ASERS

Journal of Research in Educational Sciences

Bi-annually
Volume XVI,
Issue 2(20) Winter 2025
ISSN: 2068 – 8407
Journal DOI: https://doi.org/10.14505/jres

ASERS Publishing

Who we are ...

ASERS Publishing was founded in 2010 and is one of the first online academic publishers of Romania.
Its objectives are excellence in research, scholarship, and education by publishing worldwide. We publish for a range of audiences in print and digital formats that work best for them.
Our academic publishing program serves scholars, instructors, and researchers, publishing and discovering research across subject areas in the social sciences, sciences, and humanities.

Mission

ASERS Publishing serves the needs of the academic community with a mission to help them accelerate knowledge for a better, more innovative and knowledge-intensive world, and we do this by facilitating critical information and decision-making for customers in the global research and academic ecosystems, and making science open – so that scientists can collaborate better and innovate faster.
Research is the foundation of modern society and it's thanks to advances in science that we enjoy longer, healthier and more prosperous lives than ever before in human history. We want to make science even more powerful by ensuring it is openly available. This way, society will be able to generate more knowledge and accelerate innovation, health and prosperity for all.

Our core values are a golden thread through everything we do, and are key to achieving our mission. We are unwavering in our mission and act with urgency to make all science openly available. We know that immediate, free access to high-quality research is vital for achieving this. With open access publishing, knowledge is being shared immediately, without restrictions, enabling researchers to collaborate better and innovate faster to solve the critical challenges we face as humanity.

Scientists power modern society. It is thanks to advances in science that we enjoy longer, healthier and more prosperous lives than ever before in human history. That's why every decision we make starts with one consideration: what's best for the researcher – the people doing the science? Our role is to provide the world's scientists with the best, the fastest, and the most efficient publishing experience. This approach forms the basis of our core values, which we believe are crucial to achieving our mission.

Collaboration. We are dedicated to building communities and working together to empower each other.

Innovation. We are bold in our decision-making, driven by technology and innovation, and constantly making improvements.

Because informed decisions lead to better outcomes. We are proud to help research and academic community advance science and improve health outcomes by combining quality information and data sets with analytical tools to facilitate insights and critical decisions. We help researchers share knowledge, collaborate and make discoveries. We deliver insights that help universities, research institutions, governments and funders achieve strategic goals**.** ASERS Publishing supports R&D-intensive corporations, research institutes and universities with data-led insights that drive innovation.

Serving the research community. We place the researcher at the center of everything we do.

Driving confidence in research. Building on more than 15 years of collaboration with the research community we help ensure that quality research can be accessed, trusted, shared and built upon to accelerate progress in society. We work to ensure information is validated and discoverable so that your work can make a difference, and we create tools and platforms to help you showcase your impact.

Helping people access quality research. We are rigorous in everything we do and committed to only the highest quality standards. Every researcher has a fundamental right to publish in the way that works for them. We offer you a choice of publishing open access or through the subscription model depending on what is right for you.

Inclusion diversity, and equity in research. By bringing together diverse ideas and different perspectives gender, race, ethnicity, and geography we can help you drive progress more effectively. With a diverse group of people creating solutions for customers worldwide, ASERS Publishing is helping to lay the groundwork for more diverse communities in science, and helping you create a more equitable future.

Vision

Our vision is to support the seekers, the changers, the innovators, those who see knowledge as a force for a better, more innovative and knowledge intensive world and building bridges to knowledge - supporting the development of ideas through the research process.

Volume XVI, Issue 2(20), Winter 2025

ASERS Publishing
http://www.asers.eu/asers-publishing
Journal of Research in Educational Sciences
https://journals.aserspublishing.eu/jres
ISSN 2068 – 8407
Journal DOI: https://doi.org/10.14505/jres
Issue DOI:
https://doi.org/10.14505/jres.v16.2(20).00

Table of Contents

The Journal is designed to promote scholars' thought in the field of education with the clear mission to provide an interdisciplinary forum for discussion and debate about education's most vital issues. We intend to publish papers that contribute to the expanding boundaries of knowledge in education and focus on research, theory, current issues and applied practice in this area.

The Editor in Chief would like to invite submissions for the **Volume XVII, Issue 1(21), Summer 2026** of the **Journal of Research in Educational Sciences** (JRES).

The primary aim of the Journal has been and remains the provision of a forum for the dissemination of a variety of international issues, empirical research and other matters of interest to researchers and practitioners in a diversity of subject areas linked to the broad theme of educational sciences.

The aims and scope of the Journal includes, but is not limited to; the following major topics as they relate to the Educational Sciences:

- Educational Psychology;
- Engagement and Community;
- Leadership in Education;
- School Improvement;
- Human Resources in Education;
- Education and Information Science;
- Global strategies in Higher Education;
- Learner's Needs in the 21st Century;
- The Role of Education in The Globalization World;
- Technology-Based Learning.

All papers will first be considered by the Editors for general relevance, originality and significance. If accepted for review, papers will then be subject to double blind peer review.

Deadline for Submission:	25th May 2026
Expected Publication Date:	June 2026
Web:	https://journals.aserspublishing.eu/jres
E-mail:	jres@aserspublishing.eu

https://doi.org/10.14505/jres.v16.2(20).01

From Developers to Players: Exploring the Dual Impact of Game-Based Learning on Student Engagement, Learning and Skill Development

Alice CANAVESI
University Carlo Cattaneo, Italy, and Intelligence, Complexity and Technology Lab (ICT Lab), LIUC - Carlo Cattaneo University, Castellanza, Italy
acanavesi@liuc.it

Francesco BERTOLOTTI
University Carlo Cattaneo, Italy
fbertolotti@liuc.it

Laura CATTANEO
University Carlo Cattaneo, Italy
lcattaneo@liuc.it

Giacomo BUONANNO
University Carlo Cattaneo, Italy, and Intelligence, Complexity and Technology Lab (ICT Lab), LIUC - Carlo Cattaneo University, Castellanza, Italy
buonanno@liuc.it

Luca MARI
University Carlo Cattaneo, Italy, and Intelligence, Complexity and Technology Lab (ICT Lab), LIUC - Carlo Cattaneo University, Castellanza, Italy
lmari@liuc.it

Aurelio RAVARINI
University Carlo Cattaneo, Italy
aravarini@liuc.it

Article info: Received 22 September 2025; Revised 6 October 2025; Accepted 23 October 2025; Published 30 December 2025.

Abstract: This paper explores the implementation of Game-Based Learning (GBL) in higher education, with a particular focus on the dual role of students as developers and players of educational games. In particular, the study investigates the extent to which GBL leverages intrinsic motivation, fosters engagement, consolidates prior knowledge, and cultivates essential 21st-century skills in undergraduate students that designed, developed, and played interactive digital gamebooks on STEM disciplines.

A mixed method based on surveys and focus groups was leveraged to assess the diverse learning outcomes and the educational impact of the project.

Findings highlights that the GBL approach significantly increased student engagement, fostered deeper learning in IT-related skills, and provided valuable experiences in competences such as project management and teamwork. However, the integration of advanced academic content in Mathematics and Statistics was perceived as less effective, posing challenges to knowledge acquisition and consolidation.

This study shows that while GBL is highly effective in promoting motivation and skill development, further refinement is needed to align content complexity with learning objectives.

Keywords: Game-Based Learning (GBL), student engagement, 21st-Century skills, intrinsic motivation, higher education, STEM Education

JEL Classification: I21; I23; C00.

Introduction. Digital Natives and New Educational Approaches

Growing up in an interconnected, multi-tasking and rapidly changing environment, the new generation of digital native students is difficult to engage with, as they tend to be more autonomous in seeking out information, tools and feedback from multiple sources, and in constructing their own learning. Educators now recognize the need to redesign pedagogical practices that engage digital natives by leveraging primarily on their intrinsic motivation (Thurston, 2018; Rezai, 2025) and promote diverse (*e.g.*, cognitive, social and emotional) abilities while enhancing deep learning (Hartt *et al.* 2020; Rossouw and Steenkamp, 2025; Maceiras *et al.* 2025).

The use of games in educational contexts is one of these innovative practices that exploit the form of collaborative experience. Game-based learning (GBL) has been acknowledged by scholars as a promising solution in higher education (Belova and Zowada, 2020) to motivate higher education students and improve their learning performance across diverse disciplines (Khoo *et al.* 2025).

The present paper deals with the implementation in higher education of a GBL practice based on digital non-linear and interactive storytelling. The purpose is to investigate the extent to which GBL leverages students' intrinsic motivation to engage with the subject, consolidates prior knowledge and facilitates the development of new knowledge and skills by exploring students' role as both developers and players. In the following sections, this paper aims to: (i) overview the recent literature on GBL; (ii) explore different theories underlying this methodology and present a possible theoretical framework for the purpose of the study; (iii) describe the design, implementation and use of gamebooks by students of a private university in the dual role of developers and players; and (iv) discuss the main learning outcomes obtained through a pilot course.

1. Theoretical Background

1.1 Game-Based Learning

The term 'game-based learning' (GBL) has a broad meaning that refers to the use of games designed for learning purposes (Aydin and Cakir, 2022). The main purpose of these games, whether on digital support or not, is to promote the active participation of learners and facilitate learning by including challenges, rewards, interaction and feedback (Alsawaier, 2018). Examples of GBL activities include educational videos (*i.e.*, videos used as an educational resource), simulations, digital games, quizzes and crossword puzzles (Tham and Tham, 2012).

The widespread use of games in the digital format for educational purposes has been made possible due to the adoption of ICT tools in education as well as to the increasing acknowledgement of the importance of playing in learning. These games are designed with the aim of supporting learning by identifying a balance between educational and entertainment purposes. They are generally characterized by activities, rules and constraints that replicate those of the real world, to immerse the player in a challenging and verisimilar environment (Ariffin *et al.* 2014). Digital GBL solutions can enable reflective, experiential, and intriguing learning environments (Bygstad *et al.* 2022), affecting students' emotions and providing them with a memorable learning experience (Wang *et al.* 2025). Depending on the type of task they train for, games can be classified as skills and procedures learning games, action games, role-playing games and strategy games (Hays, 2007).

GBL is not limited to lower educational levels but has been shown to be useful even at university level. As such, it is an attractive and interesting area of research which is still limited in terms of outcomes (Greipl *et al.* 2020; Khoo *et al.* 2025).

GBL differs from gamification. Whereas the former is based on games or game-like environments with a focus on a certain learning activity, gamification consists in the integration or addition of game-design elements, game mechanics or game thinking in non-game contexts (Plass *et al.* 2015). In other words, GBL uses games as part of the learning process while gamification turns the learning process into a game (Al-Azawi *et al.* 2016).

1.2 The Impact of GBL in Education

In the past decade, a growing number of empirical studies have been conducted to explore GBL effectiveness in various domains of education such as STEM subjects (Hung *et al.* 2014; Zhao *et al.* 2022; Wang *et al.* 2022; Chen *et al.* 2025). When compared to traditional methods, it has been shown that GBL can lead to interesting learning outcomes (Wouters *et al.* 2013; Freeman *et al.* 2014, Boyle *et al.* 2016):

(i) Being an inherent aspect of human social behavior, gameplay has the potential to serve as a motivational tool for students (Felicia, 2010; Plass *et al.* 2015; Khoo *et al.* 2025): it provides an enjoyable social experience, encourages a feeling of independence and capability and engages learners both emotionally and cognitively (Hartt *et al.* 2020). This notion is grounded in Self-Determination Theory (SDT), according to which inner motivation is fostered by the three psychological needs of autonomy (*i.e.*, the need to feel in control of one's

own actions and decisions), competence (*i.e*., the need to feel effective and capable of achieving desired outcomes) and relatedness (*i.e*., the need to feel connected to others and to experience meaningful relationships) (Deci and Ryan, 1985). When intrinsic motivation is triggered in the first place and possibly but not necessarily integrated with extrinsic motivational components such as grades or rewards, learners are spurred on to actively engage in the designated activities (Hartt *et al.* 2020).

(ii) Many studies denote GBL as an effective learning methodology that fosters students' deep learning, facilitating the acquisition, consolidation and retention of knowledge (Ariffin *et al.* 2014; Al-Azawi *et al.* 2016; Jääskä *et al.* 2022; Maceiras *et al.* 2025). GBL facilitates knowledge acquisition by creating an interactive environment where students can engage with the content and take an active role in the educational activity. Games often require students to apply what they have learned repeatedly, a factor that, as Roediger and Butler (2011) explain, strengthens long-term memory through the testing effect. Finally, knowledge retention is also enhanced through GBL because it leverages both repetition and the emotional engagement associated with gameplay (Plass *et al.* 2015).

(iii) GBL has been shown to promote skills' development (Echao and Romero, 2017; Bakhsh *et al.* Anggoro, 2025). Qian and Clark (2016) have defined 21st century skills as a wide range of competences ranging from learning and innovation (*i.e*., critical thinking, problem solving, creativity, collaboration and communication) to information, media and technology. Educational games are believed to influence the development of these competences, not only during the play itself, but also during the design and development of the game (Qian and Clark, 2016). The latter two stages often require the development of a strategy, a leadership style, and social interactions, as they require participants to diverge, trade, and mediate between each other to produce an effective educational game (Schrier *et al.* 2024). This leaves space to synergetic learning and facilitates teamwork as the ability to collaborate effectively and efficiently towards the achievement of a common goal (Boyatzis and Ratti, 2009; Ku *et al.* 2013), on which engagement itself is grounded (Ryan and Deci, 2000a, Ryan and Deci, 2000b). Moreover, the collective development of gamebooks requires learners to handle a challenge and find a common solution to a problem (Chiong, 2010), therefore contributing to the development of creativity, communication (Qian and Clark, 2016), critical thinking and problem-solving abilities (Salas and Piccolo, 2009; Al-Azawi *et al.* 2016). Lastly and most importantly, when students participate in the process of designing and developing an educational game in the digital format, they also experience a complex task of project management, which refers to a combination of competences far beyond the technical skills. The organization of workflow and compliance with internal and external deadlines foster the development of leadership, strategic and business management skills as well as experiential reflective learning through 'learning by doing'. These are deemed essential project management competences (Ramazani and Jergeas, 2015; Cicmil and Gaggiotti, 2018) and are requested by organizations.

Despite the potential of GBL to increase motivation, stimulate learning, and develop higher-order skills, the effectiveness of these games has been mainly investigated for students playing games, with very little investigation of students developing the games. In other words, further investigations on how GBL affects learning, motivation and 21st century skills development in the dual role of player and developer are recommended.

1.3 Research Purpose

As highlighted in the previous section, while GBL has been extensively studied, there is still a gap in understanding its full potential in the learning process, especially in contexts where students engage in roles beyond pure participation.

To further explore the potential of GBL in fostering intrinsic motivation and enhancing the learning process and outcomes a dedicated learning project was designed to involve students not only as learners but also as creators within the GBL environment. The dual role played by students, acting simultaneously as developers and players, provides a unique dimension to our study. It allows us to observe and analyze the impacts of GBL from a holistic perspective, considering both the process of game creation and the experience of game-based learning. This duality adds a layer of complexity to the traditional GBL approach, enriching the learning experience and offering deeper insights into both the efficacy and the dynamics of game-based education.

1.3.1 The Context of The Study: The Learning Project ImpaRa

The learning project, named 'ImpaRa' (in Italian 'Imparare Raccontando', 'Learning by storytelling'), consisted in the design, implementation and use of 11 gamebooks by 34 first-year Management Engineering undergraduate volunteers at an Italian university, during a period of approximately four months (February – May 2023).

The project was characterized by three sequential macro-phases:

(iv) Macro-phase 1: design and implementation by participants (developer students) of interactive, web-playable non-linear stories within which to embed appropriate problems of Mathematics, Statistics, and Informatics;

(v) Macro-phase 2: experimentation of the stories developed in macro-phase 1, where all first-year students of the Management Engineering degree course (player students) were involved;

(vi) Macro-phase 3: collection of opinions on the experience, both of developer students and player students, followed by the analysis and the critical synthesis of data;

The game developers were coordinated by four team leaders (*i.e.*, two full-professors, a lecturer and a research fellow), all of them teaching the subjects targeted by the project. Students participating in the project were rewarded with up to 5 additional course points (towards the maximum grade of 30 at Italian universities).

Students' training was an essential aspect for the development of the game and was guaranteed throughout the duration of the project. Before starting macro-phase 1 students were trained both in the use of Twine, and in storytelling. During macro-phase 1, from mid-March to mid-May, students were mentored and facilitated by two young faculty members who acted as project team leaders. Two progress meetings were scheduled to review both the content and technical-informatics issues.

The gamebooks developed by students consist in non-linear, interactive, web-playable stories based on role-playing dynamics. Each gamebook was implemented using Twine, a free, open-source software system, which provides a high-quality nonlinear story development and use experience without the cost of purchasing licenses. Twine provides different developing frameworks, from which Harlowe was selected for its simplicity.

The gamebook narrative, typically in the fantasy or detective genre, is in the second person singular. The story proceeds in a branching and nonlinear way, so that it is not possible to read pages in their numerical order. The text alternates with evocative pictures of immersive and captivating scenarios of the setting, divided into various sections. At the end of each chapter, players are asked to solve a problem to continue their adventure. These challenges require disciplinary skills in Data Science; more than one solution is correct, with different solutions leading to different paths, in the realistic logic that the same problem can be solved in multiple ways. If playing in groups, students can interact with each other for the identification of a common solution and develop social skills. On receiving instant feedback on the correctness of the solution, the players are directed to the next challenge, based on the answer provided.

1.3.2 Research Objectives

The literature indicates that GBL fosters a sense of competence and emotional involvement making learning enjoyable and engaging, (Felicia, 2010; Hartt *et al.* 2020); it consolidates prior knowledge and promotes deep learning (Al-Azawi *et al.* 2016); it cultivates higher-order skills like critical thinking, problem-solving, creativity, and collaboration through social interaction and teamwork (Qian and Clark, 2016; Boyatzis and Ratti, 2009); and, additionally, it develops new competences including project management, leadership, and strategic business management through game design and development processes (Ramazani and Jergeas, 2015; Cicmil and Gaggiotti, 2018).

On the basis of this evidence, the gamebooks were designed and implemented with the purposes of investigating to what extent GBL:

(vii) Leverages intrinsic motivation to foster engagement and interest in the subjects (for both developers and players);

(viii) Consolidates prior knowledge acquired during the year through the courses of Mathematical Analysis, Experimental Data Analysis and Statistics and Informatics (for both developers and players);

(ix) Cultivates higher-order skills demanded by the workplace in the 21st century (for developers, only);

(x) Develops new IT-related knowledge and skills related to the use of Twine (for developers, only).

The achievement of these objectives served as the basis of the gamebooks' design and was subsequently evaluated through assessment questionnaires that were administered to the students who participated in the game, both as developers and as players, to measure learning motivation and effectiveness.

2. Materials and Methods

For the purpose of the study two separate analyses were conducted in order to gain a comprehensive understanding of the effectiveness of the GBL approach on students' motivation and learning outcomes.

The achievement of the objectives from (i) to (iv) was evaluated, firstly, through two different web-based assessment questionnaires that were administered to the students who participated, respectively as developers and as players, at the end of courses (June); and secondly, through three focus groups that were conducted by one of the researchers with 3 to 6 developer students who volunteered to take part in this further investigation in the following months (October and November).

The semi-structured questionnaire for developers consisted in 16 questions aimed at investigating the students' interest and engagement in developing the gamebook, the major challenges encountered, the main takeaways in terms of higher-order skills development, and the perceived effectiveness on their learning outcomes (*i.e.*, consolidation of prior knowledge and acquisition of new competences related to the IT)

The semi-structured questionnaire for players consisted in 8 questions aimed at investigating, on the one hand, the quality of gamebooks in terms of storytelling, disciplinary content, IT and graphic implementation and, on the other hand, the perceived effectiveness on their learning outcomes (*i.e.*, consolidation of prior knowledge acquired during the year and increase of interest in the subjects' content).

Focus groups in the form of 'informal discussions among selected individuals about specific topics were conducted to obtain perceptions about the overall project. In particular, 12 broad and open-ended questions were derived from both questionnaires and developed further to explore students' experiences more deeply both as developers and as players and to assess the overall perceived effectiveness of the project.

2.1 Participants

The participants in this project were first-year students of the bachelor's degree in Management Engineering attending the same courses of mathematics, statistics, and computer science in the university under study. Of the 34 students involved in the innovation project, 29 attended the final in-person session in June and took part in the initial analyses through questionnaires, resulting in a response rate of approximately 85% of the total population.

Out of these 29 respondents, 13 volunteered to participate in further analyses conducted through focus groups during the months of October and November, representing approximately 45% of the total population.

2.2 Data Collection and Analysis

Prior to the distribution of questionnaires and the conducting of focus groups, the researcher requested honest responses by guaranteeing anonymity and confidentiality for all participants and by ensuring that any sensitive information would not be disclosed.

With regards to questionnaires, two URLs were provided to participants, one containing the form for developers and one the form for players. Both questionnaires were completed by the end of the event. This approach enabled the electronic collection of data, leading to effective data entry and integration, as well as time reduction. Data were analyzed by one of the researchers and then revised by a second researcher to achieve better accuracy. The surveys are available in the additional material of the paper.

With regards to focus groups, all discussions were recorded with prior consent of participants and then transcribed for a more accurate analysis of data. An approach of theoretical coding was adopted to attribute explicit 'codes' for the generation of themes and sub-themes, which allowed us to capture the essence of observations and to gain meaningful insights (Onwuegbuzie *et al.* 2009). Also in this case, the analysis of data was first performed by one researcher and then refined by a second researcher.

2.3 Data Validity, Reliability and Triangulation

Data validity pertains to the accuracy of data measurement, while reliability denotes the consistency of the research approach across various researchers and projects (Gibbs *et al.* 2007; Creswell and Plano Clark, 2018).

The students completed the questionnaires immediately after the end of the project; this allowed them to recall the information easily, given the short time window.

Although the questionnaires may have been completed with some pressure, as students were asked to do so by their supervisors, they provided the basis for further investigation, which was conducted in the following months through focus groups and helped gain better insights into the topic. The time window between the quantitative and the qualitative analysis allowed us to assess whether the students' perceptions of the overall usefulness and effectiveness of the project as well as the positive work dynamics that emerged through the questionnaires were confirmed.

In the quantitative analysis, there was no direct intervention of the researchers that might have influenced participants' responses. In the qualitative analysis, the intervention of the researcher (*i.e.*, moderator) was limited to ensuring the achievement of the research objectives.

Data triangulation, namely a research process involving multiple individuals, data sources and methods that allow for a deeper understanding of the subject, was ensured: first, by corroborating the quantitative analyses based on questionnaires with the qualitative analysis deriving from the focus groups; second, by always including a second researcher to review the analysis of data; third, by investigating the phenomenon at different times with respect to the end of the project.

3. Research Results

3.1 Quantitative Analysis

In this section, we present a statistical analysis of data obtained from two separate questionnaires: one related to the experience evaluation and the other related to the stories' peer-to-peer evaluation. Although the analysis yields valuable insights, it is crucial to acknowledge its limitations. The sample size is notably small, rendering the analysis less robust. This limitation in data volume is attributable to the experimental setting: the project, being an innovative venture within a small university context, involved only a limited number of student participants. Consequently, the analysis should be interpreted as providing initial evidence to be confirmed through the focus groups, rather than a foundation for independent conclusions.

3.1.1 Developing Experience Questionnaire

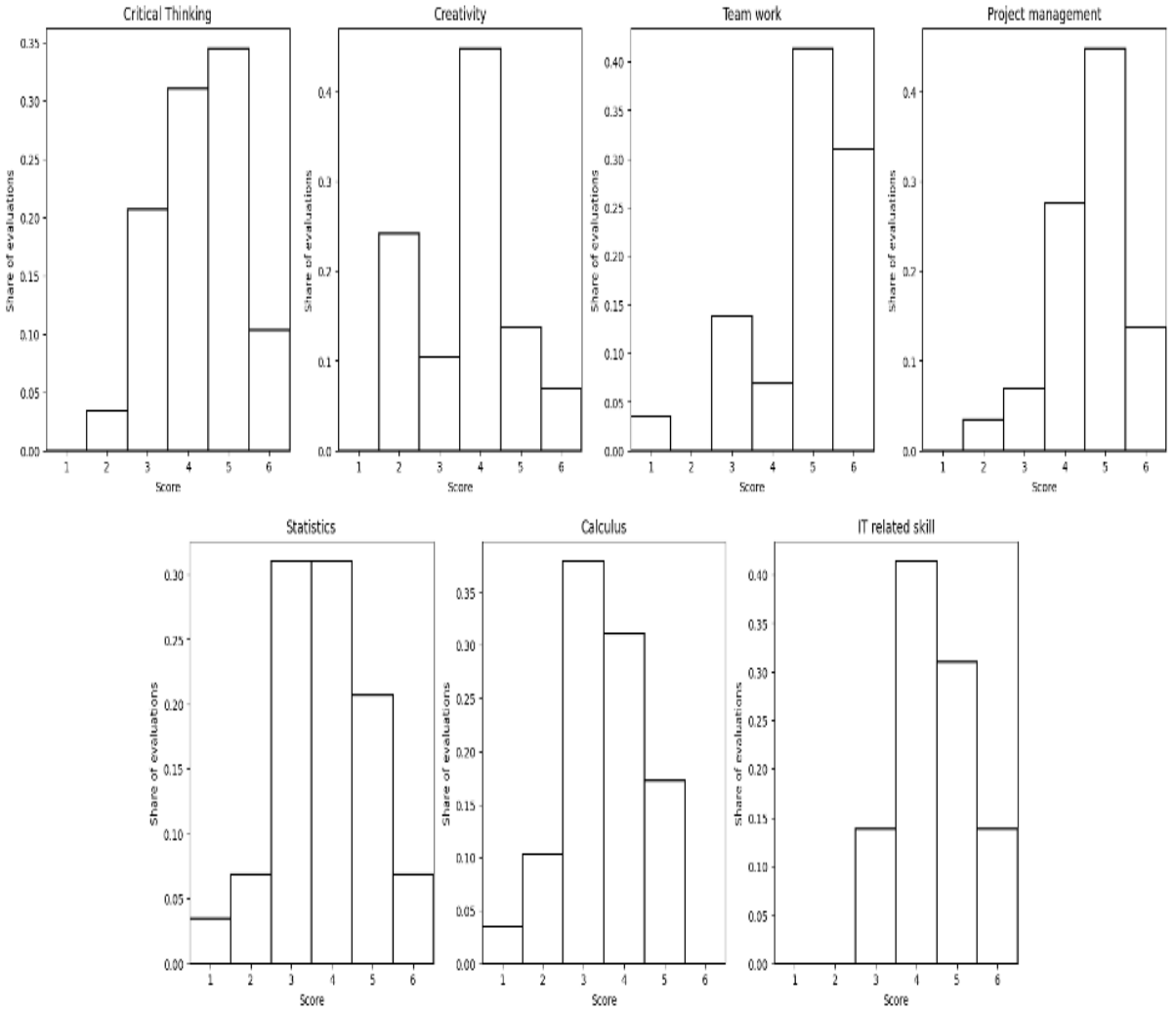
The questionnaire designed to evaluate the experiencers' perceptions was administered to students immediately after the event's conclusion but prior to the public disclosure of the results, as a strategy intended to mitigate potential biases. It comprised two types of questions: quantitative evaluation of specific features via scoring, and qualitative feedback through open-ended questions.

