Project Albus Taiwan 2025

Academic Report

Sheng-Li Junior High School Pilot on AI-Integrated Teaching and Learning

Prepared by

Institute of Education

National Yang Ming Chiao Tung University

Research Lead: Dr. Jerry Chih-Yuan Sun

Academic Research Team: National Yang Ming Chiao Tung University Research Team

In collaboration with Dalberg Advisors, Google for Education,

Hsinchu County Government Education Bureau, MediaTek, and Shou Yang Digital Technology

March 2026

Team Introduction

The academic research team for the NYCU component of Project Albus Taiwan 2025 comprised four core members who jointly supported research design, bilingual and AI-integrated pedagogy, field coordination, data collection, and academic report preparation.

Dr. Jerry Chih-Yuan Sun **Research Lead** Distinguished Professor and Director, Institute of Education / Center for Teacher Education, National Yang Ming Chiao Tung University. Led the overall academic direction of the pilot, including research framing, methodology, interpretation of findings, and report oversight.	**Dr. Lu-Chun (Regine) Lin** **Bilingual and Pedagogical Advisor** Associate Professor, Graduate Institute of TESOL, and Director, Center of Higher Education Accreditation for Teaching, National Yang Ming Chiao Tung University. Contributed expertise in language and literacy development, bilingual/EMI teacher education, and AI-enhanced pedagogy.
Dr. Shih-Jou Yu **Educational Technology Advisor** Educational technology researcher with experience in interactive feedback, wearable technology, AR/VR, chatbots, and learning motivation research. Supported technology-enhanced instructional design, implementation planning, and interpretation of classroom innovation practices.	**Chou-Yu (Joe) Chen** **Graduate Researcher** Master's student, Digital Learning Program, Institute of Education, National Yang Ming Chiao Tung University; Google Certified Trainer. Handled field coordination, instrument administration, classroom observations, interview organization, data processing, and primary drafting of the academic report.

Table of Contents

I. Introduction

1. Background and Rationale

As schools accelerate digital transformation, “technology integration” is no longer just about adding devices—it is increasingly about redesigning teaching, learning, and assessment so that digital tools genuinely strengthen learning outcomes and learner agency (Zawacki-Richter et al., 2019). This shift is becoming even more salient in the era of generative AI, where learners can access powerful language-based support for searching, summarizing, drafting, and idea generation in seconds (Kasneci et al., 2023). At the same time, research has warned that without clear guidance, such tools may encourage superficial task completion, weaken metacognitive monitoring, and create a misleading sense of competence, especially when students treat AI output as “the answer” rather than as a starting point for reasoning (Kasneci et al., 2023). For school systems, the practical question is therefore not whether AI will enter classrooms, but how to integrate it in ways that develop students’ digital and AI-related competencies while maintaining authentic thinking, ethical judgment, and responsible use (Zawacki-Richter et al., 2019).

Within this landscape, the Albus Project is positioned as an evidence-building initiative focused on how a coherent ecosystem of digital learning tools, professional development, and classroom implementation can support measurable changes in teacher practice and student competency over time. Critically, sustainable adoption is not simply a “tool issue,” but a learning-design issue: students benefit most when technology use is paired with meaningful tasks, clear roles, timely feedback, and structures that promote autonomy and productive collaboration (Bond et al., 2020). This emphasis aligns closely with the research showing that learners’ engagement in technology-mediated settings is shaped by motivational beliefs and learning processes, and that effective digital learning depends on designs that support self-regulation and meaningful participation rather than passive consumption (Sun & Rueda, 2012). In addition, Sun’s work has highlighted how learning analytics and process data can help educators see patterns of participation and interaction that are often invisible in traditional classrooms, enabling more targeted instructional support (Sun et al., 2018).

From a competency perspective, Chiu et al. (2022) have also demonstrated that "technology learning support" can foster students' digital literacy through motivational mechanisms consistent with self-determination theory—particularly when learning environments strengthen competence, autonomy, and relatedness.

Against this research backdrop, the 2025 Albus Taiwan pilot was designed to support schools in building practical, classroom-ready pathways for AI-integrated teaching and learning, while simultaneously generating rigorous evidence through mixed data sources. Research One focuses on strengthening students' digital literacy, establishing best practices for AI integration in school contexts, and supporting self-directed inquiry and problem-solving; it also examines pre–post differences in students' digital literacy after AI tools are integrated into teaching and learning activities. Research Two further examines how different AI learning strategies relate to students' learning motivation and performance, extending the evaluation beyond tool exposure alone to instructional design choices. In this pilot, participating student cohorts were drawn from grades 7–8 at Hsinchu County Sheng-Li Junior High School, Zhubei City, and were divided into two research samples defined by class membership and age range (approximately 12–14 years old).

Building on these aims, this report synthesizes multiple forms of evidence—including teacher and student pre/post surveys, classroom observations, and teacher-authored lesson plans—to provide a comprehensive yet practice-oriented account of implementation and outcomes. Conceptually, the report treats changes in digital/AI-related competencies as inseparable from changes in learning experience, because engagement and motivation are not "extra outcomes," but core mechanisms through which technology-supported learning produces (or fails to produce) meaningful growth (Bond et al., 2020). This stance is consistent with Sun's research, which underscores that technology-enhanced learning effectiveness depends on how tools are embedded into pedagogy, feedback cycles, and learner participation structures, rather than on the presence of technology itself (Sun et al., 2018; Sun & Rueda, 2012).

The research questions guiding this report are as follows:

1. What changes were observed in teachers' and students' perceptions, readiness, and competencies related to digital and AI-supported teaching and learning during the Albus implementation period?
2. To what extent did the two studies generate changes in learning-related outcomes, including digital literacy and classroom learning processes in Study 1, and reading-related performance and motivational/process outcomes in Study 2?
3. How did instructional design, participation structures, AI/tool use, and teacher facilitation shape the teaching and learning experiences observed across the two studies?

2. Overview of Project Albus Taiwan Pilot

2.1 Pilot Title, Research Team, and Partners

This research is implemented under the formal project title: "Project Albus: Education Digital Transformation Pilot" (阿爾巴斯計畫：教育數位轉型試點計畫). The research is conducted under the Institute of Education, National Yang Ming Chiao Tung University (NYCU).

In line with the Albus collaboration model, the study also plans to share de-identified data with key project partners, including Google for Education (GFE), Dalberg, and MediaTek, with explicit safeguards such as unlinking personal identifiers and ensuring confidentiality standards for any shared dataset.

2.2 Two Complementary Studies Under One Pilot

The Taiwan pilot consists of two studies designed to address complementary dimensions of AI-enabled digital transformation in school contexts:

- Study 1 (Research One): Project Albus Main Study
  Focus: students' digital literacy and self-directed learning capacity through the introduction and guided use of Google teaching and AI-enabled tools in classroom learning activities.

- Study 2 (Research Two): AI-Supported Collaborative Application in English Reading
  Focus: the influence of Google AI learning tools on English learning, implemented as a structured instructional intervention in English classes.

A key design feature is that the participant groups for Studies 1 and 2 are mutually exclusive, explicitly intended to reduce learner burden and prevent cross-condition contamination between the research activities.

3. Pilot Interventions and Learning Tools

3.1 Core Digital and AI Tools Used in the Pilot

In Study 1, students use school-provisioned accounts to access and engage with the following AI-enabled digital learning tools:

1. Google Workspace for Education
2. Gemini
3. NotebookLM

These tools are embedded in teacher-designed learning activities. The design intention is not merely to "increase usage," but to support students' information access, collaborative work, and feedback cycles through teacher-facilitated classroom routines.

3.2 Study 1: Structure of Learning Activities and Evaluation Touchpoints

Study 1 follows a staged flow consistent with Albus-style evaluation logic (pre → implementation → post), combining surveys and learning activities with follow-up qualitative inquiry:

- Pre-stage: student pre-survey (including basic background and digital literacy items).
- Implementation: 2 lessons (45 minutes each) where students participate in learning activities supported by teacher-guided tool use.
- Post-stage: student post-survey (including learning motivation and satisfaction items) and semi-structured interviews.

3.3 Study 2: Instructional Duration and Participation Requirements

For Study 2, participation requires completion of a structured instructional sequence over the two weeks. The program includes two sessions and a pre-test/post-test, with each session lasting 45 minutes. This design supports more sustained instructional exposure and enables evaluation of learning changes over a defined intervention window.

II. Research Methods — Study 1 (Research One): Project Albus Main Study

1. Study Overview and Research Design

Study 1 (Project Albus Main Study) examined the effectiveness of a school-wide digital transformation pilot implemented at Sheng-Li Junior High School (Hsinchu, Taiwan). The study adopted a mixed-methods, pretest–posttest evaluation design to assess changes in students' and teachers' readiness for digital learning and AI-supported learning, as well as perceived impacts on classroom engagement and teaching/learning processes.

Study 1 was the "main study" in the Project Albus pilot and was implemented independently of Study 2, with the two studies using mutually exclusive student samples (i.e., students in Study 2 were not included in Study 1).

Data collection and analysis were organized around three foundational attributes of digital transformation used across Project Albus:

1. Attitude and mindset toward technology/AI in learning and teaching
2. Knowledge (digital literacy, comfort with Google tools, and AI familiarity)
3. Capability and behavior (actual usage patterns, perceived effectiveness, and barriers)

In addition, Study 1 examined perceived changes in outcomes such as student engagement, collaboration, independent learning, creativity, and teacher productivity/time use, consistent with the pilot's transformation goals.

