items = 1.00	Report item-level appendix result; all exceed 0.50 threshold
Post-test mean	One table prints 26.42; raw total 185/7 and narrative give 26.43	Use 26.43
Gain-score SD	Table prints 0.01, inconsistent with published gains 3,5,10,5,12,5,7	Do not use 0.01; secondary recalculation SD = 3.20
Total embedded-test mean	Table prints 18.95 despite total 168/7 = 24.00	Do not use 18.95; report lesson means and E1 = 80.0

Note: Where arithmetic from participant-level or appendix data can resolve a discrepancy, the auditable value is used. Unresolvable inconsistencies are disclosed rather than silently corrected.

4. RESULTS

4.1 Expert Validation of the E-portfolio

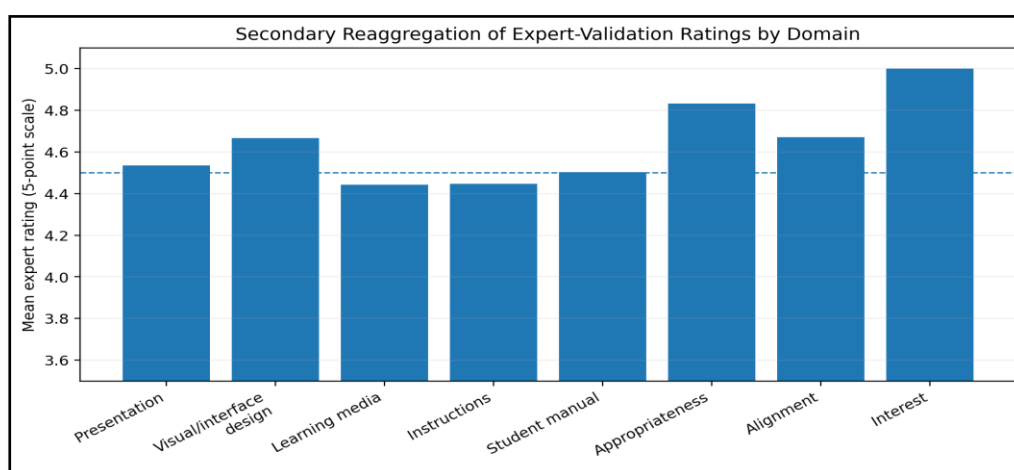
The report rated the overall media/design quality of the e-portfolio as very good ($M = 4.53$, $SD = 0.26$) based on three design/media experts. Across the 20 appendix items, several features reached the ceiling rating of 5.00, including access speed, correspondence between illustrations/video and vocabulary, clarity of illustrations, video image clarity, video sound clarity, support for self-directed learning, and the sequencing guidance in the student manual. The lowest design item concerned whether assessment was sufficiently comprehensive and aligned with objectives ($M = 3.67$, $SD = 0.58$), although this still fell within the report's 'good' band.

Content quality was rated even more strongly ($M = 4.83$, $SD = 0.29$) by three content experts. Vocabulary-objective appropriateness, vocabulary accuracy, category accuracy, vocabulary interest, vocabulary variety, and the ability of activities to support rapid learning were all rated at 5.00. The lowest content item concerned the ordering of vocabulary within lessons ($M = 4.33$, $SD = 1.15$). The instrument therefore showed a very strong content profile while revealing specific design and sequencing elements that could be targeted in later refinement.

Table 3: Expert-validation Results

Validation domain	Mean rating	Status
Design: Presentation	4.53	Secondary reaggregation of appendix item means
Design: Visual/interface design	4.67	Secondary reaggregation of appendix item means
Design: Learning media	4.44	Secondary reaggregation of appendix item means
Design: Instructions	4.45	Secondary reaggregation of appendix item means
Design: Student manual	4.50	Secondary reaggregation of appendix item means
Content: Appropriateness	4.83	Secondary reaggregation of appendix item means
Content: Alignment	4.67	Secondary reaggregation of appendix item means
Content: Interest	5.00	Secondary reaggregation of appendix item means
Overall design quality	4.53 (SD 0.26)	Reported
Overall content quality	4.83 (SD 0.29)	Reported

Remark: Domain means are secondary arithmetic reaggregations of reported item means. Overall panel means and SDs are reported findings.