Participants rated motivation and six other key items, each of them standing for a skill or a knowledge set that was expected to be stimulated by the participation in the project: creativity, teamwork, project management, statistics, mathematics, and IT related skills (details about the questions can be found in the supplementary materials). The questionnaire utilized a six-point scale for item evaluation, ranging from 1 (the related skill did not increase at all) to 6 (the related skill increased a lot), with specifics detailed in Appendix A. The aggregated results of this evaluation are presented concisely in Table 1 and in Figure 1.

Table 1. Mean and standard deviation of the items collected in the developing experience questionnaire

Item	Mean	Std Dev
Critical Thinking	4.28	1.03
Creativity	3.69	1.20
Team work	4.82	1.25
Project management	4.61	0.96
Statistics	3.79	1.18
Mathematics	3.48	1.02
IT related skills	4.45	0.91
Motivation 1	4.83	0.76
Motivation 2	5.17	1.00

Figure 1. Frequency distribution of the items collected in the developing experience questionnaire



The central value of the employed rating scale was set at 3.5. Table 1 revealed that the mean score for nearly all evaluated items surpassed this midpoint, indicating a generally positive perception of the didactic experience. Notably, only the evaluation of mathematical skills approximated this central value, suggesting a more nuanced participant response in this specific area. This overall trend implies a positive impact of participation in such educational experiences.

The observed effect was particularly pronounced in the areas of project management and teamwork. These skills are intuitively enhanced in a group work setting, where participants are required to collaborate towards a deliverable within a set deadline, involving several intermediate stages. Considering that the participants were first-year engineering students, this activity probably represented their initial exposure to university-level group and project work, offering substantial learning opportunities. Similarly, the presence of a high score regarding critical thinking can be related to: the young age of the students, who have not yet been exposed to many problem-solving tasks which need it; and the quality of the experience, which is reflected in its proper design.

Conversely, the questionnaire indicated lower scores in the core academic domains the students were expected to strengthen: statistics and mathematics. This suggests that knowledge consolidation in these areas was perceived as less effective compared to the acquisition of other skills. The disparity may stem from the actual sensation of learning. In project work and teamwork, the continuous, intense experience of collaborating and striving to meet deadlines dominates the learning process. During the development of a software-based game book, the feeling of improving IT related skills is pervasive throughout the whole implementation process. In contrast, the application of statistics and mathematics, while crucial, occupies only a fraction of the activity time. Consequently, even though there is a certain process of knowledge consolidation occurring, it is perceived to a lesser extent.

Figure 2. Frequency distribution of the items related to motivation collected in the developing experience questionnaire

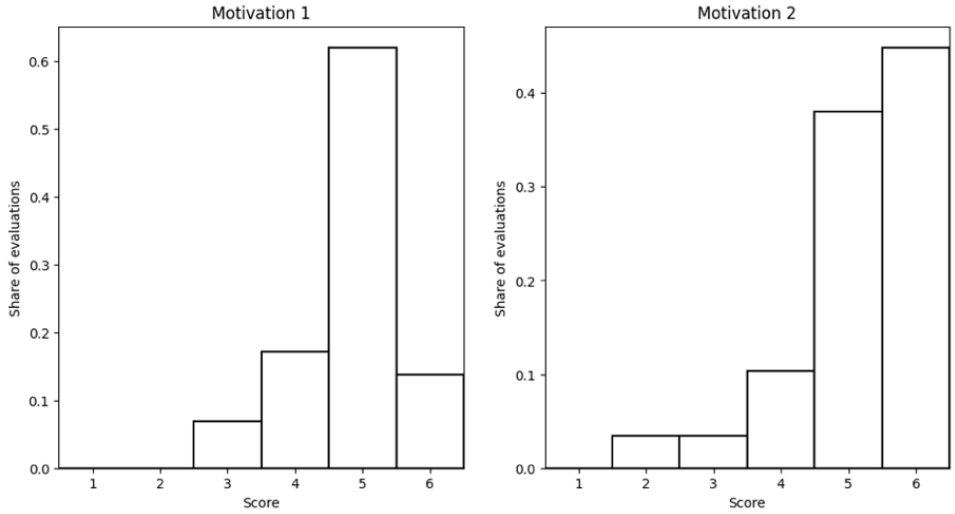
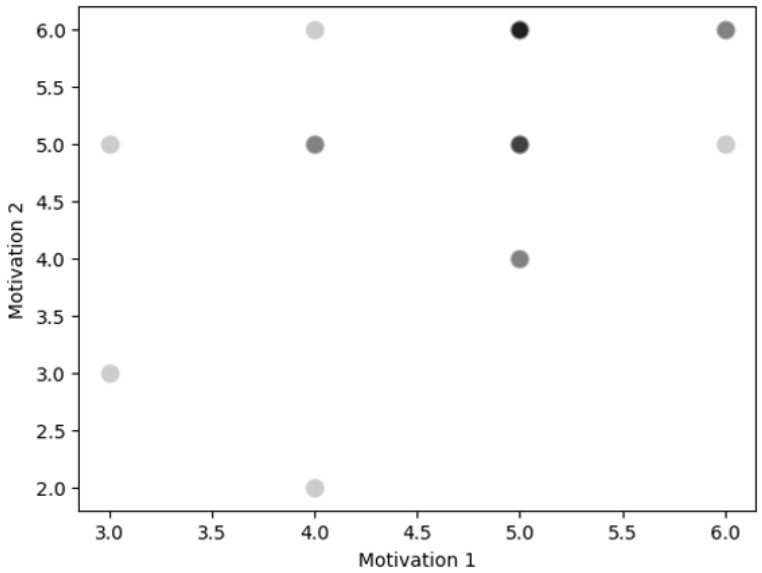


Figure 3. Scatter plot of the two motivation items employed in the survey



Motivation is a highly multifaceted and subjective concept that would require a separate theoretical framework. It has been measured in the present study with two questions, as a proxy of the students' level of engagement in and satisfaction with the project.

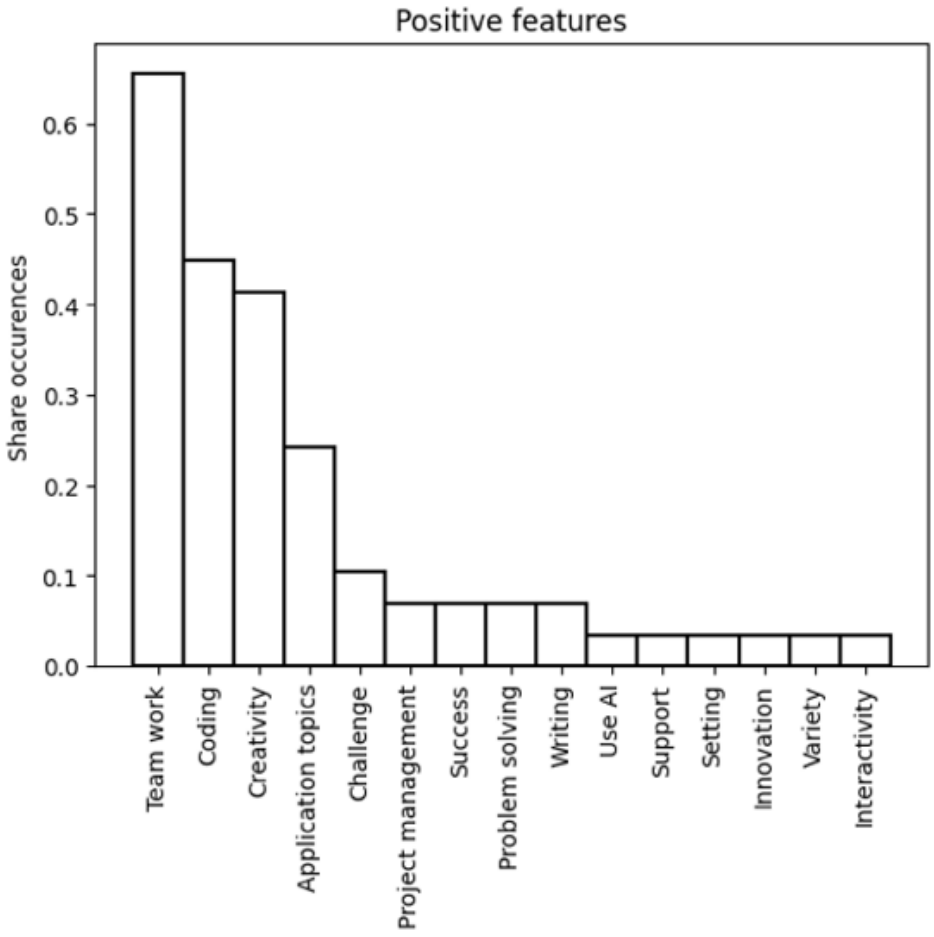
In order to ensure the accurate identification of motivation, two different metrics were utilized. The histograms in Figure 2, like those previously presented, exhibit a left-skewed distribution, indicating that the majority of participants exhibited high motivation on both questions used for measurement. Figure 3 illustrates that the two metrics are consistent with each other, given that they grow together. A point is denser when multiple observations are present, which can occur because integer values were used in the questionnaire.

The open-ended question in the questionnaire solicited participants to freely identify three positive aspects of their project involvement, aiming to capture their spontaneous perceptions without guiding their responses. The diverse array of responses, indicative of their individual experiences and priorities, is systematically presented in Table 2 and Figure 4.

Table 2. Occurrences of the features appearing in the open-ended question of the developing experience questionnaire

Feature	# occurrences
Teamwork	19
Coding	13
Creativity	12
Application of topics	7
Challenge	3
Project management	2
Success	2
Problem-solving	2
Writing	2
Use of AI	1
Support	1
Setting	1
Innovation	1
Variety	1
Interactivity	1

Figure 4. Share of the answers in which a feature occurs in the question regarding the positive elements of the developing experience. Given that they were asked up to three features per answer, the sum exceeds the unit



The distribution of positive elements identified by the students, as extracted by researchers from the open-ended responses, follows a Pareto-like pattern. Notably, 75% of the occurrences were concentrated within the first 25% of the categories. This significant clustering indicates a high degree of consensus among students regarding the positive aspects of the project, suggesting that its objectives and benefits were perceived

consistently and unambiguously by the majority of participants. This uniformity in perception highlights the project's clear scope and its impact on the students' learning experience.

This further analysis serves a dual purpose. Firstly, it corroborates certain findings obtained from the closed-ended questions, thereby reinforcing their validity. Secondly, it unveils additional insights that were not discernible through the closed-ended questions alone.

The confirmatory aspect of the analysis is evident in specific elements: teamwork, coding (representative of IT-related skills), and creativity rank among the top three positions. This alignment strongly reinforces the previously depicted perceptions with regards to the consistency in participants' views regarding these key aspects of the project experience.

Additionally, the open-ended responses reveal three elements not previously identified in the closed-ended questions.

First, over 20% of the students specifically mentioned the application of disciplinary topics in developing challenges as a positive element. This indicates that, while the average perception of specific topic learning was not overwhelmingly strong, a notable minority of students experienced a significant impact. This divergence suggests variability in how different students perceive and value the integration of theoretical knowledge into practical challenges.

Secondly, the concepts of 'success' (completion and achievement in tasks) and 'challenge' (the level of difficulty of the game) emerged from the responses. These elements can be linked to motivation, an aspect not explicitly measured in the closed questions. This finding prompts further investigation into the role of motivation in such projects. Specifically, it raises questions about whether motivation drives student participation in voluntary projects, whether participation in these projects enhances motivation, or if both dynamics coexist. Understanding this interplay is crucial, as it could significantly influence the design and appeal of similar educational initiatives in the future.

Thirdly, the aspect of critical thinking, although mentioned by only two students, still emerged in the analysis. This suggests that while critical thinking played a role, it was perceived as a less prominent element. Considering the project's primary focus on enhancing creativity rather than direct problem-solving, this outcome aligns with expectations, indicating that critical thinking was not a primary perceived benefit in an ingenuity context.

3.1.2 Playing Experience Questionnaire

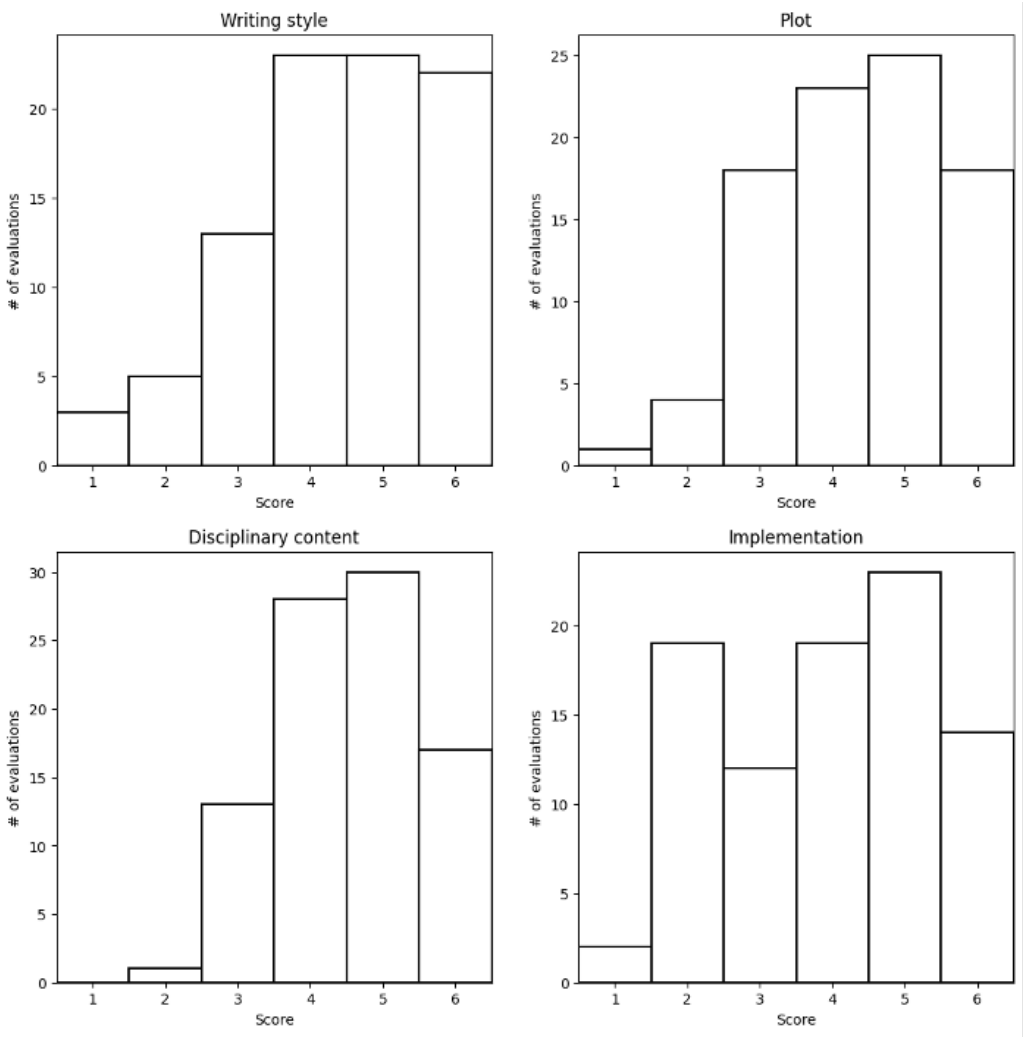
The second questionnaire focused on players peer-evaluating each story as players, not as developers. While analyzing story quality is not essential for assessing the learning effects of the design and implementation process, it is informative in determining whether the stories were positively or negatively received, as the quality of the book games produced can cast doubt on the actual development of the targeted skills, regardless of the students' self-reported perceptions. If the end product is poorly received, it could challenge the notion that the educational objectives were effectively met.

A total of 11 stories were developed, each by a different group of the student participants. These stories collectively garnered 86 evaluations, averaging 7.82 evaluations per story. While the gameplay experience was collaborative, the evaluation process was individualized, acknowledging the importance of personal experience and perception in the assessment. Each questionnaire was administered immediately after the completion of each game-story, capturing the immediate, raw responses of the players. This approach had its merits and drawbacks. On the positive side, it ensured that the initial, most impactful impressions were recorded, which is vital in determining whether the story's level was perceived as adequately engaging by the students. However, this method introduced an asymmetry in the judgment process; stories played later were evaluated against a benchmark set by earlier stories, which the first stories did not have. Despite this, the emphasis was on capturing the overall impression rather than conducting a rigorous statistical analysis, which the low numerosity of evaluation per story would anyway make impossible.

Table 3. Mean and standard deviation of the features identified by the playing experience questionnaire

	Mean	Std
Written style	4.62	0.59
Story	4.56	0.54
Disciplinary contents	4.41	0.84
Implementation	4.23	0.73
Overall	4.56	0.44

Figure 5. Frequency distribution of the features identified by the playing experience questionnaire



Similar to the previous questionnaire, each story underwent evaluation through a scoring system ranging from 1 to 6, where 1 indicated a very low score and 6 a very high score for specific features of the story.

Figure 5 illustrates the distribution of scores for each specific feature evaluated across the stories, while Table 3 consolidates these results. Importantly, both the chart and the table represent statistics from all 86 evaluations, rather than the mean scores for each individual story.

Observations reveal that the mean distribution score for each feature exceeds 4, indicating a predominantly positive assessment from the students. This suggests that the overall quality of the educational activity's output was, on average, sufficiently high. Additionally, the distributions exhibit a left-skewed tendency across some features, which implies a frequent occurrence of highly-rated plots, writing styles and disciplinary content. This trend potentially confirms the fact that the project effectively fostered students' creativity and their ability to articulate and express generated ideas. This outcome highlights the project's role not only in enhancing technical skills but also in developing students' creative and communicative competencies.

Moreover, the high mean value and skewed distribution observed in the disciplinary content evaluations suggest that the specific competences targeted by the project were effectively utilized in developing most of the book game. This indicates that, even though these subjects may not have been explicitly learned during the project, the students were able to apply them successfully in this context, demonstrating their ability to consolidate prior knowledge.

Lastly, the notable score in implementation, although the lowest among all categories yet still above 4, corroborates the enhancement of IT skills through the project. Given that Twine was a novel tool for all participants, the high-quality implementation of the book games implies effective learning and skill acquisition. This outcome suggests that the project successfully facilitated the development of new IT competencies among the students, as evidenced by their ability to proficiently use an unfamiliar technology in a practical application.

3.2 Qualitative Analyses

In this section, we present a qualitative analysis of data based on three focus groups conducted with a small number of students with the aim of corroborating the results of the questionnaires and gaining valuable insights. Each focus group lasted from 90 to 120 minutes. The first involved 3 students, the second involved 4 students and the last one involved 6 students. A moderator (*i.e.*, a researcher) facilitated an open and uninhibited dialogue (Powell and Single, 1996) throughout all sessions. 12 open-ended questions were designed for the discussion guide, but, where appropriate, the moderator was receptive to relevant unprepared issues raised by participants and encouraged equal participation of all members.

This methodology allowed us to explore a specific set of issues regarding the participants' views and experiences on the project and to assess their overall satisfaction.

3.2.1 Main Themes Identified

RQ1: How would you evaluate the project experience as a developer?

Nearly all participants stated that developing a gamebook was a positive and interesting educational experience, although in practice it was quite complex and challenging, especially because it took place during a very busy exam period and required a lot of time and effort.

In general, the project itself did not seem to significantly impact the learning of math and statistics, mostly because the challenges included in the game were simple, but it was certainly more useful for computer science since it required learning to use a software from scratch (*i.e.*, Twine). Moreover, it helped students learn to organize themselves and better manage an intense workload and group dynamics.

These are some of the most significant statements:

RQ2: How would you evaluate the project experience as a player?

The final in-class event where students were asked to play the gamebooks of their peers was found to be useful by some, as it was seen as an opportunity to compare their work with that of others.

Most participants, however, reported that it could have been handled differently. Two important sub-themes mentioned here were the lack of time and the noise in the classroom, which prevented students from focusing on the stories and giving their peers more reasoned feedback. The challenges of math and statistics presented in the gamebooks proved to be simple in most cases, and, when not, they were unrelated to the story and therefore less attractive. These different perspectives are presented with the statements below:

RQ3: Did you enjoy developing a gamebook?

All but two students enjoyed developing a gamebook, especially because it was an optional project that they voluntarily pursued, as reported in the following lines:

Overall, the part of designing the game played the biggest role in improving the students' motivation rather than simply playing the game, as already highlighted in the literature by Divjak and Tomic (2011). Qualitative analyses also showed that, when investigating the overall satisfaction and commitment to the project, these were related to elements of intrinsic motivation. In most cases, students said they were satisfied with the implementation of the gamebooks because of their interest and curiosity about the nature of the project itself, the challenge to themselves to be able to complete what they had started, and, finally, the sense of self-determination and self-actualization that they had developed along the way.

RQ4: Has your approach and/or interest in the subjects (i.e., statistics, mathematics and computer science) changed? If yes, how?

Everyone agreed on the fact that the project was too short to trigger a change in approach or to increase interest in the subjects involved, especially with regards to mathematics and statistics. A different argument emerged for computer science, as learning how to use a new software (*i.e.*, Twine) greatly affected students' motivation.

RQ5: Did the project provide you with a better understanding of the content of these subjects (i.e., statistics, mathematics and computer science)? If yes, in what way?

With respect to mathematics and statistics, the project did not help consolidate prior knowledge acquired during the year, expect for one student. The challenges related to these two subjects that were included in the game were considered simple as they were initially designed for high-school students; on a general basis, they did not require a further review of the course programs.

With respect to computer science, a somewhat different perception emerged among students.

Some of them stated that the project allowed them to consolidate computer skills related to the use of Python of the previous semester; others perceived Twine as a completely different software, to the point that they talked about acquiring new skills and content.

RQ6: What was the teamwork experience like and what did it allow you to learn?

Team dynamics proved to be partially positive. Overall teamwork helped students understand how to manage a collective project and divide tasks according to skills and personal interest. In most cases, the dynamics also increased the sense of personal responsibility towards the group, taking away a considerable amount of time from exam preparation. However, some respondents pointed out two major challenges: having to take on the work of others in case of free-riding by one or more team members, and managing the repeated slowdowns due to the re-subdivision of tasks, which turned out to be uneven throughout the project due to an initial lack of information.

RQ7: How do you evaluate the experience of managing a deadline project and what are the major difficulties you encountered?

The main aspects that emerged in this case were the usefulness of the project in terms of good time management and proper organization of the workload, especially considering the exam period, as well as the difficulties that arose due to the failure of some individuals to provide their contribution.

Some also mentioned that the project supervisors provided useful timing guidelines to manage the workload.

RQ8: How difficult was it to choose the disciplinary content for the challenges included in the stories?

According to all participants, choosing the disciplinary content for the challenges and connecting it to the story with meaning proved to be one of the biggest challenges. The eligible topics were limited, also considering that the stories were supposed to be aimed at high-school students. A group started with the story and then tried to insert the challenges, but soon realized that they had to do the opposite: starting with the challenges and then building the plot around them.

RQ9: Did you enjoy devising the story from scratch and developing it into the form of a gamebook?

Creativity came out as a central aspect in fueling students' engagement. Most respondents stated that the fact of developing a story from scratch and with few constraints in the plot was of particular interest. Only three students admitted that they were not very inventive and leveraged on a tool of generative artificial intelligence (*i.e.*, Chat GPT and autoGPT) to receive support regarding some critical junctures and the mode of exposure.

RQ10: Is there anything else that you feel you have learned from the project and group work that you didn't learn from books or classes?

The only aspect that emerged besides the skills mentioned above (*i.e.*, creativity, teamwork, project management and critical thinking) was problem solving. Having to complete one's training independently, especially on the IT side, and to manage a project that had never been done before led to the emergence and management of numerous difficulties. The nature of the project itself required students to face difficulties with determination and to try to identify and implement the best solution in each situation.

RQ11: Did you find it difficult to learn how to use Twine and do you think it can be useful now?

Students stated that learning how to use Twine was challenging, not because of the software interface but rather because of the programming logic behind it. At the beginning, a high level of difficulty was perceived, which gradually decreased as time went by, partly with self-training and partly with the support of tutors (*i.e.*, second-year students of the Management Engineering degree program). Once again, generative artificial intelligence helped do some extent in setting the right codes for the implementation of the story on Twine.

RQ12: Do you have any particular comment or suggestion on the project itself?

When the students were asked this question, several issues were raised.

The training on Twine was found to be useful because it introduced some elements of diversity from the software they had been using. For this reason, they would have appreciated more hours of training dedicated to this topic.

In contrast, the training part on storytelling was generally perceived as not particularly valuable because, although it was interesting, it was too scattered and complex compared to the practical project. For most students, it would have been more useful to have the story of an already existing gamebook modeled and explained as an example. Some others disagreed, arguing that the added value of the project lay in the autonomous learning and the self-reliance.

Another important theme that emerged in this part of the analysis was that of responsibility and self-reliance: students really appreciated being able to do a project with very few constraints. The freedom and autonomy that characterized the work turned into personal satisfaction once the project was successfully completed.

Three critical aspects were also highlighted by respondents. The first aspect is related to time, which many felt was too tight to do the project well, especially considering that there were exam sessions during the months of March, April and May. From their point of view, it would have been useful to have different timelines: set up the work in December in order to start working on it in late January and February, instead of March.

The second aspect concerns the faculty and peer evaluation, as these were found to be quite poorly explained at the beginning of the project as well as at the very end.

The last aspect has to do with the purpose of the work, which was not made clear enough at the beginning of the project. Some respondents pointed out that it would have been interesting to put more emphasis on the final output of the project, presenting it to peers in the classroom but also, and especially, outside in high schools. These represent key factors to watch out for in the future, as they may affect students' motivation and engagement.

4. Discussions

In Section 3, we outlined the primary objectives of the research, which we aimed to achieve through the design and implementation of the gamebooks. After summarizing the results of both quantitative and qualitative analyses, it is evident that:

(i) There was a noticeable increase in student involvement and enjoyment in learning through GBL, particularly during the game creation process. The majority of participants exhibited high levels of motivation in both the gamebooks' development and play. Elements of intrinsic motivation clearly emerged when investigating the level of satisfaction and commitment to the project through the focus groups. During these groups students talked about their interest and passion for the nature of the project, their sense of responsibility and autonomy in managing roles and deadlines, their desire to improve their skills and feel competent in the assigned activities, their challenge to themselves to finish what they started even during the exam period, their sense of self-fulfilment, and, finally, their perception that the meaning and purpose of the project went far beyond the extra points in the exam. This underscores the effectiveness of GBL in making learning more engaging, not only playing games, but also developing them. Although some students initially participated due to extrinsic motivation, such as additional points towards their final grades, they ultimately found that the interest and engagement they developed were the most rewarding aspects of the project. The students expressed enjoyment in developing a gamebook and highlighted the satisfaction derived from joining and successfully completing a voluntary project that turned out to be quite challenging. Having participated as developers, the game phase of the book did not remain an isolated and decontextualized experience, but could also serve as a form of peer evaluation for students. In fact, most of the interviewed students emphasized that, during this phase, they focused not only on completing the story but also on studying the strengths and weaknesses compared to what they themselves had developed.