2. Research Procedure

2.1 Implementation timeline (pilot cycle)

The Study 1 implementation followed a phased pilot structure:

July–August 2025: Planning and research design

The research team finalized the evaluation framework, survey instruments, classroom observation and interview protocols, and consent procedures. During this phase, the team also coordinated with school administrators and participating teachers to align the pilot with regular school operations.

September 2025: Teacher baseline (pre-pilot) measurement and pilot preparation

Teacher pre-surveys were administered to establish baseline information on attitudes toward technology and AI, readiness for digital teaching, and existing patterns of Google tool use. During this stage, participating teachers also engaged in preparation activities, including tool familiarization, classroom planning, and implementation alignment.

December 2025: Student baseline measurement and initial classroom implementation

Before classroom activities began, student participants completed baseline measures, including background items and indicators related to digital literacy, tool readiness, and prior experience with technology-supported learning. After the baseline survey, the pilot moved to classroom implementation using Google Workspace for Education and selected AI-supported tools based on lesson needs and teacher readiness.

December 2025 to January 2026: Ongoing pilot implementation and qualitative data collection

During the implementation period, teachers integrated Google tools and selected AI tools into authentic classroom instruction. At the same time, the research team collected qualitative evidence through classroom observations, post-observation teacher interviews, and student interviews in order to better understand how the pilot unfolded in practice.

January 2026: Endline (post-pilot) measurement

At the end of the pilot cycle, teacher and student post-surveys were administered.

Additional observations and interviews were also conducted in the late stage of implementation to capture more stabilized classroom routines, perceived outcomes, and remaining barriers.

2.2 Intervention components (what was implemented)

Study 1 evaluated an integrated set of support aimed at improving digital learning conditions at the school level:

1. Digital infrastructure & platform adoption
   - Use of Google Workspace for Education accounts for teaching/learning workflows (e.g., Classroom, Docs, Drive, Slides, Forms).
   - Use of school devices (e.g., Chromebooks) in participating classrooms.
2. AI tool exposure and classroom use
   - Gemini and NotebookLM were introduced as AI supports for learning preparation, depending on lesson needs and teacher readiness.
   - AI use was framed as teacher-guided and learning-focused (e.g., information search, scaffolding, feedback), rather than replacing student thinking.
3. Teacher professional development and coaching
   - Teachers received structured training and ongoing support across the semester (certification-oriented learning, tool workshops, device training, and coaching).
4. Student and parent enablement
   - Students received AI-related learning sessions aligned with practical classroom use.

Parents were engaged through online communication/training to strengthen home–school understanding of digital learning and AI use.

2.3 Study 1 data collection flow (student-facing procedure)

Study 1 was implemented in three stages to minimize disruption to regular instruction:

- Stage 1: Pre-activity preparation (≈ 30 minutes)
    1. Research team briefing on purpose, procedures, and participant rights
    2. Collection of student and legal guardian consent (only consented participants were included in research datasets)
    3. Student completion of baseline measures (demographics + digital literacy/tool readiness indicators)
    4. Verification that students could access the relevant accounts/tools for learning activities
- Stage 2: Learning activity implementation (≈ 45 minutes, typically 1–2 lessons)
    - Teachers conducted instruction using Google tools and, when appropriate, AI tools (Gemini/NotebookLM).
    - AI use was primarily positioned for:
        - Supporting student information search and inquiry
        - Supporting group activities
        - Enabling more timely feedback and learning support
- Stage 3: Post-activity reflection and endline measurement (≈ 50 minutes)
    1. Student completion of endline measures (e.g., learning motivation and satisfaction indicators; plus broader pilot perception measures in the post-survey)
    2. Semi-structured student interviews (voluntary participation; individual or small group depending on scheduling feasibility)

3. Participants

3.1 Study setting

The study was conducted at Sheng-Li Junior High School, a public junior high school in Hsinchu, Taiwan, during the 2025 academic year pilot cycle.

3.2 Student participants (Study 1 sample)

- Target population: 283 students, Grade 7 and Grade 8 students participating in the school's Albus digital transformation implementation.
- Exclusion principle: Students allocated to Study 2 (the experimental English-learning study) were excluded from Study 1 to maintain mutually exclusive samples.
- Participation basis: Voluntary participation with required student + guardian consent for inclusion in research datasets.

Analytical approach for matching: Pre–post survey comparisons were conducted using matched records.

3.3 Teacher participants

- Target population: 21 teachers who engaged in the pilot's professional development and implemented Google tools (and, where applicable, AI tools) in instruction.
- Coverage: Teachers represented a cross-disciplinary cohort (e.g., languages, ICT, sciences, arts, social studies, and integrated activities), reflecting the pilot's school-wide transformation focus.
- Participation basis: Voluntary participation in surveys and interviews; data were analyzed at an aggregated level to protect privacy.

3.4 Qualitative subsamples (observations and interviews)

To understand "how" and "why" changes occurred (beyond survey scores), Study 1 included qualitative sampling designed for maximum variation (different subjects, grade levels, and instructional formats):

- Classroom observations: Multiple lessons across Grade 7–8, covering varied subjects (e.g., ICT, language arts, English, guidance/career, arts/music, and integrated activities).
- Post-observation teacher interviews: Short semi-structured interviews conducted immediately after observed lessons to clarify instructional intent, tool selection rationales, and perceived student responses.
- Student interviews: Semi-structured interviews (voluntary), designed to include diverse learners and to capture both enabling factors (what worked) and constraints (what made learning difficult).

4. Research Tools and Data Sources

Study 1 integrated four complementary data streams to support triangulation and improve validity.

4.1 Quantitative instruments: Teacher and student surveys (pre/post)

Study 1 used the Google for Education Survey (Core + AI) (Teacher and Student versions), administered at baseline and endline. The instrument set captured:

1. Background variables
    - Role-related or student demographic indicators (e.g., gender; teaching experience; subject area)
2. Attitude and mindset
    - Perceptions of the value of technology/AI in classrooms
    - Post-survey reflection on whether attitudes shifted during the pilot
3. Knowledge/comfort with tools
    - Self-reported comfort levels with Google tools (e.g., Classroom, Docs, Drive, Slides, Forms; and AI tools such as Gemini where applicable)
    - Familiarity measured via ordinal confidence categories (from “not comfortable” to “extremely confident/trained”)

4. Capability and behavior
    - Frequency/time using technology in classrooms
    - Devices used and perceived device suitability for learning/teaching
    - Tool usage patterns during the pilot (post-survey emphasis)
5. Barriers and disruptions
    - Obstacles to using technology and AI-powered technology
    - Common disruptions (e.g., connectivity, hardware constraints, classroom management friction)
6. Perceived classroom impact
    - Student engagement and classroom experience (student self-report; teacher perception of student engagement)
    - Student-centered learning emphasis (teacher report)
    - Perceived changes in learning processes (e.g., collaboration, independent learning, feedback speed, creativity, completion rates)
    - Recommendation intentions (e.g., likelihood to recommend Google solutions/devices)

4.2 Qualitative instruments: Classroom observation protocol

A structured observation template was used to ensure consistency across observers and lessons. It recorded:

- Lesson metadata: date/time, grade/subject, activity topic, lesson rationale and objectives
- Observation domains (field notes):
    1. Perspective & usage of technology (teacher practice)
    2. Student perspective & usage (ease of use, differentiation needs, participation patterns)

3. Impacts on learning and teaching (classroom management, engagement, interaction quality)

- Synthesis section: key findings and interpretation of how technology/AI supported (or constrained) lesson goals
- Actionable follow-ups: questions to ask in the teacher interview; suggestions for next observations

4.3 Semi-structured interviews: Teachers and students

Interviews were designed to contextualize survey patterns and explain mechanisms behind observed changes.

- Teacher interviews focused on:
  - Why specific tools were chosen for a lesson
  - What changed in classroom routines and instructional decision-making
  - Perceived student response patterns (engagement, collaboration, learning pace)
  - Practical constraints (time, infrastructure, classroom management) and support needs
- Student interviews focused on:
  - Which tools helped learning and why
  - How students used AI tools (and how they judged reliability)
  - What increased or reduced engagement
  - Device preferences and reasons (usability, speed, screen size, convenience)
  - Recommendations for improving digital learning conditions (e.g., connectivity, device experience)

4.4 Instructional artifacts and implementation evidence

To document the actual instructional design work produced during the pilot, Study 1 collected:

- Teacher lesson plans using a common template (topic, objectives, tools used, instructional flow, links/materials)
- Classroom implementation evidence

5. Data Processing and Analysis Plan

5.1 Quantitative analysis (survey)

Study 1 followed a structured pilot evaluation framework commonly used in school-based implementation studies:

1. Data cleaning and validation
    - Removal of duplicate entries (if present)
    - Standardization of identifiers for pre/post matching
2. Scale construction
    - Construction of composite indicators aligned to the three transformation attributes:
        - Attitude/mindset index
        - Knowledge/comfort index (e.g., Google tool confidence)
        - Capability/behavior index (usage patterns; perceived effectiveness)
3. Pre–post comparison
    - Matched pre–post comparisons for continuous/ordinal indicators
4. Exploratory subgroup checks (non-causal)
    - Grade-level or subject-area comparisons (when appropriate)

5.2 Qualitative analysis (observations, interviews, artifacts)

Qualitative data were analyzed using structured trustworthiness strategies:

- Triangulation across sources (survey + observation + interview + lesson artifacts)
- Use of "negative case" checks (actively looking for exceptions to dominant patterns)
- Separation of descriptive observation notes from interpretive claims to reduce confirmation bias

5.3 Mixed-method integration

Study 1 followed a convergent mixed-methods logic:

- Quantitative results established what changed (and the size/direction of change)
- Qualitative evidence explained how changes manifested in practice and what conditions enabled them

III. Research Methods — Study 2 (Research Two): AI-Supported Collaborative Application in English Reading

1. Study Overview and Research Design

Study 2 examined how integrating NotebookLM into English reading instruction, together with reciprocal teaching (RT) and either cooperative or individual learning formats, may influence junior high school students' learning processes and outcomes. The study adopted a quasi-experimental, 2 × 2 factorial design using intact classes (i.e., existing class groupings were preserved for feasibility and classroom stability). In line with the planned design, the study emphasized:

1. Learning format (cooperative vs. individual)
2. NotebookLM support (with vs. without NotebookLM)

2. Research Setting

To ensure tool access and consistent learning workflows, students used school-provided Google Workspace for Education accounts, and NotebookLM was included as the designated AI learning tool in the experimental conditions.