**Figure 2: Secondary Reaggregation of Expert-validation Ratings by Domain.**

4.2 Test Content Congruence

All 30 pre/post vocabulary-test items met the report's stated IOC acceptability criterion (> 0.50). Twenty-nine items had IOC = 1.00. Item 12 had IOC = 0.67 because one of the three experts assigned a neutral (0) rating while the other two assigned +1. This appendix-based result supersedes the report's more general narrative statement that the test achieved IOC = 1.00 throughout. No coefficient of internal consistency for the 30-item test is reported, so no reliability coefficient is supplied or imputed.

4.3 Vocabulary Achievement

Every participant obtained a higher post-test score than pre-test score. The seven reported score changes were +3, +5, +10, +5, +12, +5, and +7 points. The pre-test mean was 19.71 (SD = 2.69) out of 30 and the post-test mean was 26.43 (SD = 2.70). The report's Wilcoxon signed-rank analysis gave Asymp. Sig. (two-tailed) = .017 and interpreted the pre/post difference as statistically significant at the .05 level.

Table 4: Participant-level Pre-test and Post-test Vocabulary Scores

Participant	Pre-test	Post-test	Gain
1	22	25	3
2	20	25	5
3	17	27	10
4	22	27	5
5	17	29	12
6	17	22	5
7	23	30	7
Mean	19.71	26.43	6.71

Note: Scores are reported in the full research report. The maximum score was 30. The report's printed gain SD of 0.01 is inconsistent with these seven gain values and is not used.

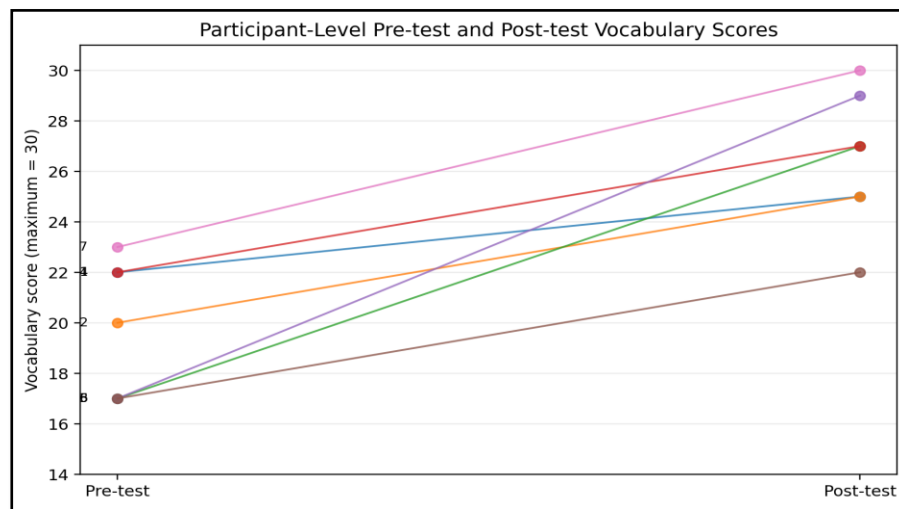


Figure 3: Participant-level Pre-test and Post-test Vocabulary Scores

4.4 Secondary Magnitude and Robustness Analysis of the Pre–Post Change

Using only the seven published score pairs, the mean gain was 6.71 points with a recalculated gain SD of 3.20. The corresponding 95% confidence interval for the mean gain was [3.76, 9.67] points. Paired Cohen's d_z was 2.10, indicating a large within-participant standardized change. Because the design lacks a control group, this effect size quantifies the magnitude of observed change within this sample; it does not identify the e-portfolio as the sole causal source of that change.

As a sensitivity check, an exact two-sided Wilcoxon signed-rank calculation from the published pairs produced $p = 0.01562$. This is close to the report's PSPP asymptotic value of .017 and leads to the same inferential conclusion at $\alpha = .05$. The

exact calculation is reported as a secondary consistency check rather than as a replacement for the original analysis.

Table 5: Reported and Secondary Pre–Post Statistics

Metric			Value	Status / interpretation
Pre-test			M = 19.71, SD = 2.69	Reported
Post-test			M = 26.43, SD = 2.70	Reported; 26.43 matches raw total 185/7
Wilcoxon (2-tailed)	Asymp.	Sig.	.017	Reported PSPP result
Mean gain			6.71 points	Reported mean; confirmed from score pairs
Gain SD			3.20	Secondary recalculation; replaces internally inconsistent printed 0.01 for analysis
95% CI for mean gain			[3.76, 9.67]	Secondary calculation
Paired Cohen's dz			2.10	Secondary standardized magnitude
Exact Wilcoxon p			0.01562	Secondary robustness check

Note: Secondary statistics use only the seven participant-level pairs printed in the original report.