(ii) Integrating game development with subjects such as Mathematics, Statistics, and Computer Science was intended to allow students to apply their knowledge in a practical, hands-on manner. The results of the 'ImpaRa' project suggest that this was more successfully achieved in the IT domain, where students had to learn how to use new software and implement a digital interactive story on it. There was little impact in areas such as mathematics and statistics, due to the little time devoted to embedding disciplinary problems in the story and the simple level of challenges initially planned for high school students. Although the mean perceived impact of specific topic learning was not exceptionally high, a substantial subset of students reported considerable benefits. This heterogeneity in responses indicates variability in how individual learners interpret and value the application of theoretical concepts to practical scenarios.

(iii) Our third objective centered on equipping students with the skills that are essential for the real-world challenges demanded by the current job market. In this regard, teamwork, creativity and communication proved to be three key competencies that were greatly enhanced by the project.

These skills are vital in any job environment and our project demonstrated that GBL can be an effective means of developing them, especially when students are engaged as game developers. The team dynamics fostered an understanding of how to manage a group project and allocate tasks based on individual skills and interests, even in stressful situations, as the project was carried out during a period when students had to prepare for their end-of-semester exams. Furthermore, the project effectively fostered students' creativity and their ability to articulate and convey novel ideas. This outcome underscores the project's significance not only in its primary objectives of fostering students' motivation but also in developing students' creative thinking and communication skills. Critical thinking, which emerged in the analysis, was mentioned by only a few students both in the survey and in the focus groups, suggesting that it was perceived as a less prominent element. This may be due to the fact that students focused their efforts on the primary focus of the project, which was that of developing and implementing the story, rather than thinking about the best challenges to include within it.

(iv) We aimed to enhance students' proficiency in handling new technology, especially through learning how to use Twine software. Coding emerged to be one of the most significant constructs in both quantitative and qualitative analyses. The students excelled in the area of IT, mastering not only Twine but also acquiring and consolidating a broader range of technical skills. In fact, their prior knowledge of Python gave them a common language and an important foundation from which to develop more complex skills in Twine and beyond. This success exemplifies the adaptability of GBL in teaching practical IT-related skills to undergraduate students.

In general, the overall quality of the educational activity's output was perceived to be sufficiently high.

4.1 Limitations

The limitations of this study are primarily related to the methodologies adopted.

With regards to the quantitative analyses, the main limitation lies in the small size of the sample, which means that the results of the analysis are not highly meaningful or generalizable. Moreover, as the questionnaire participants were junior members, who were invited to take part in the study by their professors or lecturers, they may have felt a certain degree of obligation. Nevertheless, there was no direct communication between the researchers and the participants, thus eliminating the potential of influencing their responses. The research bias was further reduced by providing the confidentiality statement, by asking semi-structured open-ended questions and by having another researcher review the data analysis performed by the first researcher, in order to achieve greater objectivity.

With regards to the qualitative analyses, a first limitation of focus groups consists in the potentially self-censoring and conforming influence exerted by the so-called 'group effect', whereby participants may follow the 'popular trend' and adjust their contributions according to the perceived location in the social expectations of the group (Powell and Single, 1996). Moreover, this methodology could encourage certain types of dominant behaviors by group members that would require constant control and possible intervention by the moderator. Last, it is argued that, compared to other types of qualitative methods (*i.e.*, in-depth interviews), focus groups may appear superficial, generating only surface information. On the other hand, leveraging explicitly on group interaction (Kitzinger, 1994), focus groups offer the advantage of uncovering the complete spectrum of perspectives held by participants and empowering them to enhance their contributions in response to points raised by others, thereby expanding on ideas that might remain underdeveloped in one-on-one interviews (Powell and Single, 1996).

A further limitation is related to the construct of motivation developed for the purpose of this study. GBL is known in the literature to increase the level of intrinsic motivation and this construct has often been measured through the adoption of appropriate and extensive theoretical frameworks, such as the ARCS Motivational Model

(Jääskä *et al.* 2022; Camacho-Sánchez *et al.* 2023; Ghani *et al.* 2024). Due to the fact that the objectives of this study were multiple (i, ii, iii, iv) and motivation was only one of them, it would have been problematic to administer a questionnaire to the students that was long and articulated enough to incorporate the ARCS motivational model or a similar framework. The two questions formulated in the survey to measure the students' overall engagement and satisfaction with the project resulted in a proxy for motivation that was then further explored during the qualitative interview. Otherwise, future studies should conduct two separate analyses, one related to motivation, which requires a separate framework, and one related to the other objectives mentioned above (consolidation of prior knowledge, cultivation of higher-order skills, and development of new technical skills) that emerged from the literature review.

Another limitation of a different nature concerns the learning outcomes we expected to achieve from the project, particularly for the courses of Mathematics and Statistics. As seen also from the results of the questionnaire analysis and from what emerged in the focus group discussions, the students did not perceive a particular advancement in their knowledge of these subjects, for different reasons. Firstly, at the beginning of the project, the target audience for the gamebook players was set to high school students. For this reason, the student developers kept the level of challenges in the gamebooks relatively low, which did not allow them to sufficiently deepen their mathematical and statistical knowledge. However, it is easily understandable that, had the final target been different, it would have been particularly complicated for the student developers to include challenges that contained particularly advanced content of math or statistics. This would indeed have forced them to use part of the project development time to independently explore some topics in depth and understand how to possibly integrate them into the storytelling. As such, it would have been complicated, also because of the limited time available for the development of the project. Nonetheless, it would be very interesting, in a learning-by-doing approach, to give students the opportunity to learn new things during the process of game development. The same consideration can also be made for the student-player. The results highlighted that, in playing the books developed by their colleagues, the students encountered a rather low and probably not very engaging level of mathematical-statistical challenges. To provide the player with an experience that is truly relevant and satisfying, it would be desirable to include challenges pushing the player to seek new knowledge in a context of critical thinking, problem-solving, and personal satisfaction for the results achieved even autonomously.

Conclusions and Further Research

The body of literature analysed in section 2 consistently highlighted the effectiveness of GBL in enhancing student engagement and deepening their understanding of complex concepts. This aligns seamlessly with the findings from our 'ImpaRa' project, where we observed a notable increase in student involvement and enthusiasm towards the subjects when engaged through gamebooks. This practical experience reinforces the theoretical framework, suggesting that GBL not only captivates students' attention but also fosters an immersive learning environment.

Furthermore, studies from other scholars indicated that GBL could act as a catalyst for intrinsic motivation, leading to a more self-directed learning approach. In 'ImpaRa', students also demonstrated a proactive attitude in collaborating and effectively managing tasks, supporting the notion that GBL can effectively translate theoretical motivation into tangible learning outcomes and engagement.

Also, the enhancement of 21st-century skills through GBL suggested by the literature finds concrete evidence in 'ImpaRa' project, where we observed significant improvement in students' project management, teamwork, and IT skills.

Overall, the present study significantly contributes to the growing body of evidence supporting GBL as a transformative tool in modern education, bridging the gap between theoretical understanding and practical application. The study novelty lies in its empirical approach, which moved beyond standard GBL to engage students as active developers of interactive gamebooks. Our findings suggest that this type of GBL is not only effective in theoretical knowledge acquisition but also crucial in acquiring practical skills necessary for the modern job environment.

Particularly, the developer role of students proved critical boosting overall involvement and successfully cultivating deep intrinsic motivation among participants, who reported a sense of self-fulfilment and purpose far exceeding initial extrinsic incentives. Furthermore, the study's outcomes underscore the project's vital role in demonstrating significant enhancement in essential skills such as project management and teamwork. Finally, the project's success in fostering practical IT proficiency, exemplified by the mastering of Twine and consolidation of technical skills, further highlights the adaptability of this GBL format in developing highly relevant competencies for undergraduate students.

The results of the questionnaires and group discussions, although limited in the number of participants, highlighted some critical issues that could be improved in the future and some aspects that should be investigated more thoroughly.

Firstly, it becomes evident that a project of this type requires a significant amount of development time, especially when engaging the student as a game developer. The time provided must be sufficient for two main activities. Initially, it should allow the student to learn the use of the Twine software at least at a basic, consolidated level. More advanced skills would be acquired during the project development in a relatively spontaneous process of learning by doing.

Secondly, it should enable the student developers to better explore the narrative structure of the story and include mathematical-statistical challenges of a more advanced level. To do so, students have to deepen their knowledge independently and understand how it can be integrated into the story, without being trivial or obvious. In this way the learning process could be really enhanced.

For these reasons, a second edition of the 'ImpaRa' project could be envisaged, which would be developed over the entire academic year and not just a few months, allowing for greater student participation. Indeed, proposing the project towards the end of the year had a somewhat distorting effect, resulting in participation mainly from students who felt already prepared and more confident, and were already motivated. Students who may have encountered difficulties during the year preferred to focus on studying and preparing for the final exams. Yet, this type of learning process could particularly help students who face some challenges. Placing them in a group work context could help them create positive relationships, which could be useful not only in completing the specific project, but generally as support and motivation for learning throughout the year.

Last but not least, data collection could also be more systematic and comprehensive, focusing on different phases of the book's development, with the administration of ad-hoc questionnaires and the possibility of more feedback from the researchers overseeing the project. The sample would be larger and less biased, ensuring a more robust final analysis, especially regarding the expected learning outcomes.

Acknowledgments:

The ImpaRa project was carried out thanks to the financial resources provided through the 'LIUC Special Projects of Education Innovation 2023' initiative.

Declaration of Competing Interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Due to the sensitive nature of the questions asked in this study, survey respondents were assured that their data would remain confidential. However, the raw data underlying this study can be made available by the authors upon reasonable request, provided that ethical and confidentiality considerations are respected.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

All participants provided written informed consent to participate in this study and to publish the research results.

This study did not require ethical approval and informed consent as it did not involve human participants, human data, or animals.

Credit Authorship Contribution Statement:

Alice Canavesi: conceptualization, investigation, methodology, project administration, formal analysis, writing – original draft, data curation, validation, writing – review and editing
Francesco Bertolotti: methodology, project administration, formal analysis, writing – original draft, data curation, validation, visualization
Laura Cattaneo: project administration, formal analysis, writing – original draft, supervision
Giacomo Buonanno: conceptualization, project administration, writing – review and editing
Luca Mari: conceptualization, project administration, writing – review and editing
Aurelio Ravarini: conceptualization, project administration, writing – review and editing

Declaration of Use of Generative AI and AI-Assisted Technologies:

The authors declare that they have not used generative AI and AI-assisted technologies during the preparation of this work.

References

[1] Al-Azawi, R., Bulshi, M., Farsi, F. (2016). Educational Gamification Vs. Game Based Learning: Comparative Study. *International Journal of Innovation, Management and Technology* (IJIMT), 7, 131-136.

[2] Alsawaier, R.S. (2018). The effect of gamification on motivation and engagement. *International Journal of Information and Learning Technology*, 35(1), 56-79. DOI: https://doi.org/10.1108/IJILT-02-2017-0009

[3] Anggoro, B., S., Dewantara, A., H., Suherman, S., Muhammad, R., R., Saraswati, S. (2025). Effect of game-based learning on students' mathematics high order thinking skills: A meta-analysis, 30(1). DOI:https://doi.org/10.1016/j.psicoe.2024.500158

[4] Ariffin, M., Oxley, A., Sulaiman, S. (2014). Evaluating Game-Based Learning Effectiveness in Higher Education. Procedia - Social and Behavioral Sciences, 123. DOI: 10.1016/j.sbspro.2014.01.1393

[5] Aydın, M.S., Cakır, N. (2022). The effects of a game-enhanced learning intervention on foreign language learning. Educational technology research and development, 70. DOI: 10.1007/s11423-022-10141-9

[6] Bakhsh, K., Hafeez, M., Shahzad, S., Naureen, B., Farid, M.F. (2022). Effectiveness of Digital Game Based Learning Strategy in Higher Educational Perspectives. *Journal of Education and e-Learning Research*, *Asian Online Journal Publishing Group*, 9(4), 258-268.

[7] Belova, N., Zowada, C. (2020). Innovating Higher Education via Game-Based Learning on Misconceptions. Education Sciences. DOI: https://doi.org/10.3390/educsci10090221

[8] Boyatzis, R. Ratti, F. (2009). Emotional, social and cognitive intelligence competencies distinguishing effective Italian managers and leaders in a private company and cooperatives. Journal of Management Development, 28, 821-838. DOI: https://doi.org/10.1108/02621710910987674

[9] Boyle, E.A., *et al.* (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games, *Computers & Education*, 178–192. DOI:https://doi.org/10.1016/j.compedu.2015.11.003

[10] Bygstad, B., Øvrelid, E., Ludvigsen, S., Dæhlen, M. (2022). From dual digitalization to digital learning space: exploring the digital transformation of higher education. *Comput. Educ.*, 182. DOI:https://doi.org/10.1016/j.compedu.2022.104463.

[11] Camacho-Sánchez, R., Serna Bardavío, J., Rillo-Albert, A., Lavega-Burgués, P. (2023). Enhancing motivation and academic performance through gamified digital game-based learning methodology using the ARCS model. *Interactive Learning Environments*. DOI: https://doi.org/10.1080/10494820.2023.2294762

[12] Chen, Y., Guo, Q., Qiu, Q., Qiao, C., & Wang, J. (2025). The Impact and Pathways of Digital Game-Based Learning on STEM Undergraduates' Engineering Concepts: Integrating Self-Determination Theory and Flow Theory. *Computer Applications in Engineering Education*, 33(4), e70066.

[13] Chiong, R. (2010). Programming with games. Special Issue on Game-based Learning. *Learning Technology Publication of IEEE Computer Society*, 12(1).

[14] Cicmil, S., Gaggiotti, H. (2018). Responsible forms of project management education: theoretical plurality and reflective pedagogies. *International Journal of Project Management*. Festschrift for Profes. J. Rodney Turner 36, 208–218. DOI: https://doi.org/10.1016/j. Ijproman.2017.07.005.

[15] Creswell, J.W., Plano Clark, V.L. (2018). Designing and Conducting Mixed Methods Research (3rd ed.). Thousand Oaks, CA: SAGE.

[16] Deci, E. L., Ryan, R.M. (1985). Intrinsic motivation and self-determination in human behavior. *Perspectives in Social Psychology (PSPS)*. DOI: https://doi.org/10.1007/978-1-4899-2271-7

[17] Divjak, B., Tomić, D. (2011). The impact of Game-based learning on the achievement of learning goals and motivation for learning mathematics - literature review. *Journal of Information and Organizational Sciences*, 35(1). Retrieved from //jios.foi.hr/index.php/jios/article/view/182

[18] Echao, O.F.S., Romero, M. (2017). Creative and collaborative problem-solving development through serious games co-creation, European Conference on Games Based Learning, Academic Conferences International Limited, 793–797.

[19] Felicia, P. (2010). What evidence is there that digital games can contribute to increasing students' motivation to learn? Brussels, Belgium: EUN Partnership AISBL

[20] Freeman, S., *et al.* (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111, 8410–8415. DOI:https://doi.org/10.1073/pnas.1319030111

[21] Ghani, A.M.T., Daud, W.A.A.W., Manan, K.A. (2024). Integration of the ARCS Motivational Model in Digital Game-based Learning for Sustaining Student Engagement in Communication. *International Journal of Religion.* DOI: https://doi.org/10.61707/sa9ded72

[22] Gibbs, G.R. (2007). Thematic Coding and Categorizing, Analyzing Qualitative Data. SAGE Publications Ltd., London. DOI: http://dx.doi.org/10.4135/9781849208574

[23] Greipl, S., Moeller, K. Ninaus, M. (2020). Potential and limits of game-based learning. *International Journal of Technology Enhanced Learning*, 12(4), 363–389. DOI: https://doi.org/10.1504/IJTEL.2020.110047

[24] Hartt, M., Hosseini, H., Mostafapour, M. (2020). Game On: Exploring the Effectiveness of Game-based Learning. Planning Practice & Research. DOI: https://doi.org/10.1080/02697459.2020.1778859

[25] Hays, R.T. (2007). Instructional Gaming Handbook: How to select and use instructional game. DOI:https://doi.org/10.1016/j.sbspro.2014.01.1393

[26] Hung, C.M., Huang, I. Hwang, G.J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, 1, 151–166. DOI: https://doi.org/10.1007/s40692-014-0008-8

[27] Jääskä, E., Lehtinen, J., Kujala, J., Kauppila, O. (2022). Game-based learning and students' motivation in project management education. *Project Leadership and Society*, 3. DOI: 10.1016/j.plas.2022.100055

[28] Khoo, Y.Y., *et al.* (2025). The impacts of game-based learning on thinking and learning in higher education context: A scoping review. *International Journal of Education in Mathematics, Science, and Technology (IJEMST*), 13(3), 623-637. DOI: https://doi.org/10.46328/ijemst.4776

[29] Kitzinger, J. (1994). The Methodology of Focus Groups: The Importance of Interaction between Research Participants. *Sociology of Health and Illness*, 16. DOI: http://dx.doi.org/10.1111/1467-9566.ep11347023

[30] Ku, H.-Y., Tseng, H., Akarasriworn, C. (2013). Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. Computers in Human Behavior, 29(3), 922–929. DOI:https://doi.org/10.1016/j.chb.2012.12.019

[31] Maceiras, R., Feijoo, J., Alfonsin, V., Perez-Rial, L. (2025). Effectiveness of active learning techniques in knowledge retention among engineering students. *Education for Chemical Engineers*, 51, 1-8. DOI:https://doi.org/10.1016/j.ece.2025.01.003

[32] Onwuegbuzie, A. J., Dickinson, W.B., Leech, N. L., Zoran, A.G. (2009). A qualitative framework for collecting and analyzing data in focus group research. *International journal of qualitative methods*, 8(3), 1-21. DOI:https://doi.org/10.1177/160940690900800301

[33] Plass, J., Homer, B., Kinzer, C. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, 50(4), 258-283. DOI: https://doi.org/10.1080/00461520.2015.1122533

[34] Powell, R.A., Single, H.M. (1996). Focus Groups. *International Journal for Quality in Health Care*, 8, 499-504. DOI: http://dx.doi.org/10.1093/intqhc/8.5.499

[35] Qian, M., Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63, 50-58. DOI: https://doi.org/10.1016/j.chb.2016.05.023

[36] Ramazani, J., Jergeas, G. (2015). Project managers and the journey from good to great: the benefits of investment in project management training and education. *International Journal of Project Management*, 33, 41–52. DOI: https://doi.org/10.1016/j.ijproman.2014.03.012

[37] Rezai, A., Ahmadi, R., Ashkani, P., Hosseini G., H. (2025). Implementing active learning approach to promote motivation, reduce anxiety, and shape positive attitudes: A case study of EFL learners. *Acta Psychologica*, 253. DOI: https://doi.org/10.1016/j.actpsy.2025.104704

[38] Roediger, H. L., Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, 15(1), 20-27. DOI: DOI: https://doi.org/10.1016/j.tics.2010.09.003

[39] Rossouw, M., Steenkamp, G. (2025). Developing the critical thinking skills of first year accounting students with an active learning intervention. *The International Journal of Management Education*, 23(1). DOI:https://doi.org/10.1016/j.ijme.2024.101086

[40] Ryan, R. M., Deci, E.L. (2000a). Intrinsic and extrinsic motivations: classic definitions and new directions, *Contemporary Educational Psychology*, 25(1), 54–67. DOI: https://doi.org/10.1006/ceps.1999.1020

[41] Ryan, R. M., Deci, E.L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. DOI: https://doi.org/10.1037/0003-066X.55.1.68

[42] Salas, E., Wildman, J., Piccolo, R.F. (2009). Using simulation-based training to enhance management education. *Academy of Management Learning and Education*, 8(4): 559-573. DOI:https://doi.org/10.5465/AMLE.2009.47785474

[43] Schrier, K., Zahradnik, A., Shaenfield, D. (2024). Playful Approaches to Leadership Development: Three Innovative Uses of Games in the Classroom. *Journal of Management Education*, 48(4), 777-801. DOI:https://doi.org/10.1177/10525629231215065

[44] Sung, H.-Y., Hwang, G.-J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43–51. DOI:https://doi.org/10.1016/j.compedu.2012.11.019

[45] Tham, L., Tham, R., (2012). Is Game-Based Learning an Effective Instructional Strategy to Engage Students in Higher Education in Singapore? A Pilot Study. *Journal of the Research Center for Educational Technology* (RCET), 8(1), 2-10. DOI: https://doi.org/10.20319/pijss.2015.s21.571581

[46] Thurston, T., N. (2018). Design Case: Implementing Gamification with ARCS to Engage Digital Natives. *Journal on Empowering Teaching Excellence*, 2(1), 5. DOI: https://doi.org/10.26077/vsk5-5613

[47] Wang, LH., *et al.* (2022). Effects of digital game-based STEM education on students' learning achievement: a meta-analysis. IJ STEM Ed, 9, 26. DOI: https://doi.org/10.1186/s40594-022-00344-0

[48] Wang, M., Zhu, J., Gu H., Zhang,J., Wu, D., Wang, P. (2025). Optimizing experiential learning in science education: The role of two-tier testing in digital game-based learning. *Entertainment Computing*, 54. DOI:https://doi.org/10.1016/j.entcom.2025.100960

[49] Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249–265. DOI: https://doi.org/10.1037/a0031311

[50] Zhao, D., Muntean, C., Chis, A., Rozinaj, G., Muntean, G.-M. (2022). Game-Based Learning: Enhancing Student Experience, Knowledge Gain, and Usability in Higher Education Programming Courses. IEEE Transactions on Education. DOI: https://doi.org/10.1109/TE.2021.3136914

Appendix A

Score	Meaning
1	Very low
2	Low
3	Lower than high
4	Higher than low
5	High
6	Very high

https://doi.org/10.14505/jres.v16.2(20).02

"What Makes a Good Class?" - Assessing University Students and Teachers' Perceptions

Ana MOURA
Centre for Research and Intervention in Education (CIIE) at University of Porto, Portugal, and
Instituto de Saúde Pública, University of Porto, Portugal, and
Laboratório para a Investigação Integrativa e Translacional em Saúde Populacional
ORCID: 0000-0003-4587-6797
ana.moura@ispup.up.pt

Carolina GOMES
Centre for Research and Intervention in Education (CIIE) at University of Porto, Portugal
ORCID: 0009-0001-1060-3855
up201904268@edu.fpce.up.pt

Teresa JACQUES
Faculty of Psychology and Education Sciences, University of Porto, Portugal, and
Centre for Psychology at University of Porto (CPUP), Portugal
ORCID: 0000-0001-9371-3779
tjacques@fpce.up.pt

Eunice MACEDO
Centre for Research and Intervention in Education (CIIE) at University of Porto, Portugal, and
Faculty of Psychology and Education Sciences, University of Porto, Portugal
ORCID: 0000-0003-1200-6621
eunice@fpce.up.pt

Mariana Veloso MARTINS
Faculty of Psychology and Education Sciences, University of Porto, Portugal, and
Centre for Psychology at University of Porto (CPUP), Portugal
ORCID: 0000-0001-6489-0290
Corresponding author: mmartins@fpce.up.pt

Article info: Received 15 September 2025; Revised 29 September 2025; Accepted 6 October 2025; Published 30 December 2025.

Abstract: This study explores the perceptions of university students, teachers, and researchers on what constitutes a 'good class' in higher education. Using qualitative data from 733 participants at a large Portuguese public university, the study applies Bronfenbrenner's ecological model to identify three key systems influencing these perceptions: microsystem, mesosystem, and exosystem. The microsystem highlights the importance of student-teacher interactions, teaching skills, and student engagement. The mesosystem emphasizes pedagogical methods, advocating for interactive and innovative teaching approaches, while the exosystem focuses on structural factors like class size and facilities. The findings reveal the critical role of these systems in shaping class quality, offering insights for improving higher education by aligning teaching practices with the expectations of students and educators. The study also calls for pedagogical innovation to address contemporary challenges and meet diverse student needs, providing valuable guidance for educators and policymakers.

Keywords: higher education; learning; student engagement; teaching method; pedagogical innovation.

JEL Classification: I21; I23; I25; I28; A22; C38.

Introduction

Current societal challenges have extended to Higher Education, requiring changes in how teaching is conceptualised and operationalised. Globalisation, changes in pedagogical theories and practices, and technological advancements (van Dijk *et al.* 2020), especially relevant in a post-pandemic period, pushed academia to a transition to andrological methods and adaptations to the digital generation, transforming how the classroom is conceptualised. Recent analyses of global higher education also highlight how structural and societal challenges continue to reshape teaching and learning practices (Tonkonog 2025).

Although the core mission of higher education remains the transmission of knowledge to prepare students for their careers and enhance human capital (Olo, Correia, and Rego 2021), several factors have shifted the focus in organizing the pedagogical offerings. These factors include changes in student profiles, an increasing number of university students, growing internationalisation, and the ambition to improve the quality of higher education (Berbegal-Mirabent, Mas-Machuca, and Marimon 2018). The importance of student and teacher involvement has been recognized even when integrating information and communication technologies (ICT) in a pedagogical manner that highlights its usefulness (Dumpit and Fernandez 2017). Given the importance of class quality in higher education and the need to engage academic communities in addressing these issues, it is essential to understand the factors that influence perceptions of a good class.

1. Framework

One way to conceptualise the interaction between teachers and students and how they bring and adapt to pedagogical innovation is using Bronfenbrenner's ecology of human development (1979, 1995) as a foundational rationale. The work of Bronfenbrenner on the analysis of individual human development emphasises the impact of interconnected environmental systems on human development over time, from immediate family and school interactions to broader cultural and societal influences (Bronfenbrenner 1979). Shortly, the microsystem includes the immediate environments where people interact directly, (*e.g.*, classrooms); the mesosystem refers to the interrelations among the microsystems (*e.g.*, family-school interactions); the exosystem encompasses broader cultural, economic, and societal influences that affect higher education, such as the policies and practices of the university administration or *curriculum* changes; and the macrosystem encompasses the several groups with shared values or beliefs in which the student is embedded, such as the culture and subculture of the academic institution. The influence of complex bioecological systems has been recognized for decades in how high-education students learn (*e.g.*, Bluteau, Clouder, and Cureton 2017) and engage (Skinner *et al.* 2022). Examining these systems also provides valuable insights into students' characteristics (Kitchen *et al.* 2019; Mulisa 2019).

Pedagogical innovation is a crucial theme in education (Law 2014). Traditional teaching methods based on transmission and memorization are being challenged and replaced by more innovative approaches (Carvalho *et al.* 2021). These new methods shift the role of the teacher towards that of a facilitator of learning. In this model, the student is central to the learning process and plays a crucial role in making decisions about their education (Carvalho *et al.* 2021; Freire 2014)

While implementing and widely accepting these new methods is necessary, it is also important to maintain their initial characteristics (Walder 2014). Pedagogical innovation contributes to creating and disseminating knowledge, helping education achieve new goals and processes and promoting significant changes in perspectives (Law 2014). This innovation has been shown to improve student motivation and deepen learning experiences (Carvalho *et al.* 2021). Authors have also noted the importance of integrating new technologies into the educational context alongside new teaching practices to avoid reverting to conservative pedagogical methods, even when they appear modern (Galeano-Salgado and Álvarez-Rivadulla 2025).