3. Participants and Group Assignment

3.1 Participants

Study 2 involved a total of 77 8th-grade students from four intact classes at Sheng-Li Junior High School. All four classes were taught by the same English teacher, which helped control for teacher effects and reduce instructional variation across conditions.

3.2 Grouping Strategy (Intact-Class Assignment)

To avoid disrupting class organization and daily routines, the four intact classes were directly assigned to four experimental conditions. This approach was explicitly intended to preserve classroom stability and support feasible implementation in a real school setting. The four instructional conditions were defined as follows:

- **AI-CR**: AI (NotebookLM) + cooperative reciprocal teaching
- **AI-IR**: AI (NotebookLM) + individual reciprocal teaching
- **CR**: no AI + cooperative reciprocal teaching
- **IR**: no AI + individual reciprocal teaching

4. Instructional Intervention and Learning Activities

4.1 Duration and Structure

The intervention spanned two weeks. The program included two sessions and a pre-test/post-test, with each session lasting 45 minutes.

4.2 Common Pedagogical Core: Reciprocal Teaching (RT)

Across all four conditions, instruction was designed around the reciprocal teaching cycle—predicting, questioning, clarifying, and summarizing—as a set of core comprehension strategies enacted through structured dialogue and guided practice (Palincsar & Brown, 1984; Rosenshine & Meister, 1994). In cooperative conditions, these steps were implemented through role-based small-group interaction and rotating leadership, consistent with established reciprocal teaching implementations that gradually transfer strategic responsibility from teacher to students (Palincsar & Brown, 1984; Spörer et al., 2009). In individual conditions, students completed the same strategy cycle independently through guided worksheets that externalized predictions, questions, clarifications, and summaries, reflecting common classroom enactments that use structured written scaffolds to maintain strategy fidelity and task focus (Spörer et al., 2009).

4.3 NotebookLM Integration (AI-Supported Conditions)

In NotebookLM conditions, the tool was used as a course-bound reading support assistant: instructional materials were uploaded in advance, ensuring that NotebookLM responses were grounded in the teacher-provided texts and supporting students' reading comprehension and discussion quality. The study also required students to preserve NotebookLM interaction traces (e.g., dialogue content and referenced evidence markers) as part of the learning record to strengthen process accountability.

4.4 Differentiated Learning Worksheets by Condition

To ensure instructional equivalence while preserving the integrity of each condition, four worksheet versions were developed. Key differences focused on (a) whether the task product was individual vs. cooperative and (b) whether NotebookLM interaction evidence was required. For example, the NotebookLM worksheet versions added fields such as “prompt / AI response / what I learned.” They required comparing students’ summaries with AI-generated summaries, while cooperative versions emphasized group consensus and shared reflection.

5. Instruments and Measures

Study 2 collected both quantitative and qualitative data to capture learning outcomes and indicators of the learning process.

5.1 Pre-test Measures (Baseline)

The pre-test phase (≈ 30 minutes) included:

- Background information items
- Self-efficacy scale (8 items)
- English reading comprehension test (20 items)

5.2 Post-test Measures (After the Intervention)

The post-test phase (≈ 50 minutes) included:

- Self-efficacy scale (8 items)
- Situational interest scale (24 items)
- Learning engagement scale (19 items)
- Cognitive load scale (10 items)
- English reading comprehension test (20 items)

5.3 Self-Efficacy Scale

The self-efficacy scale used in this study was adapted for the English reading context with reference to the Self-Efficacy for English Scale reported by Sun et al. (2015), which was originally grounded in the self-efficacy for learning and performance framework proposed by Pintrich et al. (1991). The scale consisted of 8 items and was administered at both the pretest and posttest stages. The items assessed students' perceived confidence in understanding the main ideas and important details of English texts, inferring the meanings of unfamiliar words or phrases from context, summarizing paragraph content in their own words, generating key questions to check comprehension, clarifying misunderstandings while reading, predicting upcoming text content, using appropriate learning tools to support comprehension, and believing that they could learn English reading content successfully. All items were rated on a 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree), with higher scores indicating stronger self-efficacy. Reliability of the scale was examined using Cronbach's α (Nunnally & Bernstein, 1994). In the present study, the scale demonstrated excellent internal consistency, with Cronbach's $\alpha = .94$ at pretest.

5.4 Situational Interest Scale

The situational interest scale used in this study was adapted from the revised localized version reported by Lin et al. (2021), which was developed based on the Situational Interest Scale proposed by Chen et al. (1999) and Chen et al. (2001). To fit the context of the present English reading activities, the wording of the items was adjusted according to the learning task theme, and the instrument was reviewed by two experts in digital learning prior to administration. The scale comprised 24 items across six subdimensions: exploration intention, instant enjoyment, novelty, attention demand, challenge, and total interest, with four items in each subdimension. All items were rated on a 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree), with higher scores indicating stronger situational interest during the learning activity. Reliability of the scale was examined using Cronbach's α (Nunnally & Bernstein, 1994). In the present study, the situational interest scale demonstrated excellent internal consistency (Cronbach's $\alpha = .96$). The six subdimensions also showed good to excellent internal consistency: exploration intention ($\alpha = .93$), instant enjoyment ($\alpha = .95$), novelty ($\alpha = .89$), attention demand ($\alpha = .94$), challenge ($\alpha = .89$), and total interest ($\alpha = .95$).

5.5 Learning Engagement Scale

The learning engagement scale used in this study was adapted from the Engagement Scale developed by Sun et al. (2019) and was revised to fit the present English reading activity context. Before administration, the instrument was reviewed by two experts in digital learning to establish expert validity. The scale consisted of 19 items across three dimensions: behavioral engagement (5 items), emotional engagement (6 items), and cognitive engagement (8 items). The behavioral engagement dimension assessed students' effort, attention, and persistence during the activity; the emotional engagement dimension captured students' affective responses, such as enjoyment and boredom; and the cognitive engagement dimension examined the extent to which students used strategies such as self-questioning, rereading, and comprehension monitoring. The scale included both positively keyed and reverse-coded items. All items were rated on a 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree), with higher scores indicating stronger engagement in the learning activity. Reliability of the scale was examined using Cronbach's α (Nunnally & Bernstein, 1994). In the present study, the engagement scale showed excellent internal consistency (Cronbach's α = .94). The behavioral, emotional, and cognitive engagement subdimensions also showed acceptable to excellent internal consistency, with alpha coefficients of .60, .90, and .93, respectively.

5.6 Cognitive Load Scale

The cognitive load scale used in this study was adapted from the multidimensional cognitive load instrument developed by Leppink et al. (2013), with wording adjusted for the present English reading activity and with reference to the revised version reported by Lin et al. (2021). The scale comprised 10 items across three dimensions: intrinsic cognitive load (3 items), extraneous cognitive load (3 items), and germane cognitive load (4 items). Intrinsic cognitive load referred to the perceived complexity of the learning content itself, extraneous cognitive load referred to the burden caused by the way instructions or explanations were presented, and germane cognitive load referred to the extent to which the activity supported meaningful understanding and schema construction. All items were rated on a 6-point Likert scale from 1 (strongly disagree) to 6 (strongly agree), with higher scores indicating stronger perceived cognitive load in the corresponding dimension. Reliability of the scale was examined using Cronbach's α (Nunnally & Bernstein, 1994). In the present study, the cognitive load scale showed good internal consistency (Cronbach's $\alpha = .80$). The intrinsic, extraneous, and germane cognitive load subdimensions also showed good to excellent internal consistency, with alpha coefficients of .91, .81, and .93, respectively.

5.7 English Reading Tests (Prior Knowledge and Learning Outcomes)

The English reading pretest and posttest used in this study consisted of the same 20 researcher-developed multiple-choice items, with only the item order and passage order rearranged to reduce potential memory effects while preserving comparability. Each item was scored dichotomously as incorrect, yielding a total score range of 0 to 20. Prior to formal administration, the test was reviewed and revised by two experts in digital learning and one expert in English teaching to establish expert validity. Test development was based on recent item types used in Taiwan's junior high school English reading assessments, the difficulty level and themes of the English textbook currently used at the research site, and the revised Bloom's taxonomy as the theoretical framework for classifying cognitive process levels (Anderson et al., 2001). The test blueprint covered multiple reading subskills, including main idea, detail comprehension, inference, vocabulary inference in context, referential relations, and text structure. In this study, remembering and understanding items were classified as lower-order processing, whereas analyzing items were classified as higher-order processing.