4.5 Instructional-process Performance, Efficiency, and Effectiveness

The three embedded 10-item lesson tests yielded total scores of 57, 58, and 53 across the seven students, corresponding to lesson means of 8.14 (SD = 1.07), 8.29 (SD = 0.95), and 7.57 (SD = 1.51). The third lesson therefore had the lowest mean and the greatest dispersion, suggesting that its content or task demands may have been relatively more difficult or variable for this small group. This is a descriptive interpretation; no lesson-by-lesson significance test was reported.

The total embedded-test score was 168 out of a possible 210, giving $E1 = 80.0\%$. The post-test total was 185 out of 210, giving $E2 = 88.1\%$. Thus, the reported efficiency was $E1/E2 = 80.0/88.1$, meeting the prespecified 80/80 benchmark exactly on $E1$ and exceeding it on $E2$. Pre-test performance expressed as a percentage was $Epre = 65.71\%$, while post-test performance was $Epost = 88.1\%$; the report's effectiveness measure $Epost - Epre$ was therefore 22.39 percentage points.

Table 6: Lesson-level Process Performance and Efficiency Indicators

Indicator	Lesson 1	Lesson 2	Lesson 3	Overall / post
Embedded-test total	57	58	53	168
Mean (out of 10)	8.14	8.29	7.57	—
SD	1.07	0.95	1.51	—
E1 process efficiency	—	—	—	80.0%
E2 post-test efficiency	—	—	—	88.1%
E1/E2	—	—	—	80.0/88.1
Epre	—	—	—	65.71%
Epost – Epre	—	—	—	22.39 percentage points

4.6 Learner Attitudes

Overall post-use attitude toward the electronic portfolio was very positive ($M = 4.66$, $SD = 0.44$), corresponding to the report's highest qualitative band. Three items reached a mean of 5.00: vocabulary appropriateness for learners, the ability of exercises

to check understanding, and ease of access. The lowest-rated item concerned the quality and suitability of images and videos ($M = 4.14$, $SD = 0.69$), which remained within the report's 'good' range but provides a concrete target for future redesign. Vocabulary continuity across lessons had $M = 4.43$ ($SD = 0.79$), and lesson/exercise sequencing had $M = 4.57$ ($SD = 0.79$), indicating more disagreement than the ceiling-rated items.

Table 7: Learner Attitude Ratings After Use of the E-portfolio

Item	Mean	SD
1. Can enhance vocabulary knowledge	4.57	0.53
2. Vocabulary is interesting and engaging	4.71	0.49
3. Learners can sequence the lessons	4.57	0.53
4. Vocabulary knowledge can be applied in varied ways	4.57	0.53
5. Vocabulary is appropriate for learners	5.00	0.00
6. Vocabulary is continuous across lessons	4.43	0.79
7. Lesson and exercise sequence is appropriate	4.57	0.79
8. Exercises can check learner understanding	5.00	0.00
9. Lessons are easy to access and uncomplicated	5.00	0.00
10. Image/video quality is appropriate, clear, interesting	4.14	0.69
Overall	4.66	0.44

Note: Reported results from the 10-item, five-point post-use attitude questionnaire.

4.7 Requested Analyses That Were Not Part of the Completed Study

No ablation findings, k-fold cross-validation results, calibration analysis, unseen-dataset evaluation, external validation, or model-comparison benchmarks exist in the source report. These are not omitted accidentally; they are inapplicable to the study's pre-experimental educational design. Creating such results would require new data or new experimental procedures and would violate the evidence boundary of this manuscript.

5. DISCUSSION

5.1 Interpreting the Combined Development and Outcome Evidence

The principal strength of the study is the convergence of multiple evidence streams. The e-portfolio was not evaluated solely by whether learners liked it or whether post-test scores were higher. Instead, separate expert panels evaluated its media/design quality and its content quality before the student intervention, the test was reviewed for item-objective congruence, students completed process assessments during the three lessons, vocabulary performance was measured before and after learning, and learner attitudes were recorded after use. This sequence makes the case for feasibility more credible than a single post-use satisfaction measure would.