While much has been written about innovative pedagogical practices and the need to implement them in the classroom (*e.g.*, Santos, Figueiredo, and Vieira 2019; Suyo-Vega, Fernández-Bedoya, and Meneses-La-Riva 2024), there is still no consensus on what constitutes quality teaching in higher education (van Dijk *et al.* 2020). The quest for quality in higher education is multifaceted, and there is paradoxical evidence of its most important components. Rather than viewing education as the delivery of knowledge, contemporary approaches increasingly conceptualize learning as a situated, relational, and socially mediated process (Rogoff, 2003; Vygotsky, 1978). This shift emphasizes the importance of participation, dialogue, and the co-construction of meaning in learning environments. In this light, the notion of a "good class" cannot be reduced to predefined criteria or universal metrics of effectiveness. Instead, it must be understood in relation to the social practices, roles, and interactions that take place within specific institutional and cultural contexts. The perspectives of both students and teachers are similar regarding the value attributed to the teacher-student relationship (Dicker *et al.* 2019). However, blended learning has shown equivalent learning outcomes despite significantly reduced classroom time between

30 and 80% (Müller and Mildenberger 2021). Students appreciate classes that are less lecture-focused (Güvendir 2014) and more participatory (O'Connor 2013), but still highly value the ability of the teacher to explain content clearly (Reverter-Masia and Hernandez-Gonzalez 2021). More insight is needed into how university teachers should balance these demands to integrate new pedagogical and learning theories while keeping the perception of a good class.

Understanding what constitutes a good class in higher education is multifaceted, influenced by various factors ranging from individual perceptions to broader systemic structures within educational institutions. In this context, the perceptions of both students and teachers play a crucial role in shaping the discourse around educational quality. This paper seeks to delve into these perceptions, examining and assessing the discourses around education provision, expectations and quality – and how these elements interplay to contribute to the conceptions of a good higher education class.

2. Method

The targeted population comprised the academic community of the University of Porto. Around 2200 teaching faculty and 34000 students were invited by email to participate in the study. The final sample included 733 participants. Participants were from 13 different departments and faculties of a large Portuguese public university. Four participants were affiliated with other Portuguese universities. Table 1 presents the participants' type of affiliation to the University. The majority of the sample was comprised of first year bachelor students, while researchers were the smaller group in the sample. No other information regarding the participants was collected.

Table 1. Participants' type of affiliation to the university.

Affiliation	***n***
1st Year Bachelor Students	254
2nd/ 3rd Year Bachelor Students	196
Master's Students	133
PhD Students	39
Professors	85
Researchers	4
Other*	25

* Erasmus/mobility students, students of other courses, participants that indicated more than one category, and retired professors.

The procedures for this study were performed in line with the principles of the Declaration of Helsinki. The (blinded for peer review) Ethics Committee conducted a preliminary review of our study design. Following an initial assessment confirming that data collection did not involve identifiable respondent information, the Committee determined that the study was exempt from IRB review (November 10, 2023). The Pedagogical Council then sent an email invitation to all students and teaching staff. Participants voluntarily completed an anonymous online form, which included their affiliation type within the university and the open-ended question: *"What makes a good class?"* The questionnaire remained accessible for one week.

3. Research Methodology

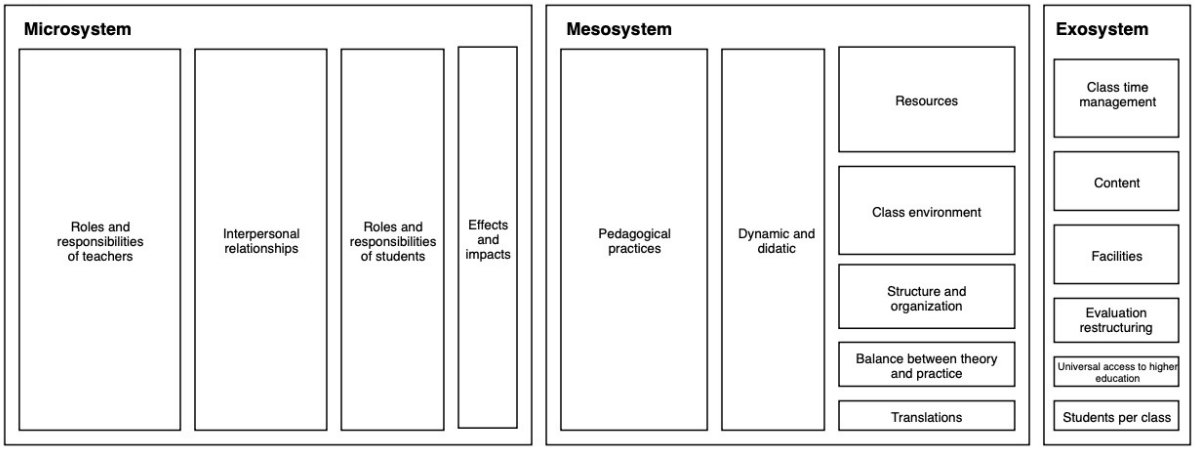
In this study, we adopted a relativist ontological position, recognizing that perceptions of a "good class" are diverse and constructed through the individual experiences of participants within their educational contexts. This ontological stance aligns with the belief that each participant's social and academic realities are dynamic and shaped by a multitude of interconnected factors (Smith, Sparkes, and Caddick 2014). As such, we approached this research with constructionist epistemology, seeking to interpret and understand the participants' experiences and perceptions from their perspectives, acknowledging the role of individual agency in shaping their understanding of classroom quality. The researchers acknowledge that these perceptions are dynamic and influenced by various interconnected factors, such as technological integration, pedagogical innovation, and the teacher-student relationship. By focusing on the participants' unique perspectives, the study aims to explore how they construct their understanding of classroom quality, using Bronfenbrenner's bioecological systems theory as a guiding framework. Reflexivity was maintained throughout to ensure an accurate and nuanced interpretation of the data.

Data were analyzed using QSR Nvivo 12 (Version 2012) to identify and comprehend patterns of meaning across the answers (Braun and Clarke 2012). Based on the research question 'What constitutes a good class for members of the academic community?' and after familiarizing with the data, a systematic coding process was carried out. Answers with similar meanings were inductively synthesised into categories and themes. After refining and defining these categories, three themes were created: 1) microsystem; 2) mesosystem; 3) exosystem. In addition to the theoretical saturation and information power model, triangulation strategies were applied to ensure the rigour and quality of the analysis. These included: i) researchers' triangulation; and ii) theory triangulation (Carter *et al.* 2014). AM and CG analysed the answers and identified the themes and categories, defining the coding framework. All other authors accompanied the process and collaborated to build the final coding tree. During the process of triangulation, the authors realized that there were vertices in common with Bronfenbrenner's ecological systems theory, therefore, the naming of the themes and data organization were aligned. All disagreements were discussed between the authors and resolved by consensus. Finally, the most illustrative *verbatim* quotes were selected and translated. (Bronfenbrenner, 1979)

4. Research Results

Results showed that three main systems influence a good class: microsystem, mesosystem, and exosystem (Figure 1).

Figure 1. Thematic map of students' and teachers' perceptions of 'what makes a good class'.



It was immediately apparent that the microsystem, which refers to the roles of teachers and students, their relationships, and the effects of classes, is the most influential system on perceptions of what constitutes a good class. Participants identified specific roles and responsibilities of teachers that significantly shape these perceptions. They generally emphasized the importance of a good teacher, which can be grouped into three sub-categories: teaching competencies, interpersonal skills, and professional attitudes. Regarding teaching competencies, participants highlighted communication skills, as well as the interest, motivation, and dedication of teachers, as central elements of a good class. Additionally, they mentioned the importance of encouraging student participation, adaptability, and the use of humor and creativity. Interpersonal skills were associated with the teacher's ability to be captivating, empathetic, and understanding, and to be available to clarify doubts. Concerning professional attitudes, participants pointed out the need for scientific rigor and proactivity.

Interpersonal relationships were frequently mentioned in participants' answers, with an emphasis on interaction. Participants believed that a good class depends on close interaction between teachers and students, which reduces power dynamics and increases students' confidence and motivation. One participant stated, 'A good class must have good interaction between students and teachers, but above all, it must not have positions of power that could reduce students' confidence in expressing their ideas and being able to learn' (A401, PhD student, Sport Science). Additionally, the values guiding interactions influence perceptions of a good class. Participants felt that classes should be framed by values such as respect, liberty, equity, and empathy. Effective communication between stakeholders and horizontal sharing logic were also identified as key elements of a good class.

Also, the roles and responsibilities of students for the existence of good classes were highlighted. Students' engagement, participation, motivation, attention, and interest seem to influence the existence of good

classes. Moreover, students' behavior in classes - specifically being quiet, using critical thinking, and respecting the teachers - seems to be a prerequisite for good classes.

The effects and impacts classes can have on students were identified as key elements of a good class, such as enhancing critical thinking and curiosity, being engaging, providing a sense of fulfillment, promoting the acquisition of new knowledge, and creating positive experiences.

Initially, the mesosystem was conceived as interactions between different microsystems where individuals are involved. However, adjustments were made to focus on interactions between the organizational systems of teachers (before and during classes) and students within this study. Participants emphasized that the quality of a class is highly dependent on the pedagogical practices employed. They highlighted interactive teaching methods such as debates, examples, application exercises, innovative approaches (*e.g.*, field trips, quizzes, games), linking content to practical applications and future professions, and addressing student doubts. Additionally, they emphasized the need for integrating challenges, using questioning techniques, fostering teamwork, providing feedback, and employing technologies to enhance class presentations. Furthermore, participants identified pedagogical practices that promote co-construction of knowledge and democratic learning processes, as well as effective class management. They stressed the importance of practices that deepen understanding and summarize learning content. Some participants also mentioned the use of the flipped classroom model.

Participants emphasized that a good class should be dynamic and didactic, fostering interest, stimulation, attention, and motivation. As one participant stated: 'Didactic and dynamic so that everyone understands the content and remains awake during the process.' (A382, Master Student, Engineering). Resources and the importance of audiovisual aids were also highlighted, such as well-structured, clear, and organized visual presentations that accompany lectures. Other mentioned resources were films, videos, illustrations, QR codes, and photographs. They underscored the value of providing materials in advance, as noted by a student: 'Good visual support (clear and appealing presentations, made available to students in advance).' (A655, Bachelor student, Medicine and Biomedical Sciences). Additionally, participants emphasized the importance of diverse and appealing resources, including written materials like books, to support learning.

Within the mesosystem, the classroom environment emerged as crucial for a good class. Participants emphasized the need for a safe, calm, and comfortable environment that fosters open communication and learning. Despite highlighting the importance of silence, participants also stressed the value of relaxed, cheerful, and creative class atmospheres. They emphasized the importance of structured and organized class sessions (preparation, delivery, and follow-up), clear objectives, content coherence, and providing support materials in advance. As one participant shared: 'In my opinion, a good class should start with a central theme followed by a brief explanation (from either teacher or students), contextualization, and discussion and reasoning exercises analyzing real-life situations/problems.' (A678, PhD student, Pharmacy).

Additionally, participants highlighted the importance of balancing theory and practice within the mesosystem of a good class. They believed that an excessive focus on theoretical content could diminish student autonomy, emphasizing the need for practical components to stimulate curiosity and active learning. As one participant noted: 'If the class is too focused on exposition, it reduces student autonomy. Mixing exposition with exploration of specific content arouses curiosity for learning.' (A116, Bachelor student, Psychology and Educational Sciences). Lastly, three mentions were made about the importance of translation for mobility students as an influencing factor of what a good class is

Finally, participants discussed the exosystem, which encompasses the influence of higher education structures, policies, and guidelines on perceptions of a good class. They emphasized the importance of effective class time management, including breaks to maintain concentration. *'A good class must have breaks because a young person cannot be 100% concentrated for more than 50 minutes'* (A300, Bachelor Student, Farmacy). Participants also highlighted the significance of structured, relevant, and scientifically rigorous content that advances academic knowledge. Infrequently mentioned factors included the importance of comfortable facilities with good acoustics, lighting, air conditioning, and seating conducive to interaction. As reinforced: *'A room with good space, comfort, climatic and technological conditions.'* (A597, Master student, Arts and Humanities). Participants also noted the restructuring of evaluations, the importance of universal access to higher education for all students (public, free, and quality higher education), and smaller class sizes to enhance learning experiences.

5. Discussions

The quality of higher education classes has been a subject of significant interest, especially following recent changes and pedagogical innovations (Núñez-Canal, de Obesso, and Pérez-Rivero 2022). The findings from this study provide a comprehensive view of what constitutes a 'good class' from the perspectives of both students and

teachers in higher education. Adapting Bronfenbrenner's ecological perspective facilitated a nuanced understanding of the multifaceted interactions between different ecological systems, illuminating how factors at different levels of influence intersect the understanding of what a good class in higher education means (B. Johnson 2008).

At the most immediate level, the surveyed academic community referred mainly to the microsystem, which encompasses the direct interactions between students and instructors, as well as the physical and social elements of the classroom itself. Despite the integration of new technologies in higher education, our results revealed that the role of the teacher remains central to student learning outcomes and engagement (Pantić 2017; Redecker 2017). This suggests that instructors who foster an engaging, supportive, and intellectually challenging microsystem can cultivate high levels of student engagement and deep learning (Shernoff, Tonks, and Anderson 2014). Therefore, it seems logical that the interaction occurring within that particular class group and with a particular teacher was the second most frequent category perceived as important for a good class experience. Learning encompasses a social aspect, which includes developing communication skills, empathy, and collaborative abilities, and is significantly enhanced through human interaction (Pianta, Hamre, and Allen 2012). The teacher-student dynamic not only drives academic achievement but also plays a pivotal role in the socio-emotional development of students, which is essential for their overall growth and future success. Recent findings corroborate that students' perceptions of respectful and supportive teacher behaviours strongly influence both engagement and well-being (Vicencio-Clarke *et al.* 2025). This positive interaction is, of course, interconnected both with the teacher's ability to engage students and with the students' engagement. The role and responsibility of the student emerged as an independent but interconnected category of the microsystem, fostered by attention, motivation and engagement. Motivation and engagement are pivotal for students' participation in class and the establishment of both student-student and teacher-student relationships (Xerri, Radford, and Shacklock 2018). Interestingly, participants pointed out that a good class must have a personal impact. Students, teachers and researchers felt that a good class has to create a sense of increasing knowledge, critical thinking, curiosity or engagement. These results align with the findings of a randomized controlled trial that demonstrated how positive first-day experiences can increase academic performance by the end of the term (Wilson and Wilson 2007).

Beyond the microsystem, the mesosystem encompasses the connections and relationships between the various settings and contexts students and teachers navigate. The mesosystem was the second theme present in most of our participants' answers, with pedagogical practices being the most frequent category. Many of the practices experienced positive involved collaborative learning (*e.g.*, working in groups; flipped classrooms). Besides increasing engagement, collaborative learning has proven beneficial in improving relationships among students and mental health (D.W. Johnson, Johnson, and Smith 2024). Furthermore, flipped classroom studies indicate that alignment between students' perceived and actual learning strategies is crucial for maximizing the benefits of these collaborative approaches (Han and Yang 2025). The incorporation of technology in the classroom was also highly mentioned. Nonetheless, this positive attribution to technology was contextualized in support of interactive and student-centered learning. In effect, when these approaches are present and the classroom is felt as dynamic, the use of technology has been shown improve learning outcomes and satisfaction (Means *et al.* 2013). However, our results show that the use of technologies and other teaching aids or resources has to be diverse. This leads to the interconnected category of balance between theory and practice. The balance between theory and practice is crucial, as excessive theoretical focus can reduce student autonomy and engagement, while practical components stimulate curiosity and active learning (Prince 2004). The use of diverse and dynamic practices while employing different pedagogical practices and balancing theory and practice requires a large dedication to structuring and organization, and participants acknowledged this in their reports. Lastly, the promotion of a comfortable and inclusive classroom environment was highlighted by the academic community. Inclusive pedagogy has been linked to better engagement and academic performance, particularly among traditionally underrepresented student groups (Felten and Lambert 2020), which seems to be the ones who suggested that a good class has the written materials translated.

Besides microsystem and mesosystem categories, broader influences and structural factors were also perceived as important to a good class. Effective time management, including regular breaks, was highlighted as vital for maintaining student concentration and engagement. The quality of classroom facilities, such as acoustics, lighting, and seating, was also mentioned as a contributing factor. Additionally, participants emphasized the importance of accessible and equitable education. Smaller class sizes and the restructuring of evaluations were suggested to enhance the learning experience. The optimal length of higher education classes and the ideal number of students per class have been subjects of considerable research over the past decade. For instance, Deslauriers *et al.* (2019) found that class periods of about 50 to 75 minutes are effective for maintaining student

engagement and cognitive performance, as shorter classes might not allow enough time for deep learning, while longer sessions risk diminishing returns in terms of attention and retention. Regarding class size, a systematic review concluded that smaller class sizes (typically under 30 students) enhance individual attention from instructors, increase student participation, and improve academic performance (Cuseo 2007).

Curiously, while we know that the overarching sociocultural, political, and economic forces that influence the educational system as a whole, Bronfenbrenner's macrosystem level of the model did not emerge as a theme from the analyzed data. This might be explained by the fact that participants chose to refer to the experiences they live most closely and tend not to prioritize societal attitudes towards higher education, government priorities and policies, and the state of the labor market can all play a role in shaping perceptions of what constitutes an optimal learning environment (Amali *et al.* 2023).

Taken together, the findings of our study reveal that the teachers and students still underscore the necessity for pedagogical innovation in higher education. Traditional teaching methods focused on transmission and memorization are increasingly being replaced by more dynamic and interactive approaches. These new methods position the teacher as a facilitator of learning, encouraging student participation and co-construction of knowledge. Innovative practices were identified as effective in promoting a deeper understanding of the material and fostering a more engaging learning environment. These practices align with the challenges and opportunities presented by the digital age and the need for higher education institutions to adapt to the evolving educational landscape.

Despite the valuable insights provided by this study, several limitations should be acknowledged. First, the study is limited by its sample size and scope, which almost exclusively includes participants from one large Portuguese public university. Although it can be assumed that the public higher education system is very similar in Portugal, this restricts the generalizability of the findings to other contexts and institutions. Second, the data collection method, relying on self-reported perceptions through an online survey, may introduce biases such as social desirability bias or response bias. Additionally, the cross-sectional nature of the study does not allow for an examination of changes in perceptions over time or the impact of specific interventions. Furthermore, while the study employs Bronfenbrenner's ecological systems theory to categorize responses, it may oversimplify the complex interactions between various factors influencing perceptions of a good class. Future research could benefit from a more nuanced analysis that considers additional variables and their interplay.

Still, this study has several strong points that enhance its contribution to the field of higher education research. The use of a well-established theoretical framework, Bronfenbrenner's ecological systems theory, provides a structured approach to analyzing and understanding the data. This theoretical grounding enables a comprehensive examination of the multiple layers of influence on perceptions of a good class. Another strength is the use of qualitative insights from open-ended responses. This approach allows for a richer, more detailed understanding of participants' views and experiences. The study also highlights the importance of both student and teacher perspectives, offering a balanced view of the factors contributing to effective teaching and learning. Finally, the emphasis on practical implications and pedagogical innovation provides actionable insights for educators and policymakers. The identification of specific teaching methods and classroom management strategies that enhance student engagement and learning can inform the development of more effective educational practices.

Conclusion

In conclusion, this study highlights the multifaceted nature of what makes a good class in higher education. The findings suggest that both micro-level classroom dynamics and macro-level institutional policies significantly influence students' and teachers' perceptions. The integration of practical components with theoretical knowledge, effective class management, and supportive learning environments is a key aspect of fostering a positive educational experience. Additionally, the push for pedagogical innovation reflects the need for higher education to evolve in response to contemporary challenges and to better meet the needs of a diverse student body. These insights provide valuable guidance for educators and policymakers aiming to enhance the quality of higher education. By understanding and addressing the factors that contribute to a good class, institutions can create more effective and engaging learning environments that support student success and academic achievement.

Acknowledgments:

The authors would like to thank all the members of the Pedagogical Council of the University of Porto Faculty of Psychology and Education Sciences for the discussion of ideas, as well as the participants for their valuable contributions.

Credit Authorship Contribution Statement:

Ana Moura: Conceptualization, Investigation, Methodology, Formal analysis, Writing – original draft, Data curation, Writing – review and editing;
Carolina Gomes: Conceptualization, Investigation, Methodology, Formal analysis, Data curation, Writing – review and editing;
Teresa Jacques: Conceptualization, Validation, Writing – review and editing;
Eunice Macedo: Conceptualization, Writing – original draft, Validation, Writing – review and editing,
Mariana V. Martins: Conceptualization, Project administration, Writing – original draft, Supervision, Validation, Writing – review and editing.

Declaration of Competing Interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Use of Generative AI and AI-Assisted Technologies:

The authors declare that they have used/not used generative AI and AI-assisted technologies during the preparation of this work.

References

[1] Amali, Nurul Amirah Khairul, Muhammad Usamah Mohd Ridzuan, Noor Hanim Rahmat, Hui Zanne Seng, and Norliza Che Mustafa. (2023). Exploring learning environment through Bronfenbrenner's ecological systems theory. *Journal of Academic Research in Progressive Education and Development,* 12 (2): 144-162. DOI: https://doi.org/10.6007/IJARPED/v12-i2/16516

[2] Berbegal-Mirabent, Jasmina, Marta, Mas-Machuca, and Frederic, Marimon. (2018). Is research mediating the relationship between teaching experience and student satisfaction? *Studies in Higher Education,* 43 (6): 973-988. DOI: https://doi.org/10.1080/03075079.2016.1212321

[3] Bluteau, Patricia, Lynn, Clouder, and Debra, Cureton. (2017). Developing interprofessional education online: An ecological systems theory analysis. *Journal of Interprofessional Care,* 31 (4): 420-428. DOI:https://doi.org/10.1080/13561820.2017.1301892

[4] Braun, Virginia, and Victoria, Clarke. (2012). *Thematic analysis*. American Psychological Association.

[5] Bronfenbrenner, Urie. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

[6] Bronfenbrenner, Urie. (1995). The bioecological model from a life course perspective: Reflections of a participant observer.

[7] Carter, N., D. Bryant-Lukosius, A. DiCenso, J. Blythe, and A. J. Neville. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum,* 41 (5): 545-547. DOI: https://doi.org/10.1188/14.Onf.545-547

[8] Carvalho, Andreia, Sergio Jesus, Teixeira, Leonilde, Olim, Sancha, de Campanella, and Teresa, Costa. (2021). Pedagogical innovation in higher education and active learning methodologies–a case study. *Education + Training,* 63 (2): 195-213. DOI: https://doi.org/10.1108/ET-11-2020-0368

[9] Cuseo, Joe. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *The Journal of Faculty Development,* 21 (1): 5-21

[10] Deslauriers, Louis, Logan S., McCarty, Kelly, Miller, Kristina, Callaghan, and Greg, Kestin. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences,* 116 (39): 19251-19257. DOI:https://doi.org/10.1073/pnas.1821936116

[11] Dicker, Ronald, Michael Garcia, Alison Kelly, and Hilda Mulrooney. (2019). What does 'quality' in higher education mean? Perceptions of staff, students and employers. *Studies in Higher Education,* 44 (8): 1425-1441. DOI: https://doi.org/10.1080/03075079.2018.1445987

[12] Dumpit, Duvince Zhalimar, and Cheryl Joy Fernandez. (2017). Analysis of the use of social media in Higher Education Institutions (HEIs) using the Technology Acceptance Model. *International Journal of Educational Technology in Higher Education,* 14: 1-16. DOI: https://doi.org/10.1186/s41239-017-0045-2

[13] Felten, Peter, and Leo M Lambert. (2020). *Relationship-rich education: How human connections drive success in college*. Johns Hopkins University Press.

[14] Freire, Paulo. (2014). *Pedagogia da autonomia: saberes necessários à prática educativa*. Editora Paz e Terra.

[15] Galeano-Salgado, Andres Mauricio, and María José Álvarez-Rivadulla. (2025). Cross-class interactions and subjective inequality: perceptions, beliefs and distributive preferences at a Colombian elite university. *Frontiers in Sociology,* 10: 1619937. DOI: https://doi.org/10.3389/fsoc.2025.1619937

[16] Güvendir, Meltem. (2014). A scaling research on faculty characteristics that higher education students prioritize. *College Student Journal,* 48 (1): 173-183

[17] Han, Feifei, and Jingpeiyi Yang. (2025). Consistency between self-reported and log data to understand students' experience of learning in flipped classrooms. *Scientific Reports,* 15 (1): 34369. DOI:https://doi.org/10.1038/s41598-025-34369

[18] Johnson, Bruce. (2008). Teacher–student relationships which promote resilience at school: A micro-level analysis of students' views. *British Journal of Guidance & Counselling,* 36 (4): 385-398. DOI:https://doi.org/10.1080/03069880802364528

[19] Johnson, David W., Roger T., Johnson, and Karl A., Smith. (2024). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching,* 25 (3&4): 85-118

[20] Kitchen, Joseph A., Ronald E., Hallett, Rosemary J., Perez, and Gwendelyn J., Rivera. (2019). Advancing the use of ecological systems theory in college student research: The ecological systems interview tool. *Journal of College Student Development,* 60 (4): 381-400. DOI: https://doi.org/10.1353/csd.2019.0035

[21] Law, Nancy. (2014). Comparing pedagogical innovations. In *Comparative education research: Approaches and methods*, 333-364. Springer.

[22] Means, Barbara, Yukie, Toyama, Robert, Murphy, and Marianne, Baki. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record,* 115 (3): 1-47

[23] Mulisa, Feyisa. (2019). Application of bioecological systems theory to higher education: Best evidence review. *Journal of Pedagogical Sociology and Psychology,* 1 (2): 104-115

[24] Müller, Claude, and Thoralf, Mildenberger. (2021). Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. *Educational Research Review,* 34: 100394. DOI: https://doi.org/10.1016/j.edurev.2021.100394

[25] Núñez-Canal, Margarita, Mª de las Mercedes de Obesso, and Carlos Alberto, Pérez-Rivero. (2022). New challenges in higher education: A study of the digital competence of educators in Covid times. *Technological Forecasting and Social Change,* 174: 121270. DOI: https://doi.org/10.1016/j.techfore.2021.121270

[26] O'Connor, Kevin. (2013). Class participation: Promoting in-class student engagement. *Education,* 133 (3): 340-344.

[27] Olo, Daniela, Leonida, Correia, and Conceição, Rego. (2021). Higher education institutions and development: Missions, models, and challenges. *Journal of Social Studies Education Research,* 12 (2): 1-25.

[28] Pantić, Nataša. (2017). An exploratory study of teacher agency for social justice. *Teaching and Teacher Education,* 66: 219-230.

[29] Pianta, Robert C., Bridget K., Hamre, and Joseph P., Allen. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In *Handbook of research on student engagement*, 365-386. Springer.