To further examine test quality, item analysis was conducted using the matched sample responses from the present study. Item difficulty (P) was calculated as the proportion of students answering each item correctly, and item discrimination (D) was estimated using the upper–lower group method in item analysis (Ebel, 1954). Following Kelley's (1939) recommendation, the upper 27% and lower 27% groups were formed based on total test scores, and item discrimination was calculated from the difference in correct response rates between these two groups. For the pretest, item difficulty ranged from .20 to .82, and item discrimination ranged from .29 to .81. Overall, the item set showed an adequate spread of difficulty and positive discrimination across items. Internal consistency of the dichotomously scored test was estimated using KR-20 (Kuder & Richardson, 1937), yielding a value of .81 for the pretest.

5.8 Semi-Structured Interviews (Qualitative Component)

To complement survey/test data and capture students' perceptions of (a) reciprocal teaching processes, (b) cooperative/individual learning experiences, and (c) NotebookLM-supported strategy use, semi-structured interviews were conducted.

The interview framework maintained comparable burden across conditions (each condition used 10 core questions + 5 probes), while selectively adapting questions to match students' actual learning experiences (e.g., role rotation in cooperative groups; NotebookLM use in AI-supported groups).

IV. Findings and Impact — Study 1 (Research One): Project Albus Main Study

1. Overview of Data and Sample

This chapter reports the key findings from Study 1 (Project Albus Main Study) using (1) teacher and student pre-/post-surveys, (2) classroom observations (7 lessons across multiple subjects), (3) teacher follow-up interviews conducted after observations, (4) a student group interview, and (5) collected lesson plans.

Survey responses (matched at the individual level):

- Students: Pre n = 283; Post n = 283 (283 matched pre–post responses from the same students were included in the analysis)
- Teachers: Pre n = 21; Post n = 21 (21 matched pre–post responses from the same teachers were included in the analysis)

2. Key Quantitative Findings (Surveys)

2.1 Changes in Students' Digital Tool Familiarity (Google Apps)

Overall pattern (students):
Students' average comfort across the core Google Apps list showed a statistically significant improvement from pre to post (3.05 → 3.18, $\Delta = 0.13$, $p = .001$, $d_z = 0.19$). Although the overall effect size was small, the result indicates that the pilot moved beyond simple exposure and produced measurable gains in students' operational readiness to learn with digital tools.

As presented in Table 1, the clearest student gains were concentrated in tools closely tied to classroom learning routines, especially Practice Sets ($\Delta M = 0.44$, $d_z = 0.34$), Google Sheets ($\Delta M = 0.28$, $d_z = 0.24$), and Read Along ($\Delta M = 0.27$, $d_z = 0.19$). This pattern supports the argument that the pilot was most effective when digital tools were embedded in concrete academic tasks rather than introduced as isolated technologies.

Students already reported relatively high comfort with Google Search at baseline; its additional increase likely reflects more purposeful academic use during the pilot. Chromebook comfort also improved significantly, suggesting that the implementation strengthened not only app familiarity but also students' readiness to participate in device-supported learning environments.

Table 1. Students' Comfort with Google Apps (Pre vs. Post, 1–5 scale; selected statistically significant changes)
(ΔM = Post − Pre; paired-samples t-test; d_z = Cohen's effect size for paired data. Following Cohen (1988), effect sizes of 0.20, 0.50, and 0.80 may be interpreted as small, medium, and large, respectively.)

Google App (Students)	*n*	Pre *M*	Post *M*	Δ*M*	*p*	d_z
Overall (mean across apps)	283	3.05	3.18	0.13	.001	0.19
Practice Sets (Google Classroom)	283	2.12	2.56	0.44	< .001	0.34
Google Sheets	283	2.31	2.59	0.28	< .001	0.24
Read Along (Google Classroom)	283	2.61	2.88	0.27	.002	0.19
Google Search	283	3.81	3.99	0.18	.016	0.14
Chromebook	283	2.75	2.92	0.17	.017	0.14

2.2 Changes in Teachers' Digital Tool Familiarity (Google Apps + Gemini)

Teachers showed larger pre-to-post gains than students, which is an encouraging finding because teacher readiness is typically the key driver of sustainable school-level implementation.

Overall pattern (teachers):
Teachers' average comfort across apps increased significantly (2.85 → 3.41, Δ = 0.57, $p < .001$, $d_z = 0.96$). This indicates a substantial shift from basic functional use toward more confident instructional use across the Google ecosystem.

Table 2. Teachers' Comfort with Google Apps (Pre vs. Post, 1–5 scale; selected statistically significant changes)

Google App (Teachers)	***n***	Pre ***M***	Post ***M***	ΔM	***p***	d_z
Overall (mean across apps)	21	2.85	3.41	0.57	< .001	0.96
Google Slides	21	2.52	3.57	1.05	< .001	0.98
Gemini	21	2.76	3.67	0.90	< .001	1.18
Google Forms	21	2.90	3.81	0.90	< .001	0.96
Chromebook	21	2.57	3.33	0.76	< .001	1.22
Google Drive	21	3.29	3.90	0.62	.009	0.64
Gmail	21	3.33	3.81	0.48	.009	0.64

Interpretation:
Teacher growth was concentrated in instruction-facing tools and workflow tools, especially Google Slides ($\Delta M = 1.05$, $d_z = 0.98$), Gemini ($\Delta M = 0.90$, $d_z = 1.18$), Google Forms ($\Delta M = 0.90$, $d_z = 0.96$), and Chromebook ($\Delta M = 0.76$, $d_z = 1.22$). This is an important implementation signal because these are the tools teachers rely on to prepare lessons, deliver instruction, manage classroom tasks, and provide feedback.

The large increase in Gemini comfort suggests that the pilot strengthened not only general digital readiness but also AI-related instructional confidence. At the same time, gains in Google Drive and Gmail indicate that the pilot supported broader operational fluency across the Google ecosystem, making it easier for teachers to integrate digital tools into everyday teaching practice.

2.3 Attitudes Toward Technology/AI: Consolidation of Positive Perceptions

Students (post-shift item):

- 78.5% reported a beneficial view (either continued to see benefits, or shifted from obstacles/indifference to a more positive view).
- 16.6% remained indifferent.
- 4.9% moved toward a more negative stance.

Teachers (post-shift item):

- 100% reported either continued positive attitudes or shifted toward recognizing benefits.

Interpretation:
The pilot appears to have stabilized and strengthened positive perceptions of technology and AI—especially among teachers. For students, the "mostly positive but not uniform" pattern is typical in school-wide pilots and suggests a realistic foundation: the program made progress while still leaving room to improve learning relevance and device experience for all learners.

2.4 Barriers and Enablers: Obstacles Generally Decreased

A key implementation indicator is whether reported obstacles shrink as training, routines, and infrastructure mature.

I. Students — Obstacles to using technology / AI in class

Technology obstacles (selected highlights):

- "No obstacles" increased slightly (45.2% → 47.0%)
- "Limited access/infrastructure" decreased (21.6% → 17.3%)
- "Limited understanding of technology" decreased (19.8% → 16.6%)

AI obstacles (normalized categories):

- "No obstacles" increased (54.8% → 56.5%)
- "Lack of awareness of AI learning tools" decreased (21.9% → 15.2%)

- “Limited access to AI resources” decreased (18.0% → 13.4%)
- “AI use not encouraged by teachers” decreased (6.4% → 2.8%)

II. Teachers — Obstacles reduced substantially (especially AI-related access/awareness)

AI obstacles (teachers):

- Limited access/infrastructure: 57.1% → 14.3%
- Limited awareness/understanding: 66.7% → 28.6%
- “No obstacles”: 0% → 14.3%

Interpretation:
The clearest “success signal” here is the sharp reduction in AI access and AI awareness barriers among teachers. This strongly suggests that the pilot’s training + hands-on practice successfully moved teachers from “interested but constrained” to “able to act.”

2.5 Student Learning Experience During the Pilot (Post-only Impact Indicators)

Students responded to a set of statements about learning processes and outcomes. Overall, results were consistently positive—especially on peer learning, pace control, independence, and collaboration.

Top student-reported impacts (% Agree):

- 61.8%: “I help and receive help from peers in class.”
- 61.5%: “I am able to learn at my own pace.”
- 60.1%: “I am able to undergo independent learning.”
- 59.7%: “I am able to collaborate more with my classmates.”
- 55.1%: “I am empowered to take charge of my own learning.”

Interpretation:
These results align closely with Study 1’s intended outcomes around self-directed learning and collaboration. The strongest signals are not merely “liking technology,” but reporting shifts in how learning happens (pace, ownership, peer support).

2.6 Teacher-Reported Student Outcomes (Post-only Impact Indicators)

Teachers rated the impact on student learning with notably high agreement rates:

Top teacher-reported impacts (% Agree):

- 81.0%: Adaptive learning helps adjust to student learning pace
- 81.0%: Students receive real-time assessment/feedback
- 76.2%: Personalized learning increases student learning motivation
- 76.2%: Students' performance improved
- 76.2%: Students grasp content quicker and more effectively

Interpretation:
Teacher perceptions are strongly aligned with the learning-design logic of the pilot: personalization, feedback, and pace adjustment. In implementation research, teacher-reported "instructional affordances" of this magnitude typically predict stronger continuation intentions.

2.7 Recommendation and Time-Saving Indicators

Teachers:

- Recommend Google Solutions (1–10): Mean = 8.52 (high)
- Recommend Chromebooks (1–10): Mean = 7.76 (moderate-high)
- Time saving: 95.2% reported Google Solutions saved time (some or significant)

Students:

- Recommend Google Solutions (1–10): Mean = 6.54 (moderate)
- Recommend Chromebooks (1–10): Mean = 6.34 (moderate)

Interpretation:

Teachers showed stronger endorsement than students. Students' more moderate ratings appeared to reflect usability and implementation conditions rather than rejection of digital learning.