The validation pattern is also informative. Content specialists rated the vocabulary package more highly overall ($M = 4.83$) than design specialists rated the media/interface package ($M = 4.53$). That difference should not be treated as a statistical panel comparison because each panel contained only three experts and the original study did not test the panels inferentially. Descriptively, however, the pattern suggests that the lexical content was a particular strength, whereas assessment alignment, some aspects of presentation, and learner support could still be refined. This

interpretation is consistent with the development-centered logic of Ngui et al. (2020), who emphasized the importance of systematically designing and evaluating an e-portfolio rather than treating the platform itself as the intervention.

5.2 Why Vocabulary Performance May Have Improved

The observed pre–post improvement is compatible with several mechanisms, although the design cannot isolate them causally. First, the e-portfolio offered repeated exposure: learners encountered vocabulary in multimedia materials and then completed embedded tests. Repetition is central to vocabulary learning, and studies of songs and audiovisual input show that frequency of exposure is associated with stronger lexical gains (Pavia et al., 2019; Peters & Webb, 2018). Second, the content was contextualized within songs, films, and social media rather than presented only as decontextualized word lists. Multimedia contexts may support form–meaning associations by providing semantic, visual, and auditory cues.

Third, the e-portfolio supported self-paced access. The report explicitly describes the e-portfolio as an online resource that learners could enter and navigate, and the design panel rated support for self-directed learning at the maximum score. This is compatible with the broader literature in which e-portfolios can promote learner ownership and self-directed learning when sufficient structure and guidance are present (Barrett, 2007; Beckers et al., 2016). Fourth, the embedded exercises created retrieval opportunities. Vocabulary learning is strengthened when learners must retrieve and evaluate lexical knowledge rather than merely re-view content, a principle that aligns with involvement-based accounts of lexical learning (Laufer & Hulstijn, 2001).

5.3 Interpretation of the Magnitude of Change

The secondary standardized effect size ($dz \approx 2.10$) indicates that the within-student score change was large relative to the variability of individual gains. However, the number should be interpreted with substantial caution. With only seven participants, standardized effects are unstable; and in a one-group pretest–posttest design, the effect incorporates not only the intervention but also possible testing effects, maturation, outside exposure, and regression-related influences. The appropriate conclusion is therefore that a large within-sample change was observed after the e-portfolio learning period—not that the e-portfolio caused a two-standard-deviation improvement compared with conventional teaching.

The exact Wilcoxon sensitivity calculation ($p = .01563$) is useful because it confirms that the report's asymptotic $p = .017$ was not an artifact of the PSPP asymptotic approximation. Yet statistical significance with $N = 7$ should not be confused with generalizability. The unanimous direction of change (all seven post-test scores exceeded pre-test scores) is descriptively notable, but replication with a larger controlled sample is necessary.

5.4 Lesson-level Variation

The third embedded lesson had the lowest mean (7.57/10) and the largest SD (1.51), compared with 8.14 (SD = 1.07) and 8.29 (SD = 0.95) for the first two lessons.

The report identifies the three thematic areas as songs, films, and social media; the third lesson corresponds to social-media vocabulary. Because the source report does not provide item difficulty indices, exposure counts, timing data, or a controlled comparison of lesson content, the cause of this pattern cannot be established. It may reflect differences in word difficulty, media quality, task construction, or simple sampling variation. The result nevertheless demonstrates how process data can identify a specific lesson for future qualitative and item-level review.

5.5 Learner Attitudes and Design Implications

Learners' overall attitude score ($M = 4.66$) supports the acceptability of the e-portfolio in this small cohort. Importantly, attitudes were not uniformly at ceiling. The strongest items concerned vocabulary appropriateness, exercise-based understanding checks, and ease of access, while image/video quality was rated lowest. This pattern is useful for design because it separates perceived pedagogical value from media-production quality. An instructional tool can be easy to access and useful for learning while still benefiting from higher-resolution, more consistently edited, or more carefully selected audiovisual resources.

The learner pattern partially converges with the expert validation. Experts rated video image and audio clarity at ceiling in the version they reviewed, whereas learners gave the broader image/video quality item 4.14. This divergence should not be interpreted as a contradiction: experts and learners answered different items from different perspectives, and the sample sizes were extremely small. Rather, it suggests that expert technical judgments and end-user experience are complementary evidence sources. Contemporary e-portfolio research similarly emphasizes that implementation quality includes both instructional design and the user experience of learners (Poole et al., 2018; Rezadoust Siah Khaleh Sar et al., 2024).