[30] Prince, Michael. (2004). Does active learning work? A review of the research. *Journal of Engineering Education,* 93 (3): 223-231.

[31] Redecker, Christine. 2017. *European framework for the digital competence of educators: DigCompEdu*. Joint Research Centre (Seville site).

[32] Reverter-Masia, Joaquin, and Vicenço, Hernandez-Gonzalez. (2021). Basic abilities a good university teacher should have: case study at the University of Lleida, Spain. *Revista Espacios,* 42 (3)

[33] Santos, Julia, Amelia Simoes, Figueiredo, and Margarida, Vieira. (2019). Innovative pedagogical practices in higher education: An integrative literature review. *Nurse Education Today,* 72: 12-17

[34] Shernoff, David J., Stephen M., Tonks, and Brett, Anderson. (2014). The impact of the learning environment on student engagement in high school classrooms. *Teachers College Record,* 116 (13): 166-177.

[35] Skinner, Ellen A, Nicolette P., Rickert, Justin W., Vollet, and Thomas A., Kindermann. (2022). The complex social ecology of academic development: A bioecological framework and illustration examining the collective effects of parents, teachers, and peers on student engagement. *Educational Psychologist,* 57 (2): 87-113.

[36] Smith, Brett, Andrew C., Sparkes, and Nick, Caddick. (2014). Judging qualitative research. In *Research methods in sports coaching*, 192-201. Routledge.

[37] Suyo-Vega, Josefina Amanda, Víctor Hugo, Fernández-Bedoya, and Monica Elisa, Meneses-La-Riva. (2024). Beyond traditional teaching: a systematic review of innovative pedagogical practices in higher education. *F1000Research,* 13: 22.

[38] Tonkonog, VV. (2025). The analysis of modern problems and challenges of Russian medical education: the sociological survey results. *Problems of Social Hygiene, Public Health and History of Medicine,* 33 (4): 706-713.

[39] van Dijk, Esther E., Jan van Tartwijk, Marieke F. van der Schaaf, and Manon Kluijtmans. (2020). What makes an expert university teacher? A systematic review and synthesis of frameworks for teacher expertise in higher education. *Educational Research Review* 31: 100365. DOI: https://doi.org/10.1016/j.edurev.2020.100365

[40] NVivo Qualitative Data Analysis Software.

[41] Walder, Anne Mai. (2014). The concept of pedagogical innovation in higher education. *Education Journal,* 3 (3): 195-202.

[42] Wilson, Janie H, and Shauna B Wilson. (2007). The first day of class affects student motivation: An experimental study. *Teaching of Psychology,* 34 (4): 226-230.

[43] Xerri, Matthew J., Katrina Radford, and Kate Shacklock. (2018). Student engagement in academic activities: a social support perspective. *Higher Education,* 75: 589-605.

https://doi.org/10.14505/jres.v16.2(20).03

Integrating Economic Philosophy into Sierra Leone's Educational System: A Comparative Strategy for Sustainable Socio-Economic Development

Emerson Abraham JACKSON
Bank of Sierra Leone
Freetown, Sierra Leone
ORCID: 0000-0002-2802-6152
ejackson1@bsl.gov.sl

Article info: Received 17 July 2025; Received in revised form 26 July 2025; Accepted 10 August 2025; Published 30 December 2025.

Abstract: This study advocates for embedding economic philosophy into Sierra Leone's educational framework, uniting technical economic expertise with critical thinking and ethical reflection to address complex socio-economic challenges. Adopting a conceptual analysis approach, it examines the discipline's relevance in the national development context and proposes a structured framework for integration, including educator training, curriculum design, and evaluation mechanisms. While recognising barriers such as resource constraints and institutional resistance, it recommends phased implementation with active stakeholder engagement. The anticipated outcome is a generation of economists proficient in theory, skilled in ethical evaluation, and capable of fostering inclusive, resilient, and sustainable development.

Keywords: economic philosophy; educational reform; socio-economic development; Sierra Leone; knowledge advancement.

JEL Classification: A22; B40; I25; O15.

Introduction

Economic philosophy extends beyond basic economic modelling or empirical policy evaluations; it rigorously analyses the ethical, social, and intellectual underpinnings of economic institutions (Jackson, Tamuke & Jabbie, 2019; Hausman, 2008). It promotes contemplation of the values, assumptions, and normative assertions that influence economic theory and practice, transcending the perception of economics as solely a technical pursuit. It provides a comprehensive and analytical framework to evaluate both the intentional and unforeseen effects of economic activities on society.

The marginalization of philosophical discourse in economics education in Sierra Leone has impeded the cultivation of economists who are both critical and culturally informed (Jackson, 2023a). Notwithstanding attempts to overhaul the education system, the emphasis has frequently been on rote memorisation and technical skills, rather than cultivating critical thinking, ethical reasoning, and a comprehensive awareness of the socio-economic factors influencing the nation's development (World Bank, 2020). This limited emphasis on technical skills restricts students' capacity to tackle the intricate challenges of poverty, inequality, governance, and globalisation that are widespread in Sierra Leone.

This study advocates for the formal incorporation of economic philosophy into the economics curriculum at both secondary and post-secondary levels in Sierra Leone. This is not simply an educational improvement but a strategic imperative for national advancement. Incorporating philosophical inquiry into economic education enables students to examine the assumptions of economic theories, engage with varied theoretical viewpoints, and contemplate the ethical implications of economic decision-making. This critical approach is vital for cultivating economists capable of significantly contributing to Sierra Leone's sustainable and inclusive development (Jackson, 2016).

The incorporation of economic philosophy into the curriculum will enhance knowledge acquisition and sharpen students' analytical skills, fostering a more economically robust and ethically informed society. This transition anticipates a future in which economic decisions are informed by reflective judgement and ethical considerations, with policies designed to emphasise justice, human welfare, and societal well-being. In the long

term, integrating economic philosophy into the educational system will cultivate leaders capable of guiding Sierra Leone towards sustained and equitable economic advancement.

This study aims to demonstrate the theoretical and practical importance of economic philosophy as a core element of Sierra Leone's educational system. It seeks to reconceptualise economic philosophy as an all-encompassing, critically engaged discipline, rather than only a technical field vital for national development. The objective is to recommend a redesign of the current economics curriculum to improve its reflective, analytical, and contextual relevance, prompting policymakers and educational stakeholders to consider these changes. The specific aims set out for this study are: (i) To examine the impact of economic philosophy on the evolution of critical economic thought and reasoning. (ii) To provide a structured framework for incorporating economic philosophy into Sierra Leone's educational curriculum, and (iii) To highlight the potential contributions of philosophically informed economic education to the socio-economic transformation of Sierra Leone.

The novelty of this research lies in its pioneering effort to position economic philosophy not merely as an academic curiosity but as a transformative driver of socio-economic progress in Sierra Leone (Jackson, 2024). While previous studies have explored curriculum reform in economics, few have directly addressed the integration of philosophical inquiry as a strategic tool for national development. By bridging the gap between technical economic instruction and reflective, values-based reasoning, this study offers an original framework for producing graduates who are both analytically rigorous and ethically grounded. It challenges conventional educational paradigms by proposing that the moral and cultural dimensions of economic decision-making are not peripheral but central to fostering inclusive growth, institutional integrity, and resilience in the face of complex global and local challenges. In doing so, it situates Sierra Leone as a potential exemplar for other developing nations seeking to harmonise economic competence with social responsibility in their educational systems.

1. Research Background

1.1. To Cultivate Critical Thinking and Ethical Reflection

Economic philosophy is essential for fostering intellectual curiosity, ethical scrutiny, and autonomous reasoning among students. The aim of promoting critical thinking extends beyond the simple use of economic models and tools; it necessitates that students thoroughly examine the foundational assumptions, values, and ethical implications that influence economic theory and practice. Through philosophical inquiry, students are prompted to scrutinise not only the technical dimensions of economic theories but also their social, political, and ethical ramifications. This ethical reflection method enables students to critically assess the effects of diverse economic policies, allowing them to recognise potential injustices or unexpected consequences stemming from these decisions. Contemporary literature increasingly acknowledges the significance of critical thinking in economics education, with scholars contending that it is vital for cultivating economists capable of addressing intricate real-world issues, where ethical considerations frequently intersect with economic analysis (Sen, 1999; Moyo, 2018). The objective of economic philosophy is to provide students with the intellectual instruments required to question assumptions, assess alternative theories, and contemplate the ethical aspects of economic systems and policies. This process is especially pertinent in Sierra Leone, where socio-economic difficulties necessitate not only technical answers but also a contemplative approach that incorporates equity, justice, and long-term sustainability in decision-making.

1.2. To Harmonise Economic Theory with Socio-Historical Realities

The secondary theoretical objective of incorporating economic philosophy into Sierra Leone's educational framework is to harmonise economic theory with the nation’s distinct socio-historical setting. Economic theory, albeit vital, must be situated within the distinct historical and social contexts of Sierra Leone, a nation characterised by its colonial heritage, post-conflict rehabilitation, and governance intricacies (Kandeh, 2008; Acemoglu, Johnson, & Robinson, 2001). The colonial history of Sierra Leone, which moulded its institutional frameworks and economic systems, persists in affecting the current economic operations. The legacies of extractive institutions and resource dependency have resulted in considerable inequalities in wealth distribution and economic power. By anchoring economic theory in these socio-historical contexts, students can more effectively comprehend the developmental issues confronting Sierra Leone and critically analyse global economic theories that may inadequately represent the local milieu. Furthermore, economic philosophy requires students to assess how policies influenced by global economic paradigms frequently neglect the historical and cultural intricacies that characterise local economies. Students will learn traditional economic models while also developing the ability to critically evaluate their application and limitations within Sierra Leone's post-colonial and post-conflict setting. The incorporation of economic philosophy into education is not alone aimed at expanding

students' comprehension of global economic theories but also at motivating them to tailor these theories to effectively confront local reality.

1.3 To Develop Economists and Policymakers Grounded on Ethical and Philosophical Principles

The primary objective of economic philosophy within the educational framework is to cultivate economists and policymakers who possess both technical proficiency and ethical awareness. In modern global and local economies, economic decisions encompass not only the maximisation of efficiency or profit but also moral considerations like fairness, equity, and justice. This is especially pertinent in a country such as Sierra Leone, where economic policies directly impact the lives of the most vulnerable populations. Integrating philosophical inquiry into economic education enables students to critically examine the ethical implications of their decisions, addressing concerns such as wealth distribution, individual rights versus collective welfare, and the long-term sustainability of economic practices. Inspired by the philosophies of Nussbaum (2011) and Sen (1999), which advocate for human-centered development, this purpose aims to cultivate a cohort of economists dedicated to promoting social justice, human dignity, and sustainable development. Graduates possessing this ethical framework will be capable of evaluating the economic ramifications of policies while ensuring that these policies foster inclusive growth, diminish inequality, and improve the welfare of all citizens. In Sierra Leone, this entails cultivating a cadre of leaders capable of addressing the nation's socio-economic difficulties with an emphasis on ethical considerations, equity, and sustainable well-being rather than immediate profits or politically convenient answers.

1.4. To Foster an Interdisciplinary Perspective

Economic philosophy promotes an integrative methodology for comprehending and tackling intricate development issues. Economic philosophy integrates economics with political science, sociology, ethics, and history, offering students a comprehensive perspective on the world, acknowledging that economic challenges are seldom isolated and are profoundly intertwined with political, social, and cultural phenomena. In Sierra Leone, an interdisciplinary approach is crucial for addressing the complex concerns that influence the nation's growth, including political instability, environmental degradation, poverty, and governance challenges. Stiglitz (2000) contends that economic policies cannot be properly formulated or executed without comprehending the wider social and political framework in which they function. Economic philosophy fosters students' ability to integrate insights from diverse disciplines, thereby enhancing their comprehension of the economic challenges confronting Sierra Leone and enabling them to propose solutions that encompass economic, political, social, and ethical dimensions. An interdisciplinary approach to the mining sector in Sierra Leone would encompass not only the economic valuation of minerals but also the environmental and social ramifications, including land rights, local community development, and the ethical considerations of resource exploitation. Economic philosophy equips students to synthesise knowledge across several fields, preparing them to tackle complex development concerns comprehensively and integratively. This interdisciplinary approach is essential for creating solutions that are technically robust, socially responsible, politically feasible, and ethically just. The goals emphasise the necessity of incorporating economic philosophy into Sierra Leone's educational framework, guaranteeing that future economists and policymakers possess not only technical expertise but also social responsibility, ethical awareness, and the ability to tackle the nation's developmental challenges from a holistic, interdisciplinary viewpoint. This integration will enable students to think critically, reflect ethically, and engage with economic concerns in a culturally relevant and internationally educated manner.

2. Research Methodology

This study employs a conceptual analysis methodology to examine the incorporation of economic philosophy into Sierra Leone's educational system. The project seeks to illustrate the potential benefits of integrating philosophical inquiry into economics education, while also confronting practical problems including resource constraints and institutional opposition. The research presents a systematic methodology grounded in the following strategies:

- Development of a Conceptual Framework: Economic philosophy serves as a vital instrument for fostering critical thinking, ethical reasoning, and a sophisticated comprehension of socio-economic reality. Prominent texts in economic philosophy, including those by Jackson (2019a & 2019b), Hausman (2008), Marx (1867), Sen (2009), and Nussbaum (2011), underpin this approach, promoting the integration of philosophical inquiry with economics to rigorously evaluate moral norms and assumptions.

- Curriculum Development and Evaluation: The study proposes a curriculum framework that includes modules such as "Ethics and Economics," "Foundations of Economic Thought," and "Economic Justice and Society". These modules will facilitate students' critical engagement with both classical and contemporary economic theories and their ethical ramifications. The curriculum design is informed by Freire's (1970) critical pedagogy, fostering dialogue and critical reflection on societal norms.
- Educator Training and Capacity Development: The study identifies the deficiency of trained educators and recommends specialised teacher training programs to provide educators with philosophical and pedagogical knowledge. This project adheres to Vygotsky's (1978) constructivist learning theory, which underscores active, social, and experiential learning.
- Strategic Implementation Framework: A slow, incremental strategy for integrating economic philosophy is proposed, commencing with pilot programs and progressive curriculum modifications. This method entails engagement with essential players, such as the Ministry of Education and professional associations, to establish consensus on the significance of economic philosophy in tackling Sierra Leone's socio-economic issues.

Engagement with Stakeholders and Institutional Assessment: Engagement with stakeholders is essential for the contextual incorporation of economic philosophy into the curriculum. The document delineates solutions to address difficulties such as curriculum congestion, opposition, and resource constraints, including the utilisation of current courses and international collaborations for resource assistance.

3. Discussions

The incorporation of economic philosophy into Sierra Leone's educational system offers a distinctive chance to tackle the nation's socio-economic issues by fostering a new generation of economists who are critically engaged and morally informed. Recent studies underscore the increasing acknowledgement that economic education should extend beyond technical proficiency to encompass a wider ethical and socio-political consciousness (Jackson, 2023). In Sierra Leone, where structural disparities endure and the economy is heavily reliant on external assistance and natural resources, this integration is especially vital. Critical thinking, as proposed by Sen (1999), is essential for students to question the existing quo and critically examine economic theories about their ethical foundations. This method not only improves technical skills but also prepares students to assess the ethical implications of economic policies and practices.

One of the primary purposes of this study is to align economic theory with Sierra Leone's socio-historical environment, profoundly influenced by its colonial legacy and post-conflict circumstances. Acemoglu, Johnson, and Robinson (2001) contend that the legacy of colonialism persists in shaping the economic frameworks of numerous African countries, including Sierra Leone. Economic philosophy offers a critical perspective for analysing these legacies and suggesting alternative frameworks that better correspond with the realities of developing nations. Recent research on post-colonial economies, including Moyo (2018), emphasises the necessity for educational reforms that enable students to interact with local historical and cultural contexts, thereby cultivating an understanding of how global economic systems have historically disadvantaged African nations. Incorporating economic philosophy into the curriculum enables students to examine these subjects comprehensively, fostering a more sophisticated knowledge of economic development and its ethical ramifications.

Moreover, incorporating philosophical inquiry into the economics curriculum immediately fulfills the necessity for economists and policymakers who possess both technical proficiency and ethical awareness. This is particularly relevant in Sierra Leone, where corruption, poverty, and inequality persistently obstruct growth (World Bank, 2020). Nussbaum (2011) supports for the integration of capacities into economic education, emphasising the enhancement of individual freedoms and possibilities for all citizens, rather than solely prioritising GDP growth. Integrating these ethical concepts into economic philosophy education will empower future economists to transcend the confines of conventional economic models and champion policies that emphasise social fairness and human dignity. This transition is essential in a nation such as Sierra Leone, where economic advancement must be more equitable to tackle the urgent challenges of inequality and underdevelopment.

Aligned with the study's purpose to cultivate an interdisciplinary perspective, economic philosophy urges students to investigate the intersections among economics, politics, sociology, and ethics. The current difficulties confronting Sierra Leone, including corruption, resource mismanagement, and political instability, necessitate a multidisciplinary approach for effective resolution (Jackson, 2021). The mining sector in Sierra Leone, a crucial component of the economy, currently confronts issues that are not merely economic but also political and ethical in nature. Incorporating economic philosophy into the curriculum would provide students with the means to

critically evaluate the moral and political aspects of these concerns, enabling them to suggest more comprehensive solutions. This interdisciplinary approach is crucial for equipping students to manage the intricacies of globalisation and its effects on local economies, especially in post-conflict contexts.

Furthermore, integrating economic philosophy via critical pedagogy and constructivist learning theory could transform the educational experience in Sierra Leone. Freire's (1970) critical pedagogy advocates for an active, participatory learning approach that invites students to interrogate society institutions and engage in discourse regarding possible alternatives. In Sierra Leone, where the education system has historically prioritised rote learning, the implementation of critical pedagogy would enable students to become more reflective and engaged participants in their education. This method has demonstrated efficacy in improving students' problem-solving skills and equipping them for real-world issues (Zhao, 2022). Utilising constructivist learning theory, which emphasises the significance of students' social surroundings in the learning process (Vygotsky, 1978), economic philosophy can be contextualised to enhance its relevance to the daily experiences of students in Sierra Leone. This method would promote enhanced engagement with the content and motivate students to apply economic theory to local concerns, like income inequality, the informal sector, and rural development.

Conclusions and Further Research

In conclusion, the integration of economic philosophy into Sierra Leone's educational framework presents a transformative pathway for cultivating graduates who are not only technically proficient but also ethically anchored and contextually aware. This study has demonstrated that embedding philosophical inquiry within the economics curriculum strengthens critical thinking, ethical reasoning, and an appreciation of the socio-historical realities that shape the nation's development trajectory. Such integration is pivotal in equipping future economists and policymakers to navigate the multifaceted challenges of a globalised economy while remaining attuned to the ethical imperatives of fairness, justice, and inclusiveness (Jackson, 2023).

By situating economic theory within Sierra Leone's historical and cultural context, the proposed approach empowers students to interrogate the moral and social implications of economic decision-making (Jackson, 2022). This is particularly vital in a society still grappling with the enduring legacies of colonialism, post-conflict recovery, and structural inequalities. Recent scholarship underscores that economic education must move beyond technical efficiency to address deeper normative concerns that influence policy legitimacy and societal well-being (Mensah & Boateng, 2024).

Moreover, the interdisciplinary emphasis of this research ensures that economic challenges are approached holistically incorporating insights from sociology, politics, ethics, and environmental studies. This multidimensional lens is essential for tackling pressing national concerns, from political instability to environmental degradation, and for fostering innovative, locally appropriate solutions (Conteh & Kamara, 2025). In this regard, economic philosophy becomes both an intellectual and practical tool for shaping sustainable and inclusive development pathways.

The adoption of critical pedagogy and constructivist learning theory further enhances the potential of economic philosophy to reframe the educational experience. By promoting active, participatory learning and contextual relevance, these approaches can dismantle the entrenched culture of rote memorisation and replace it with reflective, problem-solving mindsets. Such pedagogical reform is aligned with emerging educational models across Africa that link cognitive skills with civic responsibility (Okafor *et al.* 2024).

The novelty of this research lies in its unprecedented framing of economic philosophy not as an ancillary academic subject, but as a strategic national development tool. While prior reforms in economics education have prioritised technical competencies, few have positioned philosophical inquiry as the central bridge between technical analysis and ethical governance. This study uniquely proposes a structured, context-sensitive model that harmonises global economic theory with local socio-cultural realities, thereby producing graduates capable of advancing both economic efficiency and moral legitimacy in policy-making. In doing so, it offers Sierra Leone an opportunity to become a regional exemplar for integrating ethics and critical thinking into economic education - an approach increasingly recognised as indispensable for resilience in the face of complex global and domestic challenges (United Nations Economic Commission for Africa, 2023).

Ultimately, this intellectual and pedagogical transformation is not merely an academic aspiration but a strategic imperative. As Sierra Leone aspires to middle-income status, aligning its economic education with ethical, interdisciplinary, and context-specific principles will ensure that future decision-makers are equipped to craft policies that are technically sound, socially just, and sustainably inclusive.

Credit Authorship Contribution Statement

Emerson Abraham Jackson: Responsible for all the following: Conceptualization, Investigation, Methodology, Project administration, Software, Formal analysis, Writing – original draft, Supervision, Data curation, Validation, Writing – review and editing, Visualization, Funding acquisition.

Declaration of Competing Interest

There are no competing interests.

Declaration of Use of Generative AI and AI-Assisted Technologies

Some ideas were generated from shaping the direction of the work from DeepSeek.

References

[1] Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Analysis. *American Economic Review*, Volume 91, 5, 1369–1401. Available at: https://economics.mit.edu/sites/default/files/publications/colonial-origins-of-comparative-development.pdf

[2] Conteh, S., & Kamara, M. (2025). Interdisciplinary approaches to sustainable development in West Africa: Policy and education linkages. *African Journal of Development Studies*, 15(1), 45–63. DOI:https://doi.org/10.1080/ajds.2025.154

[3] Freire, P. (1970). Pedagogy of the Oppressed. Continuum. Available at: https://envs.ucsc.edu/internships/internship-readings/freire-pedagogy-of-the-oppressed.pdf

[4] Hausman, D. M. (2008). *Philosophy of Economics*. Stanford Encyclopaedia of Philosophy. Available at: https://plato.stanford.edu/entries/economics/

[5] Jackson, E.A. (2024). Transforming Universities for 21st Century Africa: A Case Study of Sierra Leone. *Journal of Research in Educational Sciences*, 15(2(18)), 77-82. DOI:https://doi.org/10.14505/jres.v15.2(18).01

[6] Jackson, E.A. (2023). Rethinking Epistemology: The Role of Narrative in Economics as a Social Science. Theoretical and Practical Research in Economic Fields, 14(1), 164-174. DOI:https://doi.org/10.14505/tpref.v14.1(27).13

[7] Jackson, E.A. (2023a). Philosophical Discourses on Economic Governance: An African Perspectives. *Theoretical and Practical Research in Economic Fields (TPREF*), 14(2(28)), 207-214. DOI:https://doi.org/10.14505/tpref.v14.2(28).01

[8] Jackson, E.A. (2022). Grappling with Social Sciences Thinking of (Ir)rationality Paradox during Times of Uncertainty. *Economic Insights- Trends and Challenges*, 14(3), 79-86. Available at: https://upg-bulletin-se.ro/wp-content/uploads/2023/01/8.Jackson.pdf

[9] Jackson, E.A., Tamuke, E., & Jabbie, M. (2019). Disaggregated Short-Term Inflation Forecast (STIF) for Monetary Policy Decisions in Sierra Leone. *Financial Markets, Institutions, and Risks*, 4(4), 32-48. DOI:https://doi.org/10.21272/fmir.3(4).32-48.2019

[10] Jackson, E.A. (2019a). Theoretical and Methodological Context of (Post)-Modern Econometrics and Competing Philosophical Discourses for Policy Prescription. *Journal of Heterodox Economics*, 4(2), 119-129. DOI: https://doi.org/10.1515/JHEEC-2017-0006

[11] Jackson, E.A. (2019b). On the question of the relevance of economics as a science: Postmodern Filosofia critique. *Theoretical and Practical Research in the Economic Fields*, 9(2(18)), 131-138. DOI:https://doi.org/10.14505/tpref.v9.2(18).01

[12] Jackson, E. A. (2016). Phronesis and Hermeneutics: The Construct of Social / Economic Phenomenon and their Interpretation for a Sustainable Society. *Economic Insights - Trends and Challenges*, 8 (2), 1-8. Available at: https://ssrn.com/abstract=2808452

[13] Kandeh, J. D. (2008). Unruly incumbents, donor support, and Sierra Leone's second post-conflict elections in 2007. *Journal of Contemporary African Studies*. 46(4), 603-635. Available at: https://www.africabib.org/htp.php?RID=321429729

[14] Marx, Karl. (1867). *Das Kapital: A Critique of Political Economy*. Publishing house of Otto Meissner. Available at: https://oll.libertyfund.org/titles/das-kapital-kritik-der-politischen-oekonomie-buch-1-1867

[15] Mensah, P., & Boateng, K. (2024). Reimagining economics curricula for inclusive growth: Lessons from Sub-Saharan Africa. *Journal of Economic Education and Policy*, 12(2), 101–119. DOI:https://doi.org/10.1080/jeep.2024.007

[16] Moyo, D. (2018). *Dead Aid: The Ineffectiveness of Aid and the Superior Alternatives for Afric*a. Farrar, Straus & Giroux.

[17] Nussbaum, M. (2011). Developing Capabilities: The Human Development Paradigm. Harvard University Press. Available at: https://ndpr.nd.edu/reviews/creating-capabilities-the-human-development-approach-2/

[18] Okafor, C., Adeniran, A., & Sillah, T. (2024). Critical pedagogy and socio-economic transformation in African higher education. *Comparative Education Review*, 68(3), 327–346. DOI: https://doi.org/10.1086/cer.2024.145

[19] Sen, A. (2009). *The Concept of Justice*. Harvard University Press. Available at: https://shs.hal.science/halshs-00496277v1/document

[20] Sen, A. (1999). Advancement as Liberation. Oxford University Press. Available at: http://www.c3l.uni-oldenburg.de/cde/OMDE625/Sen/Sen-intro.pdf

[21] Stiglitz, Joseph E. (2000). The Economics of the Public Sector. W.W. Norton & Company. Available at: https://books.google.com.sl/books/about/Economics_of_the_Public_Sector.html?id=dWe1QgAACAAJ&redir_esc=y

[22] United Nations Economic Commission for Africa. (2023). Rethinking education for Africa's sustainable development. UNECA Policy Brief No. 45. Available at: https://hdl.handle.net/10855/49926

[23] Vygotsky, Lev Semenovich (1978). Mind in Society: The Evolution of Advanced Psychological Processes. Harvard University Press. https://www.jstor.org/stable/j.ctvjf9vz4. International Bank for Reconstruction and Development. 2020. Sierra Leone Economic Update: The Importance of Investing in Girls.