3. Qualitative Findings (Observations, Teacher Interviews, Student Interview, Lesson Plans)

3.1 Observation Coverage: Evidence of Cross-Subject Implementation

Across 7 observed lessons, digital tools were integrated into diverse subjects and activity structures—suggesting a school-wide implementation footprint rather than a single-subject pilot.

Table 3. Overview of Observed Lessons (Study 1 Implementation Evidence)

Subject/Grade	Topic (abridged)	Core Tools/Approach (observed)
Home Economics / 7th	Kitchen safety task	Google Slides + interactive task + Kahoot; bilingual scaffolding
ICT / 8th	Arrays via Scratch	Google Classroom + Docs/Slides + Scratch; task-based group work
English / 8th	Vocabulary practice & assessment	Blended approach; gamified vocabulary practice (e.g., Quizlet-style flow)
Mother Tongue / 8th	Classical Chinese	In-class flipped approach; video-based input then practice
Guidance / 8th	Career group project	Student-centered research + collaborative production (e.g., Canva/Chat)
ICT / 8th	Scratch programming + Kahoot	Seamless workflow: instruction → practice → assessment
Music / 7th	Rhythm & composition	Chrome Music Lab integration with performance practice

3.2 Theme 1 — "Ecosystem Integration" Reduced Friction for Teachers

Across teacher interviews, a repeated theme was that Google Workspace functioned as a single connected workflow (materials, delivery, student work collection, and revisions). Teachers emphasized the practical value of centralization—especially via Drive/Classroom/Slides.

Why this matters for Study 1:
Reduced operational friction is a key prerequisite for teachers to design learning that builds students' digital literacy and self-directed learning habits. The pilot appears to have improved not only "tool availability," but instructional feasibility.

3.3 Theme 2 — Engagement Increased When Tools Supported Interaction, Gamification, or Creation

Both teachers and students described engagement as activity-dependent: devices alone do not guarantee motivation, but interactive structures consistently raised involvement. In student interviews, memorable activities included gamified learning, Canva creation, Scratch game-making, and QR-based exploration. In multiple observations, tools such as Kahoot and creative production tasks triggered visible engagement spikes.

A short student-reported infrastructure quote ("The internet is very bad") also clarifies that engagement is partly constrained by conditions of use.

3.4 Theme 3 — Movement Toward Student-Centered Learning and Differentiation

Teachers described shifting classroom dynamics when tools were used to:

- offload input (video-based or guided content)
- increase practice time
- support peer discussion and group operation
- conduct quick checks (e.g., Kahoot) to guide pacing

This theme aligns with the teacher survey result where all teachers reported increased student engagement, and most reported greater focus on student-centered learning during the pilot.

3.5 Theme 4 — AI Was Most Strongly Adopted for Teacher Preparation (Not Replacing Teaching)

Teacher interviews repeatedly positioned AI (especially Gemini) as a planning accelerator: brainstorming structures, summarizing content, drafting slides, and preparing differentiated materials. This aligns with the large increase in teacher Gemini comfort (quantitative) and the high rate of teacher-reported time saving.

Crucially, the qualitative data indicates AI was used mainly to support instructional design, rather than to substitute teacher judgement—an important point for ethical and sustainable AI integration.

3.6 Theme 5 — Constraints as "Next-Step Targets": Infrastructure, Device Fit, Bilingual Demands, and Academic Integrity

The qualitative data also surfaced improvement targets that help explain "moderate" student recommendation scores:

1. Infrastructure: network speed was the most frequently cited barrier and a major disruption baseline in student pre-survey.
2. Device fit: student feedback suggested mixed perceptions of device suitability rather than a uniform preference. While several students favored tablets such as iPads because of screen size and perceived ease of use, teacher interviews emphasized the classroom advantages of Chromebooks for typing, workflow centralization, and integration with Google Workspace. This pattern suggests that device preference was likely task-dependent and shaped by usability and infrastructure conditions.
3. Bilingual challenge: teachers noted that bilingual prompts sometimes caused hesitation, indicating the need for clearer scaffolding (especially for technical vocabulary).
4. Academic integrity: students openly described copy/paste behaviors enabled by technology, highlighting the importance of integrating process-oriented tasks, classroom norms, and AI/digital ethics as the program scales.

Interpretation:

These are not signs of failure—they are high-quality diagnostic signals that typically emerge once implementation becomes real and frequent enough for users to notice friction. In other words, the pilot progressed to a stage where optimization (not adoption) becomes the main challenge.

4. Chapter Summary (Study 1)

Across quantitative and qualitative data, Study 1 demonstrates a coherent pattern of impact:

1. Capability Gains
    - Students: significant improvement in comfort across Google Apps (especially Classroom-embedded learning tools).
    - Teachers: larger and more instruction-relevant improvements, including strong growth in Slides, Forms, Chromebook, and Gemini.
2. Barrier Reduction
    - Both students and teachers reported fewer obstacles over time, with teachers showing especially sharp reductions in AI access/awareness barriers.
3. Learning Process Improvements
    - Students reported higher autonomy, pace control, peer support, and collaboration (majority agreement across key items).
    - Teachers reported strong benefits in personalization, feedback, and perceived performance improvement.
4. Implementation Evidence
    - Observations and lesson plans confirm cross-subject integration and a shift toward more student-centered routines supported by the Google ecosystem.

V. Findings and Impact — Study 2 (Research Two): AI-Supported Collaborative Application in English Reading

This chapter presents the findings from Study 2. The study drew on pre–post English reading comprehension tests, pre–post self-efficacy measures, post-only measures of situational interest, cognitive load, and learning engagement, as well as post-intervention interview evidence across the four instructional conditions. To mirror the structure of Study 1, the findings are organized into two parts: quantitative and qualitative. The quantitative results present the overall pattern of outcomes across the four conditions, whereas the qualitative findings help contextualize how students experienced reciprocal teaching (RT), cooperation, and NotebookLM-supported learning during the intervention.

1. Matched Sample and Grouping

Study 2 examined the effects of AI support (NotebookLM) and learning mode (cooperative vs. individual reciprocal teaching) on junior high school students' English reading outcomes. The study adopted a four-group quasi-experimental design using intact classes. As shown in Table 4, the four instructional conditions were defined by crossing AI support (AI vs. no AI) with learning mode (cooperative reciprocal teaching vs. individual reciprocal teaching).

Table 4. Experimental Conditions

Group Code	AI Support	Learning Mode	*n*
AI-CR	AI (NotebookLM)	Cooperative reciprocal teaching	17
AI-IR	AI (NotebookLM)	Individual reciprocal teaching	18
CR	No AI	Cooperative reciprocal teaching	19
IR	No AI	Individual reciprocal teaching	23

Note. AI-CR = AI (NotebookLM) + cooperative reciprocal teaching; AI-IR = AI (NotebookLM) + individual reciprocal teaching; CR = no AI + cooperative reciprocal teaching; IR = no AI + individual reciprocal teaching.

2. Quantitative Findings

2.1 Reading Comprehension Outcomes

Reading comprehension outcomes were examined using Wilcoxon signed-rank tests to assess within-group pretest–posttest changes across the four instructional conditions. As shown in Table 5, only AI-IR demonstrated a statistically significant improvement in overall reading comprehension from pretest to posttest (Z = -2.44, p = .015). In contrast, AI-CR did not show a statistically significant pretest–posttest change (Z = -1.05, p = .292). Likewise, neither CR (Z = -0.55, p = .583), nor IR (Z = -0.18, p = .855), showed a statistically significant improvement. These findings indicate that, within the present intervention period, the clearest measurable gain in overall reading comprehension emerged only in the AI-IR condition.

To further clarify the nature of the reading outcomes, the reading comprehension test was divided into low-level and high-level reading comprehension. As presented in Table 6, only AI-IR showed a statistically significant improvement in low-level reading comprehension (Z = -2.06, p = .040). However, the same group did not show a statistically significant improvement in high-level reading comprehension (Z = -1.77, p = .077). For the remaining three groups, neither low-level nor high-level reading comprehension showed statistically significant pretest–posttest changes. Specifically, AI-CR showed no significant change in either low-level reading comprehension (Z = -0.25, p = .805), or high-level reading comprehension (Z = -1.42, p = .155). CR also showed no significant change in low-level reading comprehension (Z = -0.89, p = .372), or high-level reading comprehension (Z = -0.06, p = .949). Similarly, IR did not show significant change in low-level reading comprehension (Z = -0.19, p = .848), or high-level reading comprehension (Z = -0.21, p = .830).

Taken together, these findings suggest that the instructional benefit observed in Study 2 was conditional rather than uniform across conditions. More specifically, the most visible gain appeared in AI-IR, and this gain was primarily concentrated in low-level reading comprehension rather than high-level reading comprehension. This pattern suggests that, within the current instructional design and intervention period, AI may have been more effective in supporting students' basic comprehension processes than in promoting more advanced reading performance.