5.6 Comparison with Previous E-portfolio and Multimedia Vocabulary Research

The results are broadly consistent with studies that report positive learning or engagement outcomes from e-portfolios, but the present design is too small to claim equivalence with controlled studies. Chang et al. (2013) used a substantially larger sample and a non-portfolio comparison group, whereas the current project lacks a control condition. Pourdana and Tavassoli (2022) likewise analyzed a larger university EFL sample and connected portfolio use to writing development and engagement across repeated tasks. The current study contributes a different, narrower case: vocabulary learning in an engineering cohort, with dual expert validation and entertainment-related content.

The popular-media focus also aligns with audiovisual vocabulary research. Pavia et al. (2019) demonstrated vocabulary learning through songs and the benefit of repeated listening; Peters and Webb (2018) demonstrated learning from television viewing; Webb and Rodgers (2009) described the lexical demands that learners face in television programs; and Winke et al. (2010) showed how captioning can affect foreign-language video learning. The present report did not manipulate media type, captions, repetition counts, or exposure frequency, so it cannot test these mechanisms. Its contribution is instead to show a practical way of curating multiple media sources

inside one portfolio environment and evaluating that environment through content, design, process, achievement, and attitude evidence.

6. CONTRIBUTIONS

6.1 Theoretical Contribution

The study connects two bodies of theory that are often treated separately: e-portfolio learning environments and multimedia vocabulary acquisition. The e-portfolio provides a structure for learner-controlled access, organization, repeated practice, and evidence of progress; popular-media content provides contextualized lexical input. By integrating these functions, the study illustrates how portfolio technology can serve not only as an assessment archive but also as an active lexical-learning environment.

6.2 Methodological Contribution

The most transferable methodological contribution is the dual-track validation model. Media/design quality and content quality were evaluated by separate expert panels, and test-objective congruence was evaluated independently. This separation makes it possible to identify whether weaknesses lie in the learning content, the interface/media, the learner guidance, or the assessment system. The journal reconstruction further demonstrates how legacy educational reports can be audited against participant-level and appendix evidence before publication, preventing internal inconsistencies from being silently propagated.

Computational and Digital-design Contribution

The study is not computational in the sense of machine learning or predictive modeling. Its computational contribution is limited to the deployment of a web-based multimedia e-portfolio using Google Sites and Google Drive and the reproducible calculation of learning/validation metrics from retained scores. The platform architecture demonstrates a low-code approach to integrating multimedia vocabulary resources, shared files, embedded assessments, and learner access. No claim is made about algorithmic intelligence, adaptive sequencing, or automated analytics because none was implemented or evaluated.

Practical Contribution

For English instructors working with non-English-major students, the study provides a concrete design template: select contextually relevant vocabulary, organize it into thematic multimedia lessons, validate content and interface separately, include short embedded retrieval tests, measure pre/post learning, and collect end-user feedback. The expert and learner ratings also identify practical priorities for future versions: maintain easy access and clear guidance, strengthen assessment-objective alignment, review vocabulary sequencing, and improve the perceived quality of images and video.

6.3 Limitations and Future Research

The most important limitation is sample size. With $N = 7$, estimates of means, standard deviations, effect sizes, and attitudes are imprecise and susceptible to the influence of individual observations. The purposive sampling strategy further restricts generalizability. The pre-experimental one-group design lacks a control or comparison group, so the observed improvement cannot be attributed uniquely to the e-portfolio.

The study also lacks a delayed post-test, preventing any conclusion about long-term retention. The published report does not provide participant-level links to the high/middle/low RMUTSV Test sampling bands, so proficiency-moderation analysis cannot be reconstructed. No item-level pre/post response matrix is available in the report, which prevents formal item difficulty, discrimination, distractor, or internal-consistency analysis. Likewise, time-on-task, navigation traces, media exposure counts, and learning analytics were not recorded or reported.

Several reporting inconsistencies required explicit resolution. These included undergraduate versus graduate-level terminology, a qualitative label applied to a quantitative pre/post design, differing item counts in the proposal summary, the narrative IOC = 1.00 claim despite one appendix item at 0.67, a post-test mean printed once as 26.42, an implausible gain SD of 0.01, and an incorrect total process mean. These issues highlight the importance of preserving raw or participant-level outputs and auditing tables before journal submission.