[24] World Bank (2020). Sierra Leone Economic Update. Available at: https://documents1.worldbank.org/curated/en/131511593700755950/pdf/Sierra-Leone-Economic-Update-2020-The-Power-of-Investing-in-Girls.pdf

[25] Zhao, Y. (2022). Reconstruct more effectively: Prevent the learning deficit pitfall. Prospects, 51, 557–561. DOI: https://doi.org/10.1007/s11125-021-09544-y

https://doi.org/10.14505/jres.v16.2(20).04

Development of Students' Workbook with STEAM-Real World Problem to Improve Middle School Students' Problem-Solving Skills on Temperature and Heat Material

Aprilia CAHYANINGTYAS
Department of Science Education, Faculty of Mathematics and Natural Sciences
Universitas Negeri Malang, Indonesia
aprilia.cahyaningtyas.2103516@students.um.ac.id

Fatin Aliah PHANG
Department Innovative Science and Mathematics Education
Faculty of Educational Sciences and Technology
Centre for Engineering Education, Universiti Teknologi Malaysia, Malaysia
p-fatin@utm.my

Erni YULIANTI
Department of Science Education, Faculty of Mathematics and Natural Sciences
Center of Research and Innovation in STEM Education
Universitas Negeri Malang, Indonesia
Corresonding author: erni.yulianti.fmipa@um.ac.id

Article info: Received 7 July 2025; Revised 10 August 2025; Accepted 12 September 2025; Published 30 December 2025.

Abstract: Problem-solving skills are essential for students to face the challenges of the 21st century. However, based on observations and PISA 2022 data, junior high school students in Indonesia still have low problem-solving skills. The learning media used in the learning process is still conventional, still does not support problem-solving skills. This study aims to develop learning media in the form of students' workbooks with STEAM-Real World Problem content to improve students' problem-solving skills in temperature and heat material. The development followed the 4D model (Define, Design, Develop, Disseminate). The product was validated by 2 media and material experts, and readability was tested by 2 science teachers and 10 students. A total of 8 pretest and posttest questions covering problem-solving indicators were tested for validity and reliability, with valid results and a Cronbach's Alpha value of 0.665 (reliable). The empirical test involved 26 grade VII students. The effectiveness test used a quasi-experimental design with a non-equivalent control group model in two classes (28 students each). Results showed significant improvement, with an N-Gain of 61.81%. The "implementing the plan" indicator showed the highest increase, from an average score 8.81 (pretest) to a score of 68.19 (posttest). This improvement was supported by workbook activities such as designing a simple cooler, creating a poster, and calculating heat. These activities can provide space for students through experiments, visualization, creativity, and mathematical calculations. Thus, the students' workbook developed has valid, reliable, and effective criteria in improving students' problem-solving skills.

Keyword: students' workbook; STEAM; problem-solving skills; temperature and heat.

JEL Classification: I21; I23; C02.

Introduction

Education will always continue to evolve with the times. Education plays a very important role for every nation and country in creating a generation that is able to adapt to technological advances. Developments in the 21st century require students to be prepared to survive in a rapidly changing world and to create students who are able to compete in the world of education, both in academic and non-academic fields (Kain *et al.* 2024). These skills, often referred to as 21st-century skills, are the development of 4C competencies: communication, creativity, critical thinking and problem-solving, and collaboration (Herlinawati *et al.* 2024). Problem-solving skills are very

important for students in facing developments in the 21st century, so that students can effectively solve problems they encounter in their daily lives (Bessas *et al.* 2024).

Problem-solving skills are high-level thinking skills that require students to analyze problems, evaluate actions, and find appropriate solutions. Problem-solving skills can also be defined as the process of finding, organizing, and presenting information to solve problems that arise in students' daily lives (Boetje *et al.* 2024). In addition, problem-solving skills are very important in science learning to build a deeper understanding in students to train them to independently discover various concepts (Schäfer *et al.* 2024).

Indonesian students' problem-solving skills are still relatively low. Based on the results of the Program for International Student Assessment (PISA), where the average score of Indonesian students in the creative problem-solving domain is significantly below the OECD average, with most students only able to solve simple problems (OECD, 2014). This low achievement is influenced by the lack of emphasis on problem-solving learning in the Indonesian curriculum when compared to countries such as Singapore (Safrudiannur & Rott, 2019). This finding is in line with the results of PISA 2022 which showed that 59% of Indonesian students were below the minimum proficiency level in mathematics, reading, and science which are closely related to critical thinking skills and contextual problem-solving (OECD, 2023). Problem-solving skills are closely related to learning activities. If students are unable to solve problems in physics, their learning activities will be disrupted and their involvement in the learning process will tend to decline. The application of problem-solving skills in physics learning can show that they can gain meaningful learning experiences and that physics material is closely related to students' daily lives (Muñoz Alvarez *et al.* 2025).

Physics education has great potential to improve students' problem-solving skills. Temperature and heat are closely related to real-life contexts, making them important basic concepts in the physics learning process. However, many students still find it difficult to understand energy changes and heat transfer, which confirms that temperature and heat are basic concepts that students often struggle with (Brundage *et al.* 2024). Although students still find physics learning complicated, mastering physics concepts is important in finding solutions and solving problems they encounter in their daily lives. Physics learning requires structured problem-solving stages. The problem-solving stages that can be used in physics learning are the problem-solving model proposed by George Polya.

According to Pólya (1973), the stages of Problem-solving include understanding the problem, making a plan, implementing the plan, and re-checking the results obtained. There are previous studies that have implemented Polya's problem-solving stages in learning, such as in the study of Chacón-Castro *et al.* (2023) which shows that the application of Polya's problem-solving steps can strengthen students' problem-solving skills in the context of differential equations and can also be used as a reference to improve students' ability to connect concepts with real-world problem-solving in physics learning in the classroom, especially in the material on temperature and heat. However, this study has not yet integrated a conceptual or real-world approach and is not based on STEAM-real world problems.

The approach that can be used to improve problem-solving skills is the Science, Technology, Engineering, Arts, and Mathematics (STEAM) approach. The STEM approach has developed into STEAM by adding elements of art to strengthen aspects of creativity and contextual understanding in learning. According to Yulianti, *et al.* (2024) the integration of art in STEAM not only expands the problem-solving approach but also supports the achievement of Education for Sustainable Development (ESD) goals. This approach encourages students to think critically, creatively, and collaboratively in facing real-world challenges. The STEAM approach aims to equip students with the skills needed in the 21st century and has been the focus of research for the past decade. The role of the STEAM approach in the 21st century is very important in equipping students, so its implementation needs to use a learning approach that can actively involve students (Yulianti, *et al.* 2024).

In its implementation, the STEAM approach can be combined with real-world problem-based learning strategies. According to Connor *et al.* (2015) this approach focuses on student problems that are relevant to everyday life. The combination of the STEAM approach with Real World Problem-based learning can build student knowledge in developing problem-solving skills and creating a good learning atmosphere. According to Yakman (2008) the STEAM approach is an integrative model that combines the fields of science, technology, engineering, art, and mathematics in one interrelated learning framework. The STEAM approach aims to build interdisciplinary connectivity as a whole, not only focusing on mastering content, but also improving creative and critical thinking skills in improving understanding of real-world problems. The addition of arts elements includes not only visual arts, but also language, humanities, and social sciences. The addition of arts reflects the importance of the creative aspect in understanding and applying science and technology.

The implementation of STEAM-real world problems in the learning process can be optimized by using learning media that is appropriate for students. One effective medium for learning is the student workbook, as it serves as a guide for structured activities that encourage students to actively build knowledge independently through systematic activities. Student workbooks can be an important tool for students in the learning process to facilitate exploration and reflection in learning activities (Falbe *et al.* 2023). Student workbooks with STEAM-real world problem content can be found in simple concepts that exist in everyday life. These concepts are integrated into the fields of science, technology, engineering, arts, and mathematics that are relevant to technological advances and have been proven to increase students' creativity and skills in facing developments in natural sciences and technology (Filipe *et al.* 2024). Therefore, students' workbook learning media are suitable for use as an effective means of improving problem-solving for junior high school students to deepen their understanding of concepts in the topics of temperature and heat.

Based on data analysis from interviews and observations conducted with teachers and students in grades VII and VIII at a junior high school in Malang City, it was found that the learning resources used still have limitations in helping students understand the concept of temperature and heat. Learning resources generally used by teachers include student worksheets, presentations in the form of Power Point (PPT), and various other methods designed to help students understand the material. However, they still find it difficult to understand basic concepts, such as how to interpret observation results, understand the relationship between variables, and process information into a deeper understanding.

Observations with science teachers at a junior high school in Malang showed that teachers did not have media that was exactly the same as the students' workbooks. In addition, the learning media currently used has not undergone much development, especially in the context of implementing the Merdeka Curriculum Merdeka Curriculum. Generally, teachers still rely on teaching materials in the form of Student Worksheets that only contain a small amount of information and explanations related to the material on temperature and heat. This causes students to have difficulty understanding the concepts contained in the material. In addition, it has not been able to improve problem-solving skills and is not integrated with STEAM aspects. Based on the results of the needs analysis that has been carried out, teachers need a students' workbook learning medium to facilitate the improvement of students' problem-solving skills in temperature and heat material.

Based on a study by Co (2025), the use of workbooks as supplementary teaching materials to improve science learning using traditional teaching methods has been proven to significantly improve student learning outcomes in science. Thus, workbooks can improve science learning outcomes. However, this study is still limited to the use of conventional workbooks and has not yet been integrated with the STEAM approach, which links learning to real-world problems. In addition, this study is not yet based on problem-solving skills. Similar research was also conducted by Putri *et al.* (2025) stating that learning media by developing a thermodynamics physics e-book that integrates local coal production policies with the STEAM-PjBL approach has been proven to be valid, practical, and effective in improving students' computational thinking skills with an n-gain value of 0.66. However, this study has not developed a problem-solving-based students' workbook and has not linked it to real-world problems.

Based on the results of the two studies above, the author now wants to develop a product that has integrated STEAM-real world problem content and has problem-solving indicators, namely the students' workbook learning media. Based on the explanation of the problem that has been presented, this study focuses on "Development of Students' Workbooks with STEAM-Real World Problem Content to Improve Problem-solving Skills of Junior High School Students on Temperature and Heat Materials", which aims to 1) identify the need for students' workbook learning media with STEAM-real world problem content for junior high school students to improve problem-solving skills on the topic of temperature and heat, 2) develop students' workbooks with STEAM-real world problem content for junior high school students on the topic of temperature and heat to improve valid and feasible problem-solving skills, and 3) test the effectiveness of students' workbooks with STEAM-real world problem content in improving problem-solving skills of junior high school students.

1. Literature Review

1.1 Students' Workbook

Currently, the world of education has entered the era of revolution 4.0, where in this era it is required to be able to create interactive and personalized student learning experiences that encourage critical thinking, problem-solving, and creativity. Therefore, in facing the challenges of the industrial era 4.0, the transformation of human resources in the education sector must continue to develop. Educators are tasked with identifying and developing the skills needed in the face of change. The 21st century skills needed to meet the needs of education, because it can

realize a learning model that suits current needs (Saleem *et al.* 2024). One form of learning innovation that supports the development of student creativity and learning independence is the students' workbook. Students' workbook is a combination of books and student worksheet specifically designed as a learning resource for students to learn the material. Students' workbooks developed must meet the criteria of content that is in accordance with the curriculum, contextual, contains STEAM aspects, and also needs to contain assessments that reflect indicators of problem-solving skills. Students' workbooks can provide opportunities for students to practice independently, test their understanding, and support students to improve their skills in understanding the concepts they have learned. The students' workbook also contains objectives, materials, and several learning activities related to temperature and heat material (Castro-Velásquez *et al.* 2024).

According to Muliyati *et al.* (2023) the development of STEM project-based worksheets for physics learning which aims to evaluate the suitability of STEM-PjBL worksheets as learning media that can assist teachers in teaching renewable energy topics. This worksheet has several components in it that can help students understand renewable energy material effectively through project-based activities. However, the development of the student worksheet is still limited to the STEM approach that has not included Arts (A) elements, the connection to real world problems is still limited, and does not focus on problem-solving indicators. Similar research was also conducted by Annuš *et al.* (2024) to develop the Learn with M.E digital platform which includes a digital students' workbook feature aimed at supporting mathematics learning at the K-12 level. The results of the study show that digital workbook students are able to provide facilities in learning and help students learn according to their abilities. However, this research still does not integrate the STEAM-real world problem approach and there are no problem-solving skills.

1.2. Problem-Solving Skills

Teachers really need learning methods during the teaching process in order to achieve the expected targets. The methods provided must be able to make students achieve optimal learning achievements and as expected. Problem-solving skills are very important for students to overcome challenges, make the right decisions, and achieve the desired results. Problem-solving skills can be used to analyze a problem and design a solution strategy to find relevant solutions, and consider alternative results that will be obtained as needed (Tursynkulova *et al.* 2023). Problem-solving skills are used to challenge students in question their prior knowledge, increase curiosity, and encourage the search for solution ideas during the learning process (Stuppan *et al.* 2025).

According to Busyairi *et al.* (2023) the development of physics learning aids designed with the STEM-creative problem-solving model on dynamic electricity material is proven to have high validity and it can be concluded that the physics learning aids that have been developed are suitable for use in the trial stage and disseminated. This physics learning aid is used to improve problem-solving creativity in students.

1.3. STEAM

According to Portillo-Blanco *et al.* (2024) STEM education has been around since the early 1990s, STEM can create meaningful learning experiences to equip students with 21st century skills. At learning using the STEM approach requires learning related to real-world problems. According to Filipe *et al.* (2024) the change from STEM to STEAM with the addition of the element “A” (Arts) has a huge impact on the world of education. The addition of arts elements to science, technology, engineering and math learning aims to increase creativity and innovation through an arts approach, and can encourage students to experiment and create new solutions in solving problems. The STEAM approach enhances motivation, collaboration and can foster 21st century skills, making learning more meaningful and relevant to real-world challenges.

According to Trowsdale *et al.* (2024) the development of STEAM-based learning through “The Imagineerium” and “Teach-Make” projects using the Trowsdale Art-Making for Education (TAME) curriculum model focuses on real-world and transdisciplinary practices that integrate Science, Technology, Engineering, Arts, and Mathematics into project-based learning. The use of STEAM methods in the Trowsdale Art-Making for Education (TAME) approach is proven to be used in education that focuses on real-world transdisciplinary practices, so as to increase creativity and problem-solving in elementary school students.

1.4. Temperature and Heat

Physics is a branch of science that studies natural phenomena related to matter, energy, space, time, and the various forces that affect them. One important topic in physics is temperature and heat, where the concepts covered include the shape of objects due to heat, the black body principle, and heat transfer (Melita *et al.* 2023). Temperature and heat are closely related to everyday life. At the junior high school level, the material taught

related to temperature and heat includes basic concepts of temperature and heat, temperature measurement, heat transfer, and other basic thermodynamic concepts. However, the material on temperature and heat is known to be difficult for students to understand. Studies show that many students still have difficulty distinguishing between the concepts of temperature and heat, as well as connecting temperature with physical measurements (Xing *et al.* 2023).

According to Awudi *et al.* (2023), the application of the demonstration method in teaching the concept of heat has been proven to improve students' conceptual understanding, as it encourages students to be more active and critical and enables them to relate the concept to real-world problems. This study also concluded that learning about heat is proven to be effective using the demonstration method. However, this research has not yet been integrated with STEAM content and does not yet include in-depth aspects of problem-solving skills.

Broadly speaking, students' problem-solving skills still need to be improved. The use of learning media can support students in overcoming difficulties encountered during learning. Based on some of the above studies that have discussed related to the making of student worksheet that use several methods and learning models of Discovery Learning and PjBL, Problem-solving, and temperature and heat material using teaching modules based on Problem Based Learning can optimize student activeness in the learning process and help them in trying to find problem-solving and create memorable learning experiences. However, the student worksheet used only consists of activity sheets and evaluation questions. The use of Hypercontent-based Discovery Learning learning model on temperature and heat material is not conducive. This is because many students' smartphones still do not have QR-code applications and earphones to listen to learning videos properly. Some researchers only use research subjects in one class.

Based on previous research, the development of teaching materials is still conventional, has not been integrated with the STEAM approach, is not associated with the context of real world problems, and does not contain problem-solving indicators. Seeing these limitations, the author sees the need to develop products in which not only contain problem-solving indicators, but also integrate the STEAM approach and real world problems. The author develops a STEAM-real world problem student's workbook which contains temperature and heat learning materials presented in a logical order, contains exercises, questions, activities, tasks designed to help students deepen their understanding of temperature and heat material, and is integrated with the STEAM approach by using 2 classes as subjects, namely classes that use student's workbook in learning (as an experimental class) and classes whose learning does not use student's workbook (as a control class).

2. Method

This research is a development study using a Research and Development (RnD) model focused on creating learning products in the form of materials, media, tools, and learning strategies that are useful for learning rather than for testing theories. The main focus of this research is to produce innovative learning media, then test the effectiveness of the products to determine whether they are suitable for use in the learning process (Pratama, 2025). This study focuses on the development of Students' Workbook with STEAM-Real World Problem content on temperature and heat material for grade 7 junior high school. The sampling technique in this study was purposive sampling. The research and development design model used in this study is 4-D development developed by (Thiagarajan *et al.* 1974). Figure 1 depicts the research flow, which was modified based on the method introduced by Thiagarajan, it illustrates the structured steps undertaken to achieve the research objectives. This 4D model consists of 4 stages, namely define, design, develop, and disseminate or can also be abbreviated as 4D. The stages in 4D development are arranged simply, can be easily understood, simple, and its implementation takes place systematically. In addition, this development model is generally applied in the development of books or teaching materials. Furthermore, to test the effectiveness of the product, the design used in this study was a quasi-experimental design with a non-equivalent (pretest and posttest) control-group model with two groups which appears in table 1 (Creswell, 2014), namely the class that used students' workbooks (experimental class) and the control class that did not use students' workbooks.

Figure 1. Product Development Framework Using 4D Stage

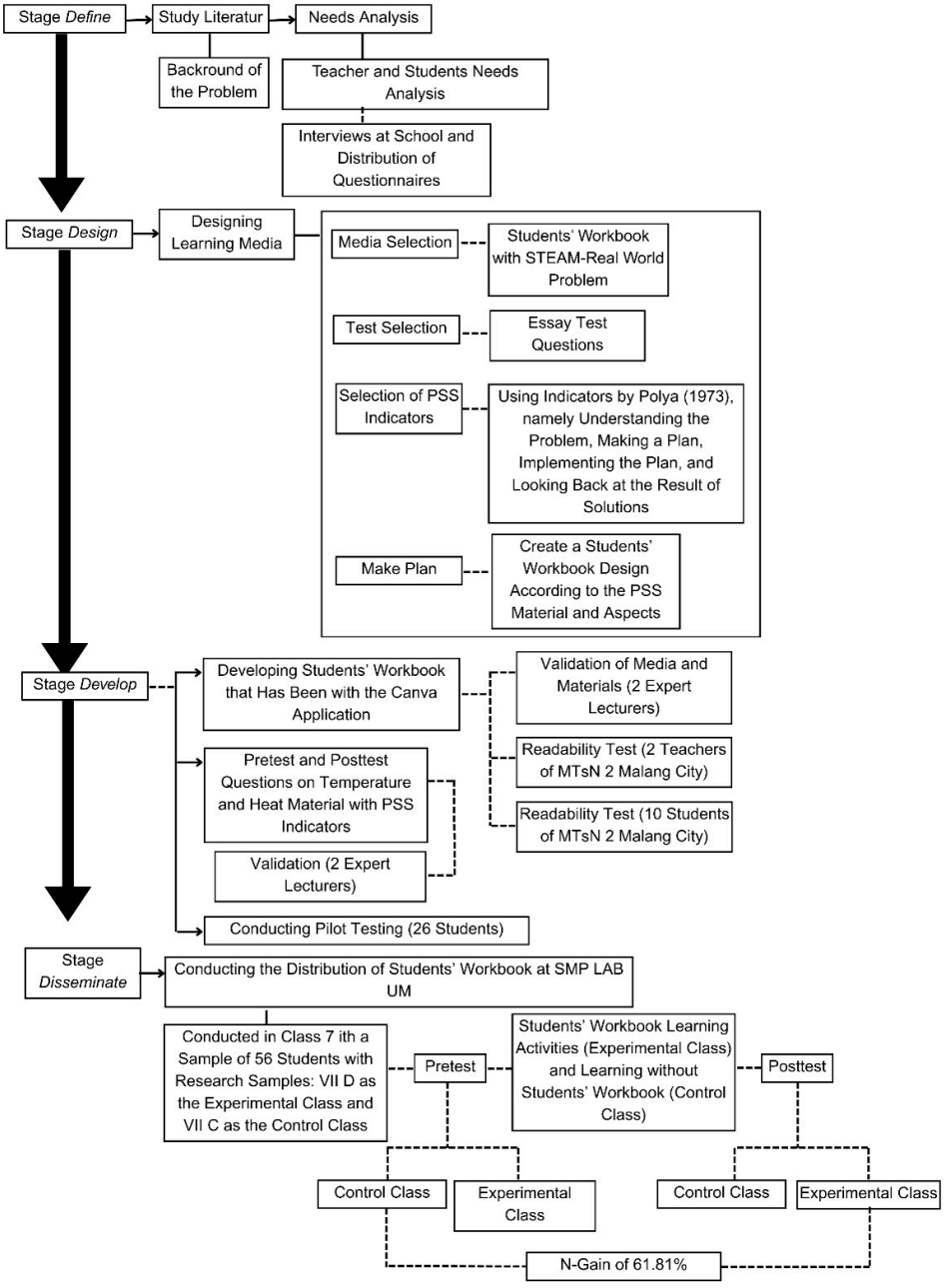


Table 1. Research Design

Group	Pretest	Treatment	Posttest	information
Experimental	O	X1	O	Was given treatment using the students' workbook
Control	O	X2	O	Conventional learning treatment uses teaching modules

Source: (Creswell, 2014)

Description:
O : Measurement (both pretest and posttest)
X_1 : Treatment (learning using students' workbook)
X_2 : Treatment (conventional learning using teaching modules)

The first stage is the defined stage which can help in determining and providing explanations related to needs and collecting information about user needs that will be included in the product. This study was conducted

in three different junior high schools in Indonesia. Junior High School A was used to analyze the initial needs for learning media, Junior High School B was used to test the validity of the media developed prior to implementation, while Junior High School C was the location for the implementation of media that had undergone the development and validity testing stages previously. The needs analysis was conducted at Junior High School A which was conducted by interviewing science teachers, grade VII students (who had not received temperature and heat learning), and providing instruments in the form of questionnaires to grade VIII students (who had received temperature and heat learning) to identify the problem-solving skills gap and the need for students' workbook media with STEAM-Real World Problem content on temperature and heat material.

The second stage is the design stage, which includes designing learning media for students' workbooks with STEAM-Real World Problem content such as determining a theme that is in accordance with the temperature and heat material so that it looks attractive to junior high school students, finding information about the material that is in accordance with the curriculum and learning outcomes of junior high school, choosing good fonts and colors so that they are easy to read, determining a cover design that reflects the theme of the material, and determining the code for the STEAM category. Then, create several activities that are arranged to support students' understanding in improving problem-solving skills.

The third stage is the development stage, which includes the creation of a product in the form of a learning media for students' workbooks with STEAM-Real World Problem content. At this stage, product development and design, such as font selection, color, placement of graphic elements, STEAM icons, and creation of book covers using the Canva application. Students' workbooks with STEAM-Real World Problem content are designed based on needs analysis, curriculum suitability, and characteristics of junior high school students. In the students' workbook, there is material on temperature and heat and several activities with STEAM-Real World Problem content to improve problem-solving skills. At the development stage, there are also supporting instruments in the form of pretest-posttest questions containing problem-solving indicators according to Pólya (1973). Table 2 shows that the questions developed consist of 8 problem-solving questions containing different indicators for each question.

Table 2. Aspects of Problem-solving Indicators in Instruments essay test and Media

Indicators Problem-Solving	No Questions on Pretest and Posttest
Understanding the Problem	1, 2, 4, 5, 6, 7, 8
Makin a Plan	1, 2, 4, 5, 6, 7, 8
Executing the Plan	1, 2, 3, 4, 5, 6, 7, 8
Looking Back at Results or Solutions	3, 4, 6

After the product is developed, the next step is the validation process by material experts and media experts, followed by a readability test by teachers, a readability test by students who are asked to assess the feasibility, content and appearance of the product. The product consisting of students' workbook learning media and pretest posttest questions is validated by two science lecturers who are experts in physics and learning media. Validation is carried out to obtain input, criticism, and suggestions from experts so that the product can be improved before being tested, so that it can increase the effectiveness, readability, and suitability of the product to the needs of students and the curriculum. At this stage, the validation results obtained quantitative data, namely the assessment of the two validators using a validation sheet with a Likert (1932) assessment scale criteria 1-4, as well as from the results of the validation of the concept truth using the Guttman scale (1944), where each correct answer is given a score of 1, while the wrong answer is given a score of 0. The validation results from the two validators are used as a reference for revision and to perfect the product before being tested and the validation results are analyzed to calculate the percentage of feasibility.

After that, the product can be declared fit for use. Furthermore, a readability test was conducted on the students' workbook media by two science teachers and 10 students, which aimed to determine whether the use of the students' workbook could be understood and used properly. In addition, there was also a trial of the questions (empirical test) on the pretest and posttest questions to ensure that the instrument had met the validity and reliability requirements. Research instruments can only be considered high quality if they have high validity and reliability (Sharma *et al.* 2024). The validity test was conducted to ensure that the instrument actually measures the intended aspects. Meanwhile, reliability was tested to determine the level of stability of the measurement results of the instrument. According to Desnita (2022), Validity indicates the extent to which an instrument is able to accurately measure what it is supposed to measure, while reliability describes the stability of measurement results when used under the same conditions. The empirical test was conducted at junior high school B to 26 grade VII students. The questions given were 8 questions that tested problem-solving skills. The data results from

the validity test using Pearson Product Moment Correlation to see the relationship between the question score and the total score. If the correlation is significant, then the question item is declared valid. While the reliability test is tested using methods such as Cronbach's Alpha. This study uses SPSS Statistics software version 25.0 in the data processing process.

The fourth stage is dissemination; at this stage the effectiveness test is carried out on the students' workbook media that has been developed with the aim of determining the extent to which the students' workbook media is successful in improving students' problem-solving skills. Data collection used 2 classes, namely class VII D which is an experimental class and was given treatment in the form of using students' workbooks, and class VII C, which is a control class and did not receive treatment. In this study, the data collection process was carried out at Junior High School C involving class VII students, each class consisting of 28 students. The study was conducted in six meetings for each class, consisting of one pretest, four learning meetings, and one posttest. In the experimental class, learning was carried out using students' workbooks containing 16 STEAM-real world problem-based activities. Student worked on 4 activities each meeting during four learning meetings, so that all activities in the students' workbook were completed gradually. The activities are designed to train problem-solving skills through observing contextual phenomena, simple experiments, group discussions, and mini projects, such as making a cooler from styrofoam and educational videos about energy efficiency. Meanwhile, in the control class, learning was delivered using conventional methods using teaching modules with the delivery of temperature, temperature scale, expansion, and heat and its transfer, without using student workbooks. This effectiveness test used pretest and posttest questions in which there were problem-solving indicators. The results of the pretest and posttest of both classes were used to collect quantitative data. Analysis of the help of the pretest and posttest results was carried out using SPSS Statistic 25.0 software as an aid. Furthermore, it was analyzed to see the improvement in student problem-solving. The analysis stage carried out first in this study was the equipment test including the normality test used to ensure data distribution and the homogeneity test to determine the differences in variance between classes. Then the Independent Sample t-Test was carried out to determine the differences in the posttest results of the experimental class and the control class. After that, an n-gain test was carried out to measure the effectiveness of student learning in improving understanding of the concept of temperature and heat by using the STEAM-real world problem student's workbook. This test compares the pretest value (before learning) with the posttest value (after learning) to determine the extent of the improvement that has occurred.