Table 5. Within-Group Pre–Post Changes in Overall Reading Comprehension

Group Code	*Z*	*p*	Interpretation
AI-CR	-1.05	.292	Not significant
AI-IR	-2.44	.015	Significant improvement
CR	-0.55	.583	Not significant
IR	-0.18	.855	Not significant

Table 6. Within-Group Pretest–Posttest Changes in Low-Level and High-Level Reading Comprehension

Group	Low-Level		High-Level		Interpretation
Code	*Z*	*p*	*Z*	*p*	
AI-CR	-0.25	.805	-1.42	.155	Not significant
AI-IR	-2.06	.040	-1.77	.077	Significant improvement only in low-level reading comprehension
CR	-0.89	.372	-0.06	.949	Not significant
IR	-0.19	.848	-0.21	.830	Not significant

2.2 Self-Efficacy

Self-efficacy was examined using Wilcoxon signed-rank tests to assess within-group pretest–posttest changes across the four instructional conditions. As shown in Table 7, none of the four groups demonstrated a statistically significant change in self-efficacy from pretest to posttest. The four groups showed no statistically significant pretest–posttest differences in self-efficacy: AI-IR ($Z = -1.30$, $p = .194$), AI-CR ($Z = -0.51$, $p = .607$), CR ($Z = -0.44$, $p = .658$), and IR ($Z = -1.10$, $p = .272$). These findings suggest that students' self-efficacy remained relatively stable across the four instructional conditions during the intervention period.

Taken together, the results indicate that the short-term intervention did not substantially alter students' self-perceived confidence in English reading across the four conditions. In other words, although AI-IR showed significant gains in reading comprehension, this improvement was not accompanied by a statistically significant increase in self-efficacy. This pattern suggests that changes in reading performance may emerge earlier than shifts in broader self-beliefs about one's reading ability.

Table 7. Within-Group Pretest–Posttest Changes in Self-Efficacy

Group Code	*n*	*Z*	*p*	Interpretation
AI-CR	17	-0.51	.607	Not significant
AI-IR	18	-1.30	.194	Not significant
CR	19	-0.44	.658	Not significant
IR	23	-1.10	.272	Not significant

2.3 Situational Interest

Situational interest was examined using Kruskal–Wallis tests to compare posttest differences across the four instructional conditions. As shown in Table 8, no statistically significant group differences were found in situational interest, $\chi^2(3) = 2.56$, $p = .465$. Likewise, no statistically significant differences were found across the six subdimensions of situational interest, including exploration intention ($\chi^2(3) = 3.60$, $p = .308$), instant enjoyment ($\chi^2(3) = 1.79$, $p = .616$), novelty ($\chi^2(3) = 4.72$, $p = .194$), attention demand ($\chi^2(3) = 3.40$, $p = .334$), challenge ($\chi^2(3) = 1.95$, $p = .583$), and total interest ($\chi^2(3) = 1.66$, $p = .645$). These findings indicated that students in AI-CR, AI-IR, CR, and IR reported comparable levels of situational interest after the intervention.

Taken together, the results suggest that differences in AI support and reciprocal teaching mode did not produce statistically significant variation in students' situational interest during the present intervention period. In other words, students' situational interest did not differ significantly across the four instructional conditions (AI-IR, AI-CR, CR, and IR). This suggests that, in the present four-group comparison, none of the instructional conditions produced significantly higher or lower situational interest than the others.

Table 8. Differences in Situational Interest Across the Four Instructional Conditions

Variable	χ^2	p	Interpretation
Situational Interest	2.56	.465	Not significant
Exploration Intention	3.60	.308	Not significant
Instant Enjoyment	1.79	.616	Not significant
Novelty	4.72	.194	Not significant
Attention Demand	3.40	.334	Not significant
Challenge	1.95	.583	Not significant
Total Interest	1.66	.645	Not significant

2.4 Cognitive Load

Cognitive load was examined using Kruskal–Wallis tests to compare posttest differences across the four instructional conditions. As shown in Table 9, no statistically significant group differences were found in cognitive load ($\chi^2(3) = 3.29$, $p = .350$). Likewise, no statistically significant differences were found across the three subdimensions of cognitive load: intrinsic cognitive load ($\chi^2(3) = 1.01$, $p = .800$), extraneous cognitive load ($\chi^2(3) = 0.43$, $p = .933$), and germane cognitive load ($\chi^2(3) = 5.20$, $p = .158$). These findings indicate that students in AI-CR, AI-IR, CR, and IR experienced comparable levels of cognitive load after the intervention.

Taken together, the results suggest that differences in AI support and reciprocal teaching mode did not produce statistically significant variation in students' perceived cognitive load during the present intervention period. In other words, the integration of AI did not appear to impose additional learning burden on students, nor did the cooperative or individual reciprocal teaching format lead to significantly different levels of intrinsic, extraneous, or germane cognitive load. This pattern suggests that the four instructional conditions were broadly similar in the cognitive processing they placed on students.

Table 9. Differences in Cognitive Load Across the Four Instructional Conditions

Variable	χ^2	p	Interpretation
Cognitive Load	3.29	.350	Not significant
Intrinsic Cognitive Load	1.01	.800	Not significant
Extraneous Cognitive Load	0.43	.933	Not significant
Germane Cognitive Load	5.20	.158	Not significant

2.5 Learning Engagement

Learning engagement was examined using Kruskal–Wallis tests to compare posttest differences across the four instructional conditions. As shown in Table 10, no statistically significant group differences were found in learning engagement, $\chi^2(3) = 2.41$, $p = .492$. Likewise, no statistically significant differences were found across the three subdimensions of learning engagement, including behavioral engagement, $\chi^2(3) = 4.20$, $p = .241$; emotional engagement, $\chi^2(3) = 2.26$, $p = .520$; and cognitive engagement, $\chi^2(3) = 1.73$, $p = .631$. These findings indicate that students in AI-CR, AI-IR, CR, and IR reported comparable levels of learning engagement after the intervention.

Taken together, the results suggest that differences in AI support and reciprocal teaching mode did not produce statistically significant variation in students' learning engagement during the present intervention period. In other words, no statistically significant differences were found among the four instructional groups (AI-IR, AI-CR, CR, and IR) in learning engagement or in its three subdimensions (behavioral, emotional, and cognitive engagement). Thus, within this four-group comparison, no single instructional condition was associated with significantly higher or lower learning engagement than the others.

Table 10. Differences in Learning Engagement Across the Four Instructional Conditions

Variable	χ^2	p	Interpretation
Learning Engagement	2.41	.492	Not significant
Behavioral Engagement	4.20	.241	Not significant
Emotional Engagement	2.26	.520	Not significant
Cognitive Engagement	1.73	.631	Not significant

2.6 Chapter Summary of Quantitative Findings

Overall, the quantitative findings suggest that the four instructional conditions produced broadly similar patterns in students' situational interest, cognitive load, learning engagement, and self-efficacy. These findings indicate that the different instructional designs did not lead to significantly different levels of motivational response or perceived task burden during the intervention period. In particular, the integration of NotebookLM did not appear to reduce students' situational interest and learning engagement, nor did it impose additional cognitive burden on learners. Likewise, none of the four groups showed a statistically significant pretest–posttest change in self-efficacy, suggesting that the use of innovative AI-supported instruction did not undermine students' confidence in their English reading ability.

In contrast to the largely stable patterns observed in the motivational and process-related measures, reading comprehension outcomes revealed a different pattern. Among the four groups, only AI-IR showed a statistically significant improvement in overall reading comprehension from pretest to posttest. Further analysis indicated that this improvement was primarily reflected in low-level reading comprehension, whereas high-level reading comprehension did not reach statistical significance. The other three groups did not show significant gains in either overall reading comprehension or its low-level and high-level components. This pattern suggests that the instructional value of NotebookLM may be conditional rather than universal. More specifically, under the present instructional design, NotebookLM appeared to be more beneficial when combined with individual reciprocal teaching, particularly for supporting students' lower-level reading comprehension processes.

Taken together, these quantitative findings suggest two main implications. First, the present AI-supported instructional design did not generate negative effects on students' situational interest, learning engagement, self-efficacy, or perceived cognitive burden. Second, although the effects were not uniform across all conditions, the AI-IR condition showed a positive effect on reading comprehension, especially at the low-level comprehension stage. Therefore, the results suggest that the educational value of NotebookLM lies not simply in whether AI is introduced, but in how it is integrated into an appropriate instructional structure. The present study's findings provide preliminary support for using NotebookLM as a feasible instructional aid in English reading, while also indicating that further refinement of instructional design may be needed to strengthen its impact across broader learning outcomes.

3. Qualitative Findings

3.1 Students Brought Emerging Reading Strategies Into the Intervention

The interview data suggest that students did not enter the intervention as strategy-free readers. Instead, they already used a range of practical reading approaches, such as previewing questions, scanning for possible answers, or reading the passage first before returning to the items. From this perspective, the intervention may have functioned less as the introduction of entirely new strategies and more as a process of making existing strategies more explicit, discussable, and transferable through reciprocal teaching.

3.2 Cooperation Functioned as Both Support and Burden

Students' accounts of cooperative learning were mixed. On the positive side, some students described group work as useful for discussing ideas together, generating better questions, and comparing interpretations before deciding on a final answer. These responses suggest that cooperation could support co-construction and improve the quality of discussion.

At the same time, cooperation was not uniformly smooth. Some responses indicated uneven participation within groups, and a few students described situations in which one member carried a disproportionate share of the work. Other students suggested that the most difficult part of collaboration was not understanding the text itself, but managing peers, assigning roles, and telling classmates what to do. In other words, part of the burden of cooperative learning appeared to come from coordination demands rather than from reading comprehension alone.

3.3 NotebookLM Was Viewed as Useful, but Its Outputs Were Often Difficult to Use Directly

Students in the NotebookLM-supported conditions generally understood the tool through an existing generative-AI frame of reference, often treating it as something similar to ChatGPT. This indicates that students did not approach NotebookLM as a completely unfamiliar technology, but rather as part of a broader AI-use mindset already present in their daily learning habits.