Future work should retain the validated development process but evaluate a larger sample with an appropriate comparison condition and preferably random or matched assignment where feasible. A delayed post-test would address retention. Item-level data would support reliability and diagnostic analyses. Process analytics could test whether frequency of media exposure, lesson order, or time-on-task predicts vocabulary gain. The e-portfolio could also be migrated to a contemporary learning-management or mobile-responsive environment while preserving the core design principles. Critically, these are future-study recommendations; no such experiments are claimed in the present paper.

7. CONCLUSION

This study developed a multimedia electronic student portfolio for English vocabulary learning and evaluated it through independent expert review, pre-post testing, process efficiency, and learner attitudes. The instrument received very high content-quality ratings and high-to-very-high design-quality ratings, with specific improvement targets identified in assessment alignment and media experience. In the seven-student pilot, all participants scored higher after the intervention; the reported Wilcoxon analysis was significant, E1/E2 met the 80/80 criterion, and post-use attitudes were strongly positive.

The most defensible interpretation is that the e-portfolio was feasible, well validated for its intended instructional use, and associated with substantial within-sample vocabulary gains. The evidence does not establish superiority over conventional instruction because there was no control group, the sample was very small, and retention was not measured. By integrating dual expert validation with learner

outcomes and by transparently reconciling source-report inconsistencies, the manuscript offers a reproducible framework for developing and evaluating small-scale technology-enhanced vocabulary interventions in higher education.

REFERENCES

- Barrett, H. C. (2007). Researching electronic portfolios and learner engagement: The REFLECT initiative. *Journal of Adolescent & Adult Literacy*, 50(6), 436–449. <https://doi.org/10.1598/JAAL.50.6.2>
- Beckers, J., Dolmans, D., & Van Merriënboer, J. J. G. (2016). E-portfolios enhancing students' self-directed learning: A systematic review of influencing factors. *Australasian Journal of Educational Technology*, 32(2), 32–46. <https://doi.org/10.14742/ajet.2528>
- Bryant, L. H., & Chittum, J. R. (2013). ePortfolio effectiveness: A(n ill-fated) search for empirical support. *International Journal of ePortfolio*, 3(2), 189–198.
- Chang, C.-C., Tseng, K.-H., Liang, C., & Chen, T.-Y. (2013). Using e-portfolios to facilitate university students' knowledge management performance: E-portfolio vs. non-portfolio. *Computers & Education*, 69, 216–224. <https://doi.org/10.1016/j.compedu.2013.07.017>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Lin, Q. (2008). Preservice teachers' learning experiences of constructing e-portfolios online. *The Internet and Higher Education*, 11(3–4), 194–200. <https://doi.org/10.1016/j.iheduc.2008.07.002>
- López-Fernández, O., & Rodríguez-Illera, J. L. (2009). Investigating university students' adaptation to a digital learner course portfolio. *Computers & Education*, 52(3), 608–616. <https://doi.org/10.1016/j.compedu.2008.11.003>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
- Ngui, W., Pang, V., & Hiew, W. (2020). Designing and developing an e-portfolio for second language learners in higher education. *International Journal of Information and Education Technology*, 10(5), 362–366. <https://doi.org/10.18178/ijiet.2020.10.5.1390>
- Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition*, 41(4), 745–768. <https://doi.org/10.1017/S0272263119000020>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577. <https://doi.org/10.1017/S0272263117000407>

- Poole, P., Brown, M., McNamara, G., O'Hara, J., O'Brien, S., & Burns, D. (2018). Challenges and supports towards the integration of ePortfolios in education: Lessons to be learned from Ireland. *Heliyon*, 4(11), e00899. <https://doi.org/10.1016/j.heliyon.2018.e00899>
- Pourdana, N., & Tavassoli, K. (2022). Differential impacts of e-portfolio assessment on language learners' engagement modes and genre-based writing improvement. *Language Testing in Asia*, 12, Article 7. <https://doi.org/10.1186/s40468-022-00156-7>
- Rezadoust Siah Khaleh Sar, H., Khalili Sabet, M., Mahdavi Zafarghandi, A., et al. (2024). Iranian teachers' and learners' attitudes towards the use of e-portfolio in second language speaking assessment. *Language Testing in Asia*, 14, Article 61. <https://doi.org/10.1186/s40468-024-00335-8>
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366. <https://doi.org/10.1111/j.1467-9922.2009.00509.x>
- Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407–427. <https://doi.org/10.1093/applin/amp010>
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65–86. <https://doi.org/10.64152/10125/44203>