3. Research Result

This study develops a product in the form of learning media, namely students' workbook with STEAM-Real World Problem content to improve problem-solving skills of junior high school students in grade VII on temperature and heat material. This students' workbook media is almost the same as student worksheet, but there are a number of advantages, namely this students' workbook is designed according to the independent curriculum for 4 meetings, in the students' workbook there is temperature and heat material combined with various activities that integrate STEAM elements (Science, Technology, Engineering, Arts, and Mathematics) and equipped with problem-solving indicators by Pólya (1973). In addition, this students' workbook is equipped with S-T-E-A-M code icons that appear in each activity, and the appearance of the students' workbook is designed attractively, colorfully, and not monotonously in order to increase interest and the active role of students in learning activities. In Figure 2. The front view of the students' workbook can be seen.

Figurea 2. STEAM integrated students' workbook front page

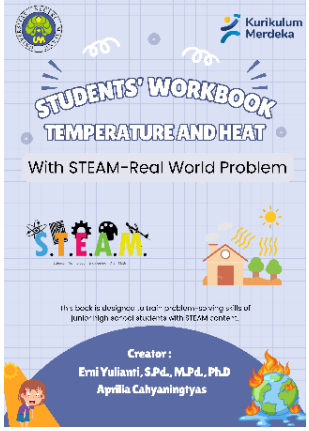
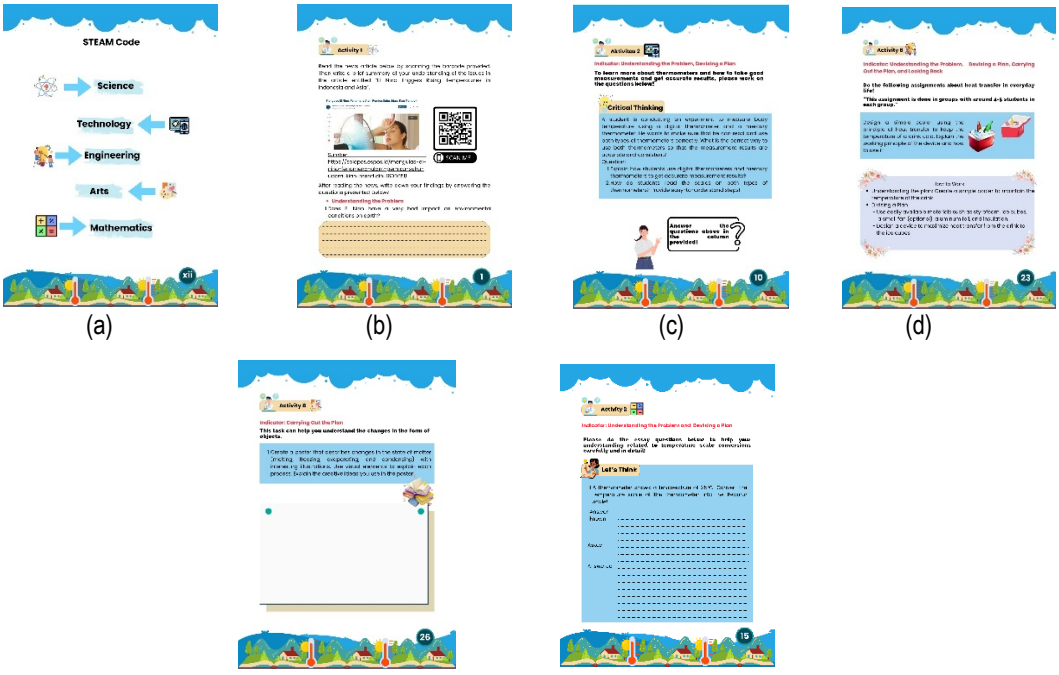


Figure 2. is the cover of the students' workbook used in the research with STEAM-Real World Problem content to improve problem-solving skills. On the front page of the students' workbook there is the title of the material, the author's name, and the approach used. An example of an activity from the students' workbook is in Figure 3.

Figure 3. STEAM Integrated Students' Workbook Activities; (a) S-T-E-A-M Icon Found in Each Activity in the Students' Workbook; (b) Activity (Science) and Problem-solving Practice "Understanding the Problem"; (c) Activity (Technology) and Problem-solving Practice "Understanding the Problem and Making a Plan"; (d) Activity (Engineering) and Problem-solving Practice "Understanding the Problem, Making a Plan, Implementing the Plan, and Rechecking the Results Obtained"; (e) Activity (Arts) and Problem-solving Practice "Implementing the Plan"; (f) Activity (Mathematics) and Problem-solving Practice "Understanding the Problem and Making a Plan"



(a) (b) (c) (d)

(e) (f)

Figure 3 is a STEAM-integrated students' workbook activity used in this study. The students' workbook also contains CP and TP, instructions for use, S-T-E-A-M code icons, and a column for writing students' answers. The questions in the students' workbook encourage students to improve their problem-solving through real-world problems. The steam aspects integrated in this study are shown in Table 3.

Table 3. STEAM Aspects

No.	Aspect	Description	Implementation in Students' Workbook
1	Science	The Science aspect focuses on understanding the natural world through observation, experimentation, and reasoning based on the sciences of physics, chemistry, and biology.	Students conduct observations and experiments on temperature and heat, changes in state, and heat transfer (conduction, convection, radiation) to understand the basic concepts of physics on the topic of temperature and heat.
2	Technology	The Technology aspect focuses on the processes, tools, and knowledge used, developed, and operated in everyday practical solutions.	Students use digital thermometers, temperature sensors, and simulations to understand and visualize heat phenomena. In addition, they also create educational videos using digital technology.
3	Engineering	The Engineering aspect focuses on the ability to design and engineer innovations to solve problems that connect the concepts of science, mathematics, and technology.	Students design and make a simple cooling device to keep the temperature of drinks cold by utilizing the principle of heat transfer.
4	Arts	The Arts aspect focuses on everything about art that can help in finding innovative and creative solutions.	Students create posters about changes in the state of matter and the greenhouse effect, and present the project results in a visually appealing and

No.	Aspect	Description	Implementation in Students' Workbook
			creative way.
5	Mathematics	The Mathematics aspect focuses on the basis of understanding patterns, quantitative relationships, geometric shapes, and the application of mathematical calculations in problem-solving.	Students calculate the amount of heat, temperature changes, and temperature conversions using physics formulas and data analysis.

Source: (Bahrum *et al.* 2017)

After the students' workbook learning media was developed, the next stage was validation. The validity data of the students' workbook that supported this research was conducted by two lecturers of the Science Education Study Program who were experts in the field of physics. This validity test aims to assess the feasibility, quality, and suitability of the learning media before the students' workbook is implemented. In media validity, material validity, the validator assesses each aspect using a 4-point Likert scale, which includes the criteria of "strongly disagree", "disagree", "agree", and "strongly agree", so that the percentage of validity results are obtained for each aspect assessed. Each indicator in the validation sheet is filled in by the validator according to his/her perception of the quality of the aspect being assessed. After the data is obtained, calculations are carried out to obtain the percentage of validity in each aspect. Both validators use a quantitative approach.

Table 4. Media Validity

No	Aspects	Percentage (%)
1	Presentation Feasibility	87.4%
2	STEAM Aspect Accuracy	87%
3	Media Functional Feasibility	85.16%
Average Result		86.52%
Description		Very Valid

The results of the students' workbook media sheet with STEAM-Real World Problem content to improve students' problem-solving skills obtained a validity of 86.52% as seen from table 4, including the very valid category. From the results of the media validity, the students' workbook is suitable for use in learning activities, as explained by (Akbar, 2016). In addition, the validator also provided general notes that the students' workbook product was good and could be tested without revision.

Table 5. Material Validity

No	Aspects	Percentage (%)
1	Suitability of Material	91.33%
2	Relevance of Source Material Used	87.25%
3	Truth of Concept	100.00%
Average Result		92.86%
Description		Very Valid

The results of the validation at Table 5 show that the students' workbook media with STEAM-Real World Problem content to improve students' problem-solving skills obtained a validity of 92.86% with a very valid category. Thus, the students' workbook that was developed has met the eligibility standards as a learning medium, in accordance with the assessment criteria put forward by (Akbar, 2016). In addition, the validator also provided a general note that the students' workbook product was good and could be tested without revision.

Table 6. Problem-solving Skills (PSS) Validity

No	Aspects	Percentage (%)
1	Problem-solving Skills	100.00%
Average Result		100.00%
Description		Very Valid

The results of the validation on Table 6 of the Problem-solving Skills (PSS) aspect using the Guttman scale, from 32 statements all were stated as "Yes" by the validator, so that the validity score was 100.00%. This means that the contents of the activities and indicators that have been developed are in accordance with Pólya's theory (1973), relevant to the STEAM approach, and support students' problem-solving skills in temperature and heat material. Based on the eligibility criteria, the results of 100.00% indicate that the problem-solving Skills indicator in the students' workbook is declared very valid. This shows that the students' workbook that was developed has met the eligibility criteria for use in learning, as the assessment criteria explained by (Akbar, 2016). In addition, the validator also provided a general note that the students' workbook product was good and could be tested without revision.

The students' workbook learning media with STEAM-Real World Problem content, after being declared valid by experts, was then tested on teachers and students, through a readability test. The readability test was conducted at Junior High School B, involving 2 science teachers and 10 students. The purpose of the readability test was to determine the extent to which the students' workbook learning media can be easily understood by students and effectively used in learning. Students were given time to read the contents of the students' workbook before being given a validation sheet. After that, students filled out the readability test validation sheet (students). The results of the student readability test showed that all students stated that the students' workbook was easy to understand and they could follow the activity instructions well. Both teachers also said that the students' workbook was suitable for use in the learning process, with the note that each activity was given an activity/activity title, or the purpose of the activity was written, the choice of language that was easy for students to understand, and it was necessary to understand the difference between objectives and indicators.

In addition to developing students' workbook learning media, researchers also developed pretest posttest questions using Polya's problem-solving indicators in the form of descriptive questions consisting of 8 questions. The development of these questions aims to improve problem-solving before and after learning using students' workbooks. Before the pretest posttest question instrument is used for the effectiveness test, a validity and reliability test is first conducted. The empirical test was conducted at Junior High School B in class 7C with 28 students. The results of the empirical test were in the form of Pearson Bivariate correlation results (validity test) and Cronbach's Alpha (reliability test).

Based on the results of the validity analysis conducted using Pearson's Bivariate correlation on the questions developed, namely 8 questions. It was found that all questions were declared valid. This is based on the results of the calculated r of each question being greater than the r table (0.388) or the significance result being less than 0.05. Furthermore, a Cronbach's Alpha reliability test was conducted. The results of the reliability test on 8 valid questions were 0.665, which is included in the high reliability category. These results indicate that the 8 questions meet the requirements for use and are suitable for use in research.

Before testing the effectiveness of the product, prerequisite tests are carried out first, namely normality tests and homogeneity tests. Table 7 presents the results of the normality test and Table 8 presents the results of the homogeneity test.

Table 7. Results of the Normality Test Pretest Posttest Experimental and Control Classes

Class	Shapiro-Wilk Test		
	Sig. Level	Sig.	Conclusion
Experiment Pretest	0.05	0.122	Normal
Experiment Posttest	0.05	0.164	Normal
Control Pretest	0.05	0.482	Normal
Control Posttest	0.05	0.363	Normal

Based on the results of Table 7, all significance values (Sig.) in the Shapiro-Wilk test on each pretest posttest data of both classes are >0.05. From these results, all data are normally distributed

Table 8. Results of the Homogeneity Test for the Pretest Posttest of the Experimental and Control Classes

Class	Levene Test		
	Sig. Level	Sig.	Conclusion
Pretest experiment control	0.05	0.897	homogeneous
Posttest experiment control	0.05	0.958	homogeneous

Based on the results of Table 8, the significance value (Sig.) of the pretest-posttest data for both classes is more than 0.05, indicating that the variance of both classes is homogeneous.

Independent Sample t-Test

The next stage, after the prerequisite test has been fulfilled, is to conduct an effectiveness test using the independent sample t-test to see the differences between classes that use students' workbooks and those that do not.

Table 9. Result Independent Sample t-Test

Class	Different Test		
	Sig. Level	Sig. (2-tailed)	Conclusion
Pretest experiment control	Sig. >0.05	0.068	equal initial ability
Posttest experiment control	Sig. <0.05	0.000	different final ability

Table 10. Result Compare Mean Test

Result Posttest		N	Mean
Experiment Class		28	64.75
Control Class		28	14.25

In the results of Table 9. shows the results of the independent sample t-Test on the pretest data with a Significance value (2-tailed) of 0.068> 0.05, this indicates that there is no significant difference between the two classes. Based on these results, it can be concluded that students have the same initial abilities, as expected in the experimental design. Therefore, the pretest data was not used in further analysis. Meanwhile, the results of the t-test on the posttest data obtained a Significance value. (Sig. 2-tailed) of 0.000, less than 0.05. Thus, it can be concluded that there is a significant difference between the two classes. In the results of Table 10. Shows that the posttest results of the experimental class (students' workbook) obtained a mean of 64.75, while in the control class (without students' workbook) obtained a mean of 14.25. Therefore, the use of students' workbooks has a very positive impact on improving student learning outcomes.

Table 11. Result N-Gain Test

N-gain Test		N-gain Percent	Category
Experiment Class		61.81%	Significant improvement

Referring to the results of Table 11. the average value (mean) of N-gain using SPSS in the class given the students' workbook treatment was 61.81% indicating a significant increase according to Hake in A.A. Setyo *et al.* (2022). The use of students' workbooks has been proven to be significant in improving the problem-solving skills of junior high school students in the material of temperature and heat. In Figure 4. It can be seen that all problem-solving indicators have increased after being given the students' workbook treatment. The indicator that experienced the highest increase was the "implementing planning" indicator, where the pretest result percentage was 8.81, while during the posttest it increased significantly to 68.19, so that it increased by 674%. Thus, these results can prove that students are able to implement the indicators well, so that they can experience significant improvement. This increase is in line with learning activities using students' workbooks. In addition, activities in students' workbooks can not only train students in improving their problem-solving skills, but also to integrate STEAM elements in real terms. Through activities in the students' workbook, students are invited to apply science in understanding concepts, technology in compiling and implementing solutions, art in expressing creative ideas, and mathematics in calculations. As shown in Figure 5. Some activities in the students' workbook that are done by students.

Figure 4. Improvement in Pretest and Posttest Results for Each Indicator in the Experimental Class

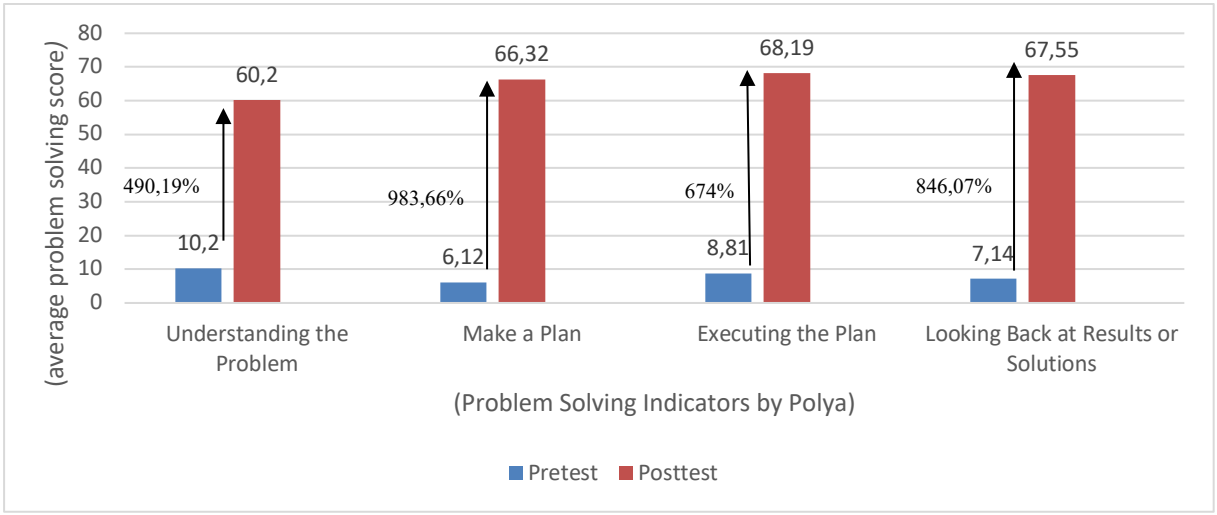
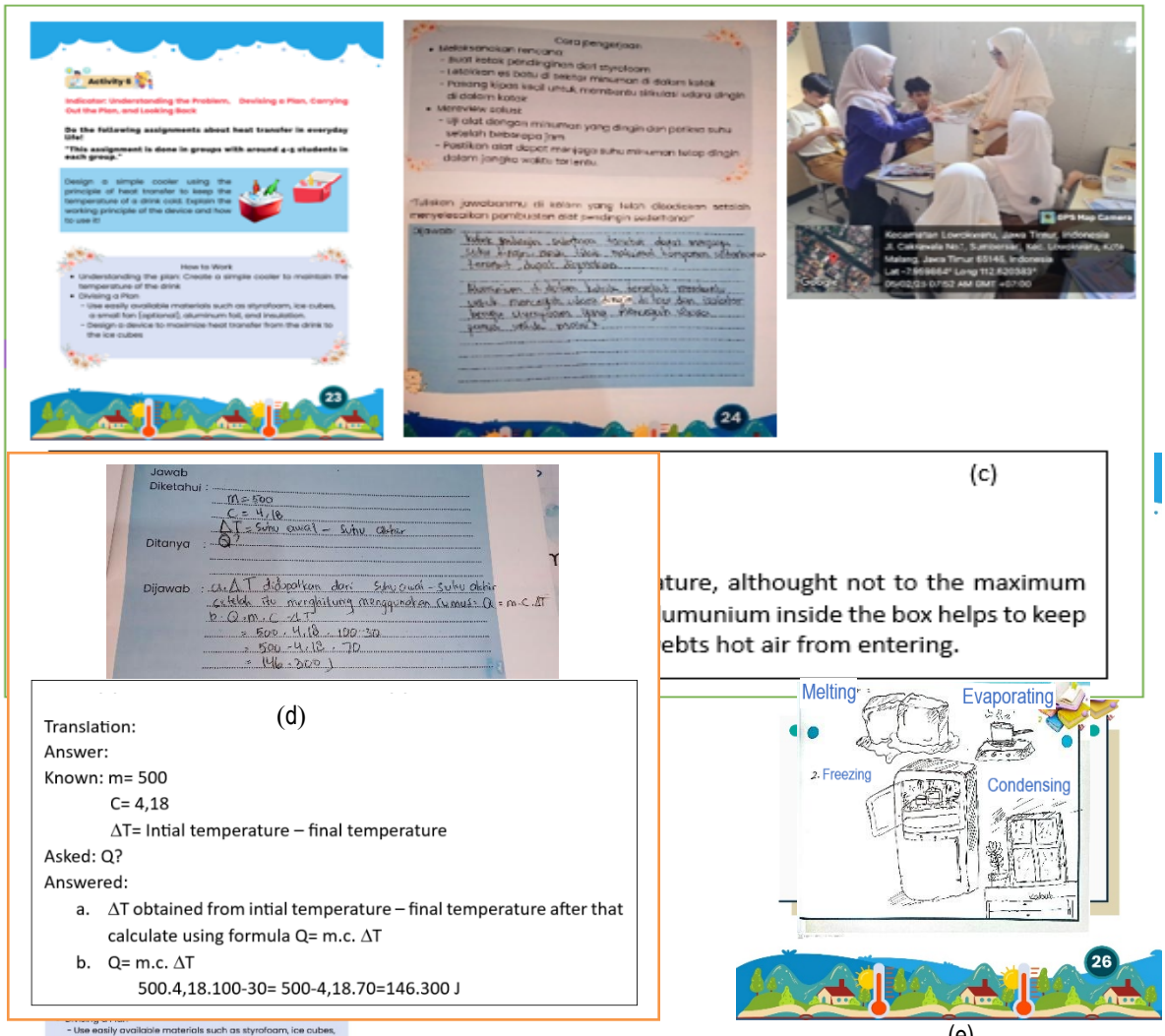


Figure 5. Several Students' Workbook Activities Towards Indicator Improvement; (a) Engineering Activities; (b) Students' Answers to the STEAM (Engineering) Aspect and Problem-solving Indicator "Understanding the Problem, Making a Plan, Carrying Out the Plan, and Rechecking the Results Obtained"; (c) Students' Process of Making Cooling Devices; (d) Students' Answers to the STEAM (Arts) Aspect and Problem-solving Indicator "Carrying Out the Plan"; (e) Students' Answers to the STEAM (Science) Aspect and Problem-solving Indicator "Understanding the Problem, Making a Plan, and Carrying Out the Plan".

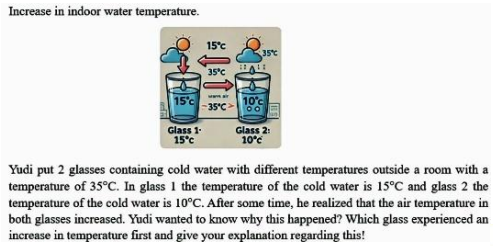


Based on the results of Figure 4. It is concluded that students' workbook is more effective in improving students' problem-solving skills than before using students' workbook. The low pretest means that students' problem-solving skills are lacking. However, after learning in the experimental class using students' workbook to improve students' problem-solving skills, there was an increase in scores (posttest). The following is a detailed explanation of students' answers to each indicator.

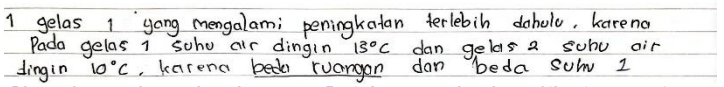
1.Understanding the problem

In the indicator of understanding the problem, the aspect that is emphasized is the students' skills in identifying important information from the problem expressed in the question. Students are given reading questions about increasing the temperature of water in a room. Based on the reading, students are asked to explain which glass experiences an increase in temperature first in the two glasses that have different temperatures outside the room with a temperature of 35°C and provide reasons. The appearance of the questions and answers for the students' pretest and posttest can be seen in Figure 6.

Figure 6. (a) Pretest and Posttest Questions on the Understanding Problem Indicator; (b) Students' Pretest Answers on the Understanding Problem Indicator; (c) Students' Posttest Answers on the Understanding Problem Indicator.

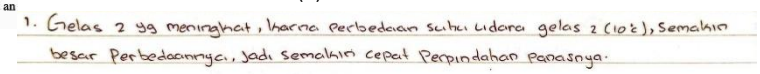


(a)



Glass 1 experienced an increase first, because in glass 1the temperature of the cold water was 15°C and in glass 2 the temperature of the cold water was 10°C, because of the different rooms and different temperatures.

(b)



Glass 2 increases, due to the difference in air temperature between glass 2 (10°C), the greater the difference, the faster the heat transfer.

(c)

It can be seen that the students' answers in Figure 6 (b), it appears that the students have not been able to understand the problem correctly. The students answered that glass 1 experienced an increase first, while the correct answer is glass 2, because it has a larger temperature difference with the room temperature, so it quickly receives heat. This shows that students do not understand the concept of heat transfer. In the posttest section with the same indicators and the same questions. It can be seen that the students' answers in Figure 6 (c) are able to provide the right answer. Based on the question about the increase in water temperature in two glasses, the students answered that glass 2 experienced an increase in temperature first because it had a larger temperature difference with the room temperature. This answer is correct and shows that students have focused on the question and understand the concept of heat transfer correctly.

The students' workbook activity that directly supports the improvement of the problem understanding indicator is Activity 7. There is a STEAM (Science) aspect. In this activity, students are asked to observe the phenomenon of changes in the state of objects and identify the types and causes. This activity can train students to recognize important information from the situation and understand the concept of heat transfer, as required in the problem understanding indicator according to Polya. The display of activity 7 and students' answers are attached in Figure 7.

Figure 7. Students' Workbook Activity Indicator Understanding Problems And Students' Answer

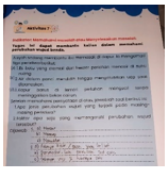


Figure 7. Students' Workbook Activity Indicator Understanding Problems and Students' Answers

Aisyah is helping her mother cook in the kitchen. She observes the following three events:

1. The ice cubes taken from the freezer slowly melt at room temperature.
2. The water in the pot boils and emits steam when heated.
3. The camphor in the closet slowly shrinks without leaving any liquid residue.

After understanding the above statement, answer the following questions:

1. What type of change of state occurs in each event?
2. What factors influence these changes?

Answer:

1. a) Melting 2. a) The presence of heat or heat involved
   b) Evaporate b) Heating, surface area of liquid
   c) Sublimate c) Vapor pressure and temperature relationship

2. Making a Plan

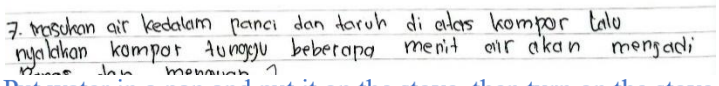
In the indicator of making a plan, the aspect highlighted is the students' skills in designing simple experiments to prove heat transfer through conduction, convection, and radiation in a kitchen environment. In this case, students are expected to be able to arrange the steps in a sequential, logical, and easy-to-understand manner. Students must also be able to clearly identify the parts of the experiment that show each heat transfer. The appearance of the students' pretest and posttest questions and answers can be seen in Figure 8.

Figure 8. (a) Pretest and Posttest Questions on the Planning Indicator; (b) Students' Pretest Answers on the Planning Indicator; (c) Students' Posttest Answers on the Planning Indicator.

Proving Heat Transfer Through Conduction, Convection, and Radiation in the Kitchen.

Mom was boiling water in a pot placed on a gas stove. After some time, she noticed that the air around the pot became warm. What are the steps you can take for a simple experiment in the kitchen to prove that heat from the stove moves through conduction, convection and radiation?

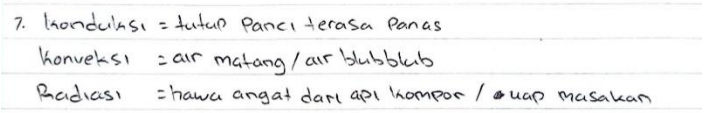
(a)



7. masukan air kedalam panci dan taruh di atas kompor lalu nyalakan kompor tunggu beberapa menit air akan mengadi

Put water in a pan and put it on the stove, then turn on the stove, wait a few minutes, the water will become hot and evaporate.

(b)



7. konduksi = tutup panci terasa panas
konveksi = air matang / air blubblub
Radiasi = hawa angat dari api kompor / uap masakan

Conduction = the lid of the pan feels hot
Convection = boiled water / blubblub water
Radiation = warm air from the stove / cooking steam

(c)

It can be seen that the students' answers in Figure 8 (b), it appears that the students have not been able to link the heat transfer process specifically through conduction, convection, and radiation, and have not mentioned the parts involved in each process. This shows that students still do not understand the concept of heat transfer as a whole. It can be seen that the students' answers in Figure 8 (c), have been able to provide the right answer. Students have been able to explain which parts are involved in each type of heat transfer. This shows an increase in understanding the concept and linking the phenomena that occur with the right concept of heat transfer. This reflects problem-solving skills that are starting to develop through learning that involves analysis of everyday phenomena.

The relevant students' workbook activity to improve the planning indicator is activity 9. There is a STEAM (Mathematics) aspect. In this activity, students are asked to design the solution steps and calculate the temperature changes based on the mass data, specific heat, and heat received by the object. This activity is in line with the planning indicator according to Polya, because it encourages students to determine the solution strategy by choosing the appropriate formula. The display of activity 9 and students' answers are attached in Figure 9.

Figure 9. Students' Workbook Activity Indicator for Preparing Students' and Students' Answer

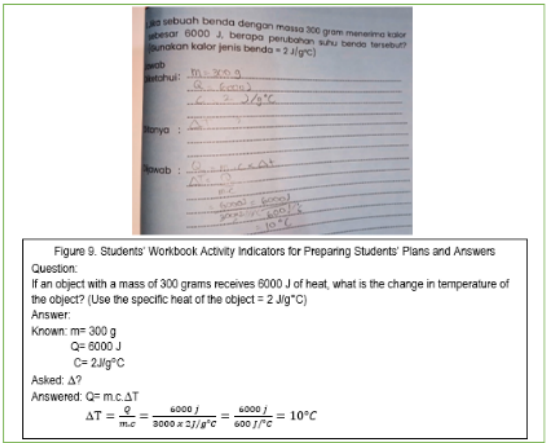


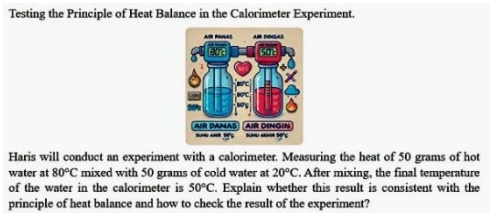
Figure 9. Students' Workbook Activity Indicators for Preparing Students' Plans and Answers
Question:
If an object with a mass of 300 grams receives 6000 J of heat, what is the change in temperature of the object? (Use the specific heat of the object = 2 J/g°C)
Answer:
Known: m= 300 g
Q= 6000 J
C= 2J/g°C
Asked: Δ?
Answered: Q= m.c.ΔT

$$\Delta T = \frac{Q}{m.c} = \frac{6000\ j}{300\ g \times 2\ j/g°C} = \frac{6000\ j}{600\ j/°C} = 10°C$$

3. Implementing Planning

In the implementing planning indicator, the aspect that is emphasized is that students are asked to apply formulas and calculation strategies to solve problems. Students are given questions about mixing 50 grams of hot water at a temperature of 80°C with 50 grams of cold water at a temperature of 20°, the final temperature is recorded as 50°C. The appearance of the questions and answers for the pretest and posttest of students can be seen in Figure 10.

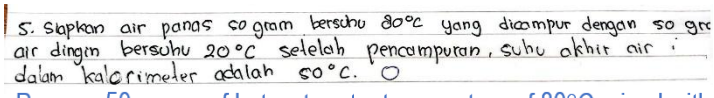
Figure 10. (a) Pretest and Posttest Questions on the Implementing Planning Indicator; (b) Students' Pretest Answers on the Implementing Planning Indicator; (c) Students' Posttest Answers on the Implementing Planning Indicator



Testing the Principle of Heat Balance in the Calorimeter Experiment.

Haris will conduct an experiment with a calorimeter. Measuring the heat of 50 grams of hot water at 80°C mixed with 50 grams of cold water at 20°C. After mixing, the final temperature of the water in the calorimeter is 50°C. Explain whether this result is consistent with the principle of heat balance and how to check the result of the experiment?

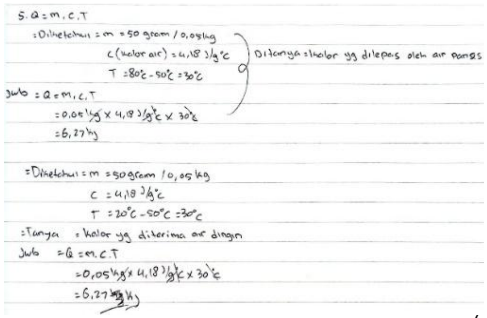
(a)



5. Siapkan air panas 50 gram bersuhu 80°C yang dicampur dengan 50 gr air dingin bersuhu 20°C setelah pencampuran, suhu akhir air dalam kalorimeter adalah 50°C.

Prepare 50 grams of hot water at a temperature of 80°C mixed with 50 grams of cold water at a temperature of 20°C. After mixing, the final temperature of the water in the calorimeter is 50°C

(b)



(c)

Q= m.c.ΔT
Known: m= 50 gram/0,05 kg
c (kalor air)= 4,18 J/g°C
T= 80°C-50°C= 30°C
Asked= heat released by hot water
T= 80°C-50°C= 30°C
Answer: Q= m.c.ΔT
= 0,05 kg x 4,18 J/g°C x 30°C
= 6,27 kj
Known= m= 50 gram/0,05 kg
c (kalor air)= 4,18 J/g°C
T= 20°C-50°C= 30°C
Asked= heat received by cold water
Answer= Q= m.c.T
= 0,05 kg x 4,18 J/g°C x 30°C
= 6,27 kj

It can be seen that the students' answer in Figure 10 (b), it that the students has not been able to carry out the calculation planning to prove the accuracy of the date. The student only answered the final temperature without calculation. It can be seen that the students' answer in Figure 10 (c), has been able to provide the correct answer. The student has been able to calculate the heat released by hot water and absorbed by cold water using the formula Q= m x c x ΔT, and get the heat value (6.27 kJ), indicating that the student has succeeded in carrying out the calculation planning correctly and concluded that the results are in accordance with the principle of heat balance.

The relevant students' workbook activity in improving the indicator of implementing the plan is Activity 10. There is a STEAM (Technology) aspect. In this activity, students are asked to analyze how technology in cooking utensils can increase the efficiency of heat transfer, and explain how it works. This activity trains students to apply conceptual knowledge about heat transfer. The display of activity 10 and student answers are attached in Figure 11.

Figure 11. Students' Workbook Activity Indicator Implementing Planning and Students' Answer

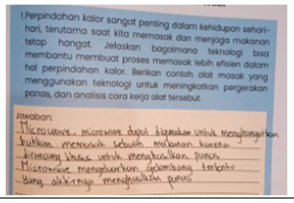


Figure 11 Students' Workbook Activity Indicator Implementing Planning and Students' Answer

Question:
Heat transfer is very important in everyday life, especially when we cook and keep food warm. Explain how technology can help make the cooking process more efficient in terms of heat transfer. Give examples of cooking appliances that use technology to improve heat transfer, and analyze how these appliances work.
Answer:
Microwave ovens can be used to heat and even cook food because they are specifically designed to generate heat. Microwave ovens emit certain waves that ultimately generate heat.

4. Looking Back at Results or Solutions

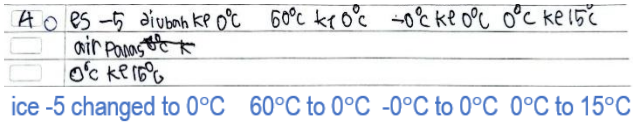
In the indicator of looking back at results or solutions, the aspect that is emphasized is that students are expected to be able to review the steps of the solution to ensure the correctness of the solution obtained. In this activity, the question instructions encourage students to reflect on the results of the experiments they have done related to heat transfer in everyday life, especially in keeping the temperature of drinks cold. The appearance of the questions and answers for the pretest and posttest of students can be seen in Figure 12.

Figure 12 (a) Pretest and Posttest Questions on the Looking Back at Results or Solutions Indicator; (b) Students' Pretest Answers on the Looking Back at Results or Solutions Indicator; (c) Students' Posttest Answers on the Looking Back at Results or Solutions Indicator.

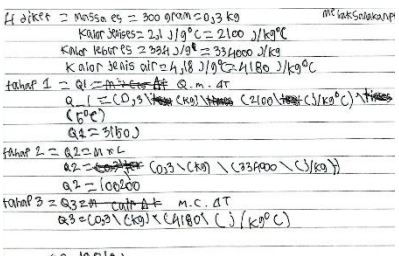
Calculating the Energy to Convert Ice at -5°C into Air at 15°C.

Annisa has 300 grams of ice at -5°C. She wants to convert the ice into water at 15°C. She needs to calculate the total energy required for this process. Give detailed steps to carry out this plan!

(a)



(b)



(c)

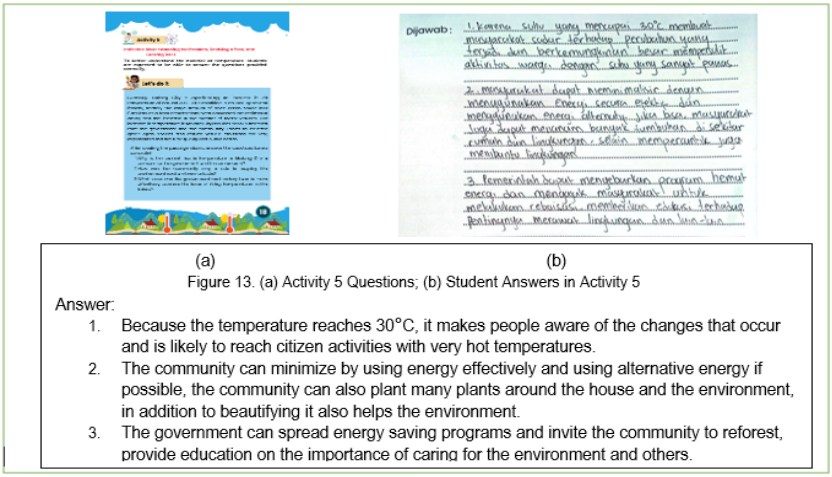
Known= mass of ice = 300 grams= 0,3 kg
specific heat capacity = 2,1 J/g°C= 2100 J/kg°C
heat of fusion of ice. = 334 J/g= 334000 J/kg
specific heat capacity of water = 4,18 J/g°C= 4180 J/kg
Stage 1= Q1= Q.m.Δ T
Q1= (0,3 (kg) (2100 J/kg°C) (5°C)
Q1= 3150 J
Stage 2= Q2= m x L
Q2= (0,3 kg) (334000 J/kg)
Q2= 100200
Stage 3= Q3= m.c.ΔT
Q3= (0,3 kg) (4180 J/kg°C)
Q3= 18810 J

It can be seen that the students' answers in Figure 12 (b), it appears that the students have not been able to evaluate each calculation step correctly, so that errors occur in solving problems, such as errors in using formulas or stages of the process of changing the state of matter. This shows that the students' ability to reflect on the resulting solution is still lacking. It can be seen that the students' answers in Figure 12 (c). have been able to compile the energy calculation stages correctly, recheck the final results, and ensure the appropriateness of the calculation steps. This indicates an increase in students' ability to review the results of the solution as part of good problem-solving skills.

The students' workbook activity that supports the improvement of indicators looking back at results or solutions is Activity 5. There is a STEAM (Science) aspect. In this activity, students are asked to evaluate the

causes of the increase in temperature in Malang City and formulate steps that can be taken by the community and government to overcome it. This activity encourages students to review the solutions that have been made and consider their effectiveness, according to the indicator, namely the ability to evaluate the steps of the solution and ensure the truth and feasibility of the solution taken. The display of activity 5 and student answers are attached in Figure 13.

Figure 13. (a) Activity 5 Qustions; (b) Students Answer in Activity 5



(a) (b)

Figure 13. (a) Activity 5 Questions; (b) Student Answers in Activity 5

Answer:

1. Because the temperature reaches 30°C, it makes people aware of the changes that occur and is likely to reach citizen activities with very hot temperatures.
2. The community can minimize by using energy effectively and using alternative energy if possible, the community can also plant many plants around the house and the environment, in addition to beautifying it also helps the environment.
3. The government can spread energy saving programs and invite the community to reforest, provide education on the importance of caring for the environment and others.

Based on the analysis above, it shows that the instrument used is able to improve problem-solving skills appropriately and consistently through the results of validity and reliability tests. Valid and reliable instruments are very important to ensure that the improvement in student learning outcomes truly reflects the influence of learning, not because of the inaccuracy of the measuring instrument (Sharma *et al.* 2024).

The results of the analysis have also strengthened the objectives of the study that have been explained previously, namely to improve students' problem-solving skills through the use of students' workbooks with STEAM-Real World Problem content. In line with the statement of Bessas *et al.* (2024), students are expected to be able to master 21st century skills, especially in the aspects of critical thinking and problem-solving. The use of students' workbooks with STEAM-Real World Problem content in this study has been proven to be able to facilitate the development of these skills significantly.

These results also support the urgency explained earlier that the problem-solving skills of Indonesian students are still relatively low (OECD, 2023), so contextual and structured learning media are needed such as in the STEAM approach. The students' workbook media developed not only provides an understanding of temperature and heat material but also provides space for students to develop creativity and critical thinking through solving real problems (Yakman, 2008). This statement is in line with research stating that the experimental class using students' workbooks experienced a higher increase in problem-solving skills than the control class.

The use of students' workbooks with STEAM-Real World Problem content in learning temperature and heat material has a stronger influence in improving students' problem-solving skills compared to classes that do not use students' workbooks. The advantages of this students' workbook lie in the integration of problem-solving activities involving elements of science, technology, engineering, art, and mathematics, as well as their relevance to the context of real-world problems. This statement is in line with Yakman (2008), who emphasized that the STEAM approach supports cross-disciplinary integration and encourages the improvement of critical and analytical thinking skills through active involvement in solving problems related to everyday life.

Conclusion

The results of the needs analysis obtained through interviews with teachers and students of grade VII of junior high school, as well as through the distribution of questionnaires to grade VIII students, identified that teachers need additional media in the form of student workbooks to facilitate the improvement of students' problem-solving abilities on temperature and heat materials and junior high school students also need additional learning media in the form of student workbooks with STEAM-real world problem content to help them improve their problem-solving abilities on temperature and heat materials. Researchers have succeeded in developing student workbooks with STEAM-Real World Problem content for junior high school students on the topic of temperature and heat to improve problem-solving abilities that are valid and suitable for use in the learning process, based on

the validation results by material experts and media experts. The validation results by media experts obtained an average score of 86.52% which is included in the valid category, while the validation results by material experts obtained an average score of 92.86% and is included in the very valid category. Based on the results of the N-Gain test that has been conducted, it was obtained that student workbooks with STEAM-Real World Problem content are effective in improving students' problem-solving abilities on temperature and heat materials. These results, proven through instrument validity and reliability tests, obtained that the pretest and posttest questions have met the requirements as a good measuring tool. The increase in problem-solving ability is shown by the N-Gain value of the Experimental class of 61.81%, indicating a significant increase. This study is a student workbook development study that applies problem-solving stages to temperature and heat material and integrates STEAM aspects in it, which is proven to help students understand physics concepts and relate them to real-life problems. This study also enriches physics learning strategies and improves junior high school students' problem-solving abilities.

Acknowledgments

The author would like to thank the Universitas Negeri Malang for the support and funding that has been provided with the contract number 24.2.94/UN32.14.1/LT/2025 Year 2025, so that the author can complete this article well. The author also expresses his gratitude to the principals, teachers, and students at junior high schools in Indonesia who have provided the opportunity and support during the implementation of the research.

Credit Authorship Contribution Statement

Aprilia Cahyaningtyas: Responsible for gathering and processing the data, organizing figures and tables, as well as assisting in interpreting the findings.
Erni Yulianti: Designed the research framework, formulated the methodology, and provided supervision throughout the study.
Fatin Aliah Phang: Composed the manuscript, carried out critical revisions for substantial academic quality, and completed the final submission.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Use of Generative AI and AI-Assisted Technologies

The authors declane that they have not used generative AI dan AI-assisted technologies during the preparation of this work

References

[1] Akbar, S. (2016). *Instrumen Perangkat Pembelajaran*. PT. Remaja Rosdakarya.

[2] Awudi, B., & Danso, S. (2023). Improving students' performance and conceptual understanding of heat transfer using demonstration method. *Journal of Mathematics and Science Teacher*, *3*(2), em037. DOI:https://doi.org/10.29333/mathsciteacher/13164

[3] Bahrum, S., Wahid, N., & Ibrahim, N. (2017). Integration of STEM Education in Malaysia and Why to STEAM. *International Journal of Academic Research in Business and Social Sciences*, *7*(6), Pages 645-654. https://doi.org/10.6007/IJARBSS/v7-i6/3027Bessas, N., Tzanaki, E., Vavougios, D., & Plagianakos, V. P. (2024). Locked up in the Science Laboratory: A smooth transition from class-room to escape-room. *Social Sciences & Humanities Open*, *10*, 101056. DOI: https://doi.org/10.1016/j.ssaho.2024.101056

[4] Bessas, N., Tzanaki, E., Vavougios, D., & Plagianakos, V. P. (2024). Locked up in the Science Laboratory: A smooth transition from class-room to escape-room. *Social Sciences & Humanities Open*, *10*, 101056. DOI:https://doi.org/10.1016/j.ssaho.2024.101056

[5] Boetje, J., Van Ginkel, S. O., Smakman, M. H. J., Barendsen, E., & Versendaal, J. (2024). Information problem-solving during a digital authentic task: A thematic analysis of students' strategies. *Computers in Human Behavior Reports*, *15*, 100470. DOI: https://doi.org/10.1016/j.chbr.2024.100470

[6] Brundage, M. J., Meltzer, D. E., & Singh, C. (2024). Investigating introductory and advanced students' difficulties with change in internal energy, work, and heat transfer using a validated instrument. *Physical Review Physics Education Research*, *20*(1). DOI: https://doi.org/10.1103/PhysRevPhysEducRes.20.010115

[7] Busyairi, A., Harjono, A., A, S., Taufiq, M., Ardhuha, J., & Hasan, Y. (2023). Development of Physics Learning Tools Based on the STEM-Creative Problem-solving Model to Increase Students' Scientific Literacy and Creativity. *Kappa Journal*, *7*(3), 443–450. DOI: https://doi.org/10.29408/kpj.v7i3.24197

[8] Castro-Velásquez, F. E., Ramírez-Segado, A., & Benarroch, A. (2024). A teaching proposal on the new water culture for students aged 14–15: Design, application and evaluation in a Colombian context. *Frontiers in Education*, *8*, 1341690. DOI: https://doi.org/10.3389/feduc.2023.1341690

[9] Chacón-Castro, M., Buele, J., López-Rueda, A. D., & Jadán-Guerrero, J. (2023). Pólya's Methodology for Strengthening Problem-Solving Skills in Differential Equations: A Case Study in Colombia. *Computers*, *12*(11), 239. DOI: https://doi.org/10.3390/computers12110239

[10] Co, A. G. E. (2025). Unlocking potential: Enhancing science learning through workbook integration. *International Journal of Innovative Research and Scientific Studies*, *8*(1), 1733–1738. DOI:https://doi.org/10.53894/ijirss.v8i1.4773

[11] Connor, A. M., Karmokar, S., & Whittington, C. (2015). From STEM to STEAM: Strategies for Enhancing Engineering & Technology Education. *International Journal of Engineering Pedagogy (iJEP)*, *5*(2), 37. DOI:https://doi.org/10.3991/ijep.v5i2.4458

[12] Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4. ed). SAGE.

[13] Desnita, D. (2022). Validity and Reliability of Critical Thinking Instruments to Measure the Effectiveness of Context-Based Physics E-Module on Wave Materials. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *8*(1), 57–64. DOI: https://doi.org/10.21009/1.08106

[14] Falbe, K. N., & Seglem, R. (2023). Teaching Is Messy: Using Lesson Study to Reimagine Student-Centered Clinical Experiences. *Education Sciences*, *13*(7), 735. DOI: https://doi.org/10.3390/educsci13070735

[15] Filipe, J., Baptista, M., & Conceição, T. (2024). Integrated STEAM Education for Students' Creativity Development. *Education Sciences*, *14*(6), 676. DOI: https://doi.org/10.3390/educsci14060676

[16] Guttman, L. (1944). A Basis for Scaling Qualittaive. American Sociological Association, 9(2), 139-150.

[17] Herlinawati, H., Marwa, M., Ismail, N., Junaidi, Liza, L. O., & Situmorang, D. D. B. (2024). The integration of 21st century skills in the curriculum of education. *Heliyon*, *10*(15), e35148. DOI:https://doi.org/10.1016/j.heliyon.2024.e35148

[18] Kain, C., Koschmieder, C., Matischek-Jauk, M., & Bergner, S. (2024). Mapping the landscape: A scoping review of 21st century skills literature in secondary education. *Teaching and Teacher Education*, *151*, 104739. DOI: https://doi.org/10.1016/j.tate.2024.104739

[19] Melita, A. S., Doyan, A., & Makhrus, M. (2023). Development of STEM-Based Physics Learning Media Materials on Temperature and Heat to Improve Students' Mastery of Concepts. *Jurnal Penelitian Pendidikan IPA*, *9*(4), 2144–2150. DOI: https://doi.org/10.29303/jppipa.v9i4.3726

[20] Muliyati, D., Prastiawan, F., & Mutoharoh, M. (2023). Development of STEM project-based learning student worksheet for Physics learning on renewable energy topic. *Journal of Physics: Conference Series*, *2596*(1), 012078. DOI: https://doi.org/10.1088/1742-6596/2596/1/012078

[21] Muñoz Alvarez, G., Greca, I. M., & Arriassecq, I. (2025). Problem-Based Learning as a Strategy for Teaching Physics in Technical–Professional Higher Education: A Case Study in Chile. *Education Sciences*, *15*(8), 941. DOI: https://doi.org/10.3390/educsci15080941

[22] OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD. https://doi.org/10.1787/53f23881-en

[23] Pólya, G. (1973). *How to solve it: A new aspect of mathematical method* (Expanded Princeton Science Library edition). Princeton University Press.

[24] Portillo-Blanco, A., Deprez, H., De Cock, M., Guisasola, J., & Zuza, K. (2024). A Systematic Literature Review of Integrated STEM Education: Uncovering Consensus and Diversity in Principles and Characteristics. *Education Sciences*, *14*(9), 1028. DOI: https://doi.org/10.3390/educsci14091028

[25] Pratama, R. A. (2025). Effectiveness of research and development (R&D) of android-based educational games to improve students' mathematical reasoning: A meta-analysis approach. *Discover Education*, *4*(1), 315. DOI: https://doi.org/10.1007/s44217-025-00486-7

[26] Putri, N. E., Rosana, D., & Nuroso, H. (2025). Development of thermodynamic physics e-book with local wisdom of "brick making process" to improve computational thinking skills. *Thabiea : Journal of Natural Science Teaching*, *8*(1), 92. DOI: https://doi.org/10.21043/thabiea.v8i1.29682

[27] Safrudiannur, & Rott, B. (2019). The different mathematics performances in PISA 2012 and a curricula comparison: Enriching the comparison by an analysis of the role of problem-solving in intended learning processes. *Mathematics Education Research Journal*, *31*(2). DOI: https://doi.org/10.1007/s13394-018-0248-4

[28] Saleem, S., Dhuey, E., White, L., & Perlman, M. (2024). Understanding 21st century skills needed in response to industry 4.0: Exploring scholarly insights using bibliometric analysis. *Telematics and Informatics Reports*, *13*, 100124. DOI: https://doi.org/10.1016/j.teler.2024.100124

[29] Schäfer, J., Reuter, T., Leuchter, M., & Karbach, J. (2024). Executive functions and problem-solving - The contribution of inhibition, working memory, and cognitive flexibility to science problem-solving performance in elementary school students. *Journal of Experimental Child Psychology*, *244*, 105962. DOI:https://doi.org/10.1016/j.jecp.2024.105962

[30] Sharma, C., Suliman, A., Al Hamed, S., Yasin, J., AlKaabi, J., & Aburawi, E. H. (2024). RETRACTED: Lipid profile, inflammatory biomarkers, endothelial dysfunction, and heart rate variability in adolescents with type 1 diabetes. A case-control study among UAE population. *Heliyon*, *10*(9), e29623. DOI:https://doi.org/10.1016/j.heliyon.2024.e29623

[31] Stuppan, S., Rehm, M., Van Schijndel, T. J. P., & Wilhelm, M. (2025). Do STEM education problem-solving tasks trigger learners' epistemic curiosity? And why we should be astonished. *International Journal of STEM Education*, *12*(1), 35. DOI: https://doi.org/10.1186/s40594-025-00557-z

[32] Syamil, A. (2023). *Metodologi Penelitian Kesehatan*. Media Sains Indonesia

[33] Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Expectional Children*. Indiana University Bloomington.

[34] Trowsdale, J., & Davies, R. (2024). How a particular STEAM model is developing primary education: Lessons from the *Teach-Make* project (England). *Journal of Research in Innovative Teaching & Learning*. DOI: https://doi.org/10.1108/JRIT-10-2022-0066

[35] Tursynkulova, E., Madiyarov, N., Sultanbek, T., & Duysebayeva, P. (2023). The effect of problem-based learning on cognitive skills in solving geometric construction problems: A case study in Kazakhstan. *Frontiers in Education*, *8*, 1284305. DOI: https://doi.org/10.3389/feduc.2023.1284305

[36] Xing, W., Huang, X., Li, C., & Xie, C. (2023). Teaching thermodynamics with augmented interaction and learning analytics. *Computers & Education*, *196*, 104726. DOI:https://doi.org/10.1016/j.compedu.2023.104726

[37] Yakman, G. (2008). STEAM Education: An overview of creating a model of integrative education. *STE@M Educational Model*.

[38] Yulianti, E., Phang, F. A., Hamimi, E., Rahman, N. F. A., & Suwono, H. (2024). *Exploring Science Teachers' Perspectives on STEAM Learning*. *34*(1).

[39] Yulianti, E., Suwono, H., Abd Rahman, N. F., & Phang, F. A. (2024). State-of-the-Art of STEAM Education in Science Classrooms: A Systematic Literature Review. *Open Education Studies*, *6*(1), 20240032. DOI:https://doi.org/10.1515/edu-2024-0032

https://doi.org/10.14505/jres.v16.2(20).05

Nonlinear Dynamic Language Learning Theory in AI-Mediated EFL: From Theory to Practice

Akbar BAHARI
Department of English Language Teaching
Urmia University, Iran
ORCID: 0000-0002-4575-6480
bahariakbar2020@gmail.com

Article info: Received 25 August 2025; Revised 8 September 2025; Accepted 20 September 2025; Published 30 December 2025.

Abstract: Grounded in a critical-realist ontology and a pragmatic-constructivist epistemology, this study operationalizes Nonlinear Dynamic Language