However, students also repeatedly reported that NotebookLM outputs were often too long, too dense, or insufficiently organized for direct classroom use. Even when prompts were simplified, the responses could still feel like a large block of information that required further sorting, interpretation, and reduction. This suggests that AI support did not simply remove difficulty; in some cases, it shifted difficulty from searching for information to processing, filtering, and restructuring AI-generated content.

3.4 AI Support Increased Efficiency, but Also Created a Risk of Superficial Completion

Another important theme concerns how students actually used AI-generated output in the worksheet process. Although NotebookLM could accelerate idea generation and provide quick support for summary or comprehension tasks, some students acknowledged that they occasionally copied AI-generated content directly into the worksheet without fully reorganizing or paraphrasing it. This indicates that AI support can simultaneously create efficiency gains and a risk of surface-level task completion.

Relatedly, students in the non-AI conditions often expressed that, if AI were to be used, they would most want it for high-load steps such as summarizing long passages or clarifying unfamiliar vocabulary. This finding indirectly supports the logic of the intervention design: students themselves intuitively located AI's value in reducing language barriers and supporting difficult steps in the reading process. At the same time, the interview data show that such support must be carefully structured if it is to promote comprehension rather than shortcut behavior.

3.5 Summary of the Qualitative Pattern

Taken together, the qualitative findings help contextualize the updated quantitative pattern observed across the four conditions. Rather than suggesting that AI support was uniformly effective or ineffective, the interview data indicate that students' learning experiences were shaped by how AI use, task structure, and participation demands were combined during the intervention. On the one hand, students recognized the potential value of NotebookLM for reducing language barriers and supporting difficult reading steps such as summarizing and clarifying unfamiliar vocabulary. On the other hand, they also reported that AI-generated responses were often too long, insufficiently organized, or easy to copy without deep processing. In cooperative settings, students further described coordination demands, uneven participation, and role-management burden, suggesting that part of the difficulty came not only from reading itself but also from managing the collaborative process. Viewed together with the quantitative findings, the qualitative evidence suggests that the instructional value of NotebookLM depends less on the mere presence of AI and more on whether the learning design provides sufficient individual accountability, structured output requirements, and support for transforming AI-generated information into personally meaningful understanding.

VI. Conclusions and Future Directions

1. Evidence-Based Conclusions Across the Two Studies

Across both studies, the Taiwan Albus pilot conveys a consistent message: digital transformation is most effective when tools are embedded in learning design, classroom routines, and accountability structures, rather than introduced as stand-alone add-ons. This conclusion is consistent with engagement-centered evidence in educational technology research (Bond et al., 2020), as well as Sun's motivation-and-engagement research showing that technology-enhanced learning effectiveness depends on how tools are integrated into participation structures, self-regulatory processes, and meaningful learning tasks rather than on technology presence alone (Sun & Rueda, 2012). In addition, digital literacy growth is more likely when the learning environment provides technology learning support that strengthens learners' autonomy and competence, as demonstrated in an SDT-based study co-authored by Sun (Chiu et al., 2022).

As summarized in Table 11, the two studies provide complementary forms of evidence for this conclusion. In Study 1, the pilot produced measurable improvements in students' and teachers' readiness to use digital and AI-supported tools, while qualitative evidence further showed that technology integration supported more interactive, student-centered, and workflow-connected teaching and learning practices. In Study 2, the quantitative findings did not show statistically significant between-group differences in situational interest, cognitive load, learning engagement, or self-efficacy; however, they also indicated that integrating NotebookLM did not reduce students' motivation or increase perceived learning burden. Most importantly, only the AI-IR showed a statistically significant improvement in overall reading comprehension, with the clearest gain concentrated in low-level reading comprehension.

Table 11. Summary of Key Findings and What They Mean

Domain	Study	Key evidence (quantitative)	Practical conclusion
Tool capability growth	Study 1	Students' overall comfort with Google Apps increased (3.05→3.18, p = .001).	Capability gains are scalable when tools are tied to tasks and feedback cycles.
Teacher readiness growth	Study 1	Teachers' overall comfort increased (2.85→3.41, p < .001), with strong gains in Slides, Forms, Chromebook, and Gemini.	Teacher readiness is a high-leverage scaling driver; ecosystem coherence reduces burden and enables redesign.
Barrier reduction	Study 1	Teacher AI barriers sharply decreased (e.g., access 57.1%→14.3%; awareness 66.7%→28.6%).	The pilot progressed from "adoption" to "optimization," indicating implementation maturity.
Student learning process	Study 1	Majority agreement on peer support, pace control, independent learning, and collaboration.	The core impact is a shift in learning processes, consistent with engagement-centered TEL findings (Bond et al., 2020).
Learning outcome pattern	Study 2	Only the AI-IR showed a significant pretest–posttest gain in overall reading comprehension (Z = -2.44, p = .015), with an additional significant gain in low-level reading comprehension (Z = -2.06, p = .040).	AI is not uniformly beneficial; its instructional value becomes clearer when paired with stronger individual accountability and structured meaning-making.
Motivation, process, and perceived burden	Study 2	No significant between-group differences were found in situational interest, cognitive load, learning engagement, and self-efficacy.	Across the four instructional conditions, neither the AI-supported nor the cooperative format was associated with lower motivation or higher perceived burden.

2. Conclusions for Study 1 (Project Albus Main Study)

2.1 The pilot produced measurable capability gains, especially where tools directly supported classroom learning tasks.

Students demonstrated a statistically significant improvement in their overall comfort with core Google Apps, with the clearest gains concentrated in classroom-embedded tools such as Practice Sets, Read Along, and Google Sheets. This pattern supports an evidence-based interpretation: digital transformation outcomes are not simply "usage increases," but reflect whether tools become instructional scaffolds that strengthen learner autonomy, task completion, and participation in structured learning routines (Chiu et al., 2022).

2.2 Teacher readiness improved even more substantially, strengthening the feasibility of sustainable scaling.

Teachers showed larger gains than students across the Google ecosystem, with particularly significant improvements in Google Slides, Google Forms, Chromebook, and Gemini. This suggests that the pilot enhanced not only teachers' general digital confidence but also their instructional and AI-related readiness. In practical terms, the strongest implementation effects emerged where the ecosystem reduced operational friction and enabled teachers to redesign learning flow, feedback processes, and classroom management rather than merely digitizing existing materials. This sustainability signal aligns with Sun and Rueda's (2012) work, which suggests that engagement in technology-mediated learning depends on motivational factors and self-regulatory processes rather than on technology use alone.

2.3 Implementation also appeared to mature over time: barriers decreased, and remaining constraints became more diagnostic than prohibitive.

The reduction of reported AI barriers among teachers suggests that the pilot successfully moved beyond the earliest stage of adoption. In implementation terms, the emergence of optimization issues, such as infrastructure reliability, device-fit decisions, bilingual scaffolding, and academic integrity routines, indicates a pilot that has reached routine use and is now ready for more evidence-informed refinement.

2.4 The most valuable "learning impact" signal was a shift in learning processes.

Student and teacher results converged on improvements in pacing, peer support, independent learning, and feedback-related classroom processes—dimensions repeatedly emphasized as central mechanisms of technology-enhanced engagement (Bond et al., 2020). This process-oriented interpretation also fits Sun's research tradition: motivational mechanisms and participation structures are not peripheral, but foundational to technology-supported learning effectiveness (Sun & Rueda, 2012).

3. Conclusions for Study 2

The findings of Study 2 suggest that the instructional value of NotebookLM was conditional rather than uniform across outcomes and instructional conditions. Most notably, only the AI-IR condition showed a statistically significant improvement in overall reading comprehension, and this gain was primarily concentrated in low-level reading comprehension rather than high-level reading comprehension. By contrast, the other three conditions did not demonstrate statistically significant gains in either overall reading comprehension or its subcomponents. These results suggest that the benefits of AI support were not automatic, but depended on how the tool was embedded within the instructional structure and learners' task responsibilities.

At the same time, the quantitative findings showed no statistically significant differences across the four conditions in situational interest, cognitive load, learning engagement, or self-efficacy. This pattern is important for interpretation. It indicates that the integration of AI did not appear to reduce students' motivation, weaken their engagement, increase perceived learning burden, or undermine their confidence in English reading. In other words, the present instructional design did not produce evidence of negative side effects in these motivational and process-related outcomes, even though strong performance gains were not observed across all conditions.

When viewed together with the qualitative findings, a more nuanced conclusion emerges. Students recognized NotebookLM as a potentially useful support tool, especially for summarizing, clarifying unfamiliar vocabulary, and reducing language-related barriers. However, they also reported that AI-generated responses could be overly long, insufficiently organized, or easy to copy without deep processing. In cooperative settings, students further described uneven participation, role-management burden, and coordination demands. These accounts suggest that the challenge was not simply whether AI “worked,” but whether the learning design sufficiently guided students to transform AI-generated information into personally meaningful understanding. From this perspective, the clearer gain observed in the individual AI-supported condition may reflect the advantage of stronger personal accountability during reciprocal teaching.

Overall, Study 2 suggests that AI can be integrated into English reading instruction without increasing cognitive burden or damaging students’ motivational responses, but its academic benefits are more likely to emerge when paired with an instructional design that emphasizes individual accountability and structured meaning-making. Therefore, the practical implication of this study is not simply that AI should or should not be used in reading instruction, but that its effectiveness depends on whether teachers provide clear task structures, output requirements, and evidence-based learning routines that prevent superficial completion and support deeper comprehension.

4. Practical Recommendations for Scaling

These recommendations are grounded in an engagement-and-mechanism perspective supported by the literature: educational technology is most likely to improve learning when it strengthens participation, feedback, and motivational processes rather than merely increasing access to tools (Bond et al., 2020; Sun & Rueda, 2012). Across the two studies, the Albus Taiwan pilot suggests that future scaling should focus not only on expanding tool use, but also on refining the instructional strategies under which technology can support meaningful learning.

4.1 Strengthen reliability foundations: infrastructure and access as conditions for equitable implementation.

Because infrastructure reliability directly shapes whether participation structures can operate smoothly, scaling should continue to prioritize stable connectivity, device readiness, and purpose-aligned access strategies. This recommendation is especially important in Study 1, where implementation quality was closely tied to device fit, workflow coherence, and reduced technical disruption.

4.2 Upgrade professional development from "tool training" to "pedagogy-with-AI" routines

Teachers' readiness gains indicate that next-phase PD should shift from operational fluency to learning-design mastery: task design, formative assessment loops, differentiation, and AI pedagogy routines. This direction is supported by SDT-oriented technology learning support research showing that environments that support autonomy and competence can strengthen digital literacy outcomes (Chiu et al., 2022).

4.3 Institutionalize AI learning rules: evidence alignment, paraphrase discipline, and academic integrity

Given the observed risk of copy–paste behaviors, AI integration should explicitly teach learners how to use AI outputs as inputs for reasoning (paraphrase-and-justify; text-evidence alignment). This recommendation is consistent with generative AI scholarship emphasizing the need for pedagogical constraints and ethical norms (Kasneci et al., 2023).

4.4 Redesign AI-supported collaboration using "accountability-first orchestration"

Because cooperative learning effects depend on implementation quality (Kyndt et al., 2013), future AI-supported collaborative designs should prioritize traceable contribution, structured role scripts, and clearer output requirements. This revision is intended to reduce coordination burden and superficial completion, while strengthening the conditions under which AI can support reciprocal-teaching strategy use more productively.

4.5 Build a lightweight learning analytics routine for ongoing improvement

A key next step is to integrate process evidence (submission traces, revision cycles, role-rotation logs) into coaching and redesign. This recommendation is consistent with Sun et al.'s (2018) study, which shows that process data can reveal participation patterns invisible to traditional classroom assessment, and with Sun et al.'s (2019) study, which shows that feedback design can shape emotional/cognitive engagement and cognitive load.

5. Future Directions

5.1 Move from "Does AI help?" to "What orchestration makes AI help?"

Future cycles should treat AI as a pedagogical variable, not a binary access variable:

- capture prompting patterns, evidence citations, and revision traces (NotebookLM + Classroom logs),
- measure role fidelity, contribution balance, and quality of strategy talk during discussion,
- test whether redesigned collaborative AI protocols improve reading outcomes, accountability, and meaningful processing of AI-supported tasks.

5.2 Position AI literacy as SRL-oriented language learning

Given the pilot's signals on accountability and perceived learning, the next stage should explicitly teach AI literacy as SRL:

- planning (what I need from AI),
- monitoring (does AI match the text?),
- evaluating (what evidence supports the claim?),
- and reflecting (what did I learn, not just what did I submit).
  This direction is directly supported by Sun's work linking situational interest, self-efficacy, and self-regulation to engagement in technology-mediated learning (Sun & Rueda, 2012), and by Sun's recent synthesis on AI and self-regulated language learning (Chang & Sun, 2024).

5.3 Strengthen teacher adoption/acceptance evidence for long-term scaling

As the pilot scales, research should track teachers' acceptance and sustained use patterns, including perceived usefulness, ease of use, and facilitating conditions across different tool types and contexts—an approach aligned with evidence on technology acceptance in learning applications (Liu et al., 2021).

6. Closing Statement

In conclusion, the Albus Taiwan pilot provides a strong foundation for scaling because it generated measurable capability gains, reduced barriers, and—most importantly—produced actionable design rules for AI integration. Study 1 shows that when Google ecosystem tools are embedded in instructional workflows, both teacher readiness and student learning processes (agency, pacing, peer support) improve in meaningful ways. Study 2 further clarifies that the impact of AI on reading is not automatic; it depends on how AI is orchestrated within classroom social structures. In particular, the clearest achievement gains emerged when AI support was paired with stronger individual accountability. In contrast, collaborative AI-supported learning appears to require more deliberate orchestration (e.g., role scripts, contribution tracing, constrained AI outputs, and evidence alignment) to support meaningful learning rather than superficial task completion. These conclusions mirror the broader SSCI literature: sustainable technology integration depends on teacher capacity and supportive conditions (Ertmer & Ottenbreit-Leftwich, 2010), and effective LLM use requires new literacies, critical evaluation, and pedagogical safeguards (Kasneci et al., 2023).

VII. Appendix

Appendix A. Study 1 Teacher Survey (Pre/Post)

Appendix B. Study 1 Student Survey (Pre/Post)

Appendix C. Study 1 Classroom Observation Protocol

Appendix D. Study 1 Teacher Interview Guide

Appendix E. Study 1 Student Interview Guide

Appendix F. Sample Lesson Plans

Appendix G. Study 2 Worksheets and Learning Task Samples

Appendix H. Study 2 Instruments and Interview Guide

VIII. References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Pearson.

Bond, M., Bedenlier, S., Buntins, K., Kerres, M., & Zawacki-Richter, O. (2020). Facilitating student engagement in higher education through educational technology: A narrative systematic review in the field of education. *Contemporary Issues in Technology and Teacher Education*, *20*(2), 318-358.

Chang, W.-L., & Sun, J. C.-Y. (2024). Evaluating AI's impact on self-regulated language learning: A systematic review. *System*, *126*, Article 103484. https://doi.org/10.1016/j.system.2024.103484

Chen, A., Darst, P. W., & P., P. R. (1999). What constitutes situational interest? Validating a construct in physical education. *Measurement in Physical Education and Exercise Science*, *3*(3), 157-180. https://doi.org/10.1207/s15327841mpee0303_3

Chen, A., Darst, P. W., & Pangrazi, R. P. (2001). An examination of situational interest and its sources. *British Journal of Educational Psychology*, *71*(3), 383-400. https://doi.org/10.1348/000709901158578

Chiu, T. K. F., Sun, J. C.-Y., & Ismailov, M. (2022). Investigating the relationship of technology learning support to digital literacy from the perspective of self-determination theory. *Educational Psychology*, *42*(10), 1263-1282. https://doi.org/10.1080/01443410.2022.2074966

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Ebel, R. L. (1954). Procedures for the Analysis of Classroom Tests. *Educational and Psychological Measurement*, *14*(2), 352-364. https://doi.org/10.1177/001316445401400215

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, *42*(3), 255-284. https://doi.org/10.1080/15391523.2010.10782551

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T.,…Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kelley, T. L. (1939). The Selection of Upper and Lower Groups for the Validation of Test Items. *Journal of Educational Psychology*, *30*(1), 17-24. https://doi.org/10.1037/h0057123

Kuder, G. F., & Richardson, M. W. (1937). The Theory of the Estimation of Test Reliability. *Psychometrika*, *2*(3), 151-160. https://doi.org/10.1007/BF02288391

Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning: Do recent studies falsify or verify earlier findings? *Educational Research Review*, *10*, 133-149. https://doi.org/10.1016/j.edurev.2013.02.002

Leppink, J., Paas, F., Vleuten, C. P. M. V. d., Gog, T. V., & Merriënboer, J. J. G. V. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058-1072. https://doi.org/10.3758/s13428-013-0334-1

Lin, H. C. S., Yu, S. J., Sun, J. C. Y., & Jong, M. S. Y. (2021). Engaging university students in a library guide through wearable spherical video-based virtual reality: Effects on situational interest and cognitive load. *Interactive Learning Environments*, *29*(8), 1272-1287. https://doi.org/10.1080/10494820.2019.1624579

Liu, Y., Sun, J. C.-Y., & Chen, S.-K. (2021). Comparing technology acceptance of AR-based and 3D map-based mobile library applications: a multigroup SEM analysis. *Interactive Learning Environments*, *31*(7), 4156-4170. https://doi.org/10.1080/10494820.2021.1955271

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117-175. https://doi.org/10.1207/s1532690xci0102_1

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. The University of Michigan.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*(4), 479-530. https://doi.org/10.3102/00346543064004479

Spörer, N., Brunstein, J. C., & Kieschke, U. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and Instruction*, *19*(3), 272-286. https://doi.org/10.1016/j.learninstruc.2008.05.003

Sun, C.-Y., Lin, C. T., & Chou, C. (2018). Applying learning analytics to explore the effects of motivation on online students' reading behavioral patterns. *The International Review of Research in Open and Distributed Learning*, *19*(2), 209-227. https://doi.org/10.19173/irrodl.v19i2.2853

Sun, J. C.-Y., Chang, K.-Y., & Chen, Y.-H. (2015). GPS sensor-based mobile learning for English: An exploratory study on self-efficacy, self-regulation and student achievement. *Research and Practice in Technology Enhanced Learning*, *10*, Article 23. https://doi.org/10.1186/s41039-015-0024-y

Sun, J. C.-Y., & Rueda, R. (2012). Situational interest, computer self-efficacy and self-regulation: Their impact on student engagement in distance education. *British Journal of Educational Technology*, *43*(2), 191-204. https://doi.org/10.1111/j.1467-8535.2010.01157.x

Sun, J. C.-Y., Yu, S.-J., & Chao, C.-H. (2019). Effects of intelligent feedback on online learners' engagement and cognitive load: the case of research ethics education. *Educational Psychology*, *39*(10), 1293-1310. https://doi.org/10.1080/01443410.2018.1527291

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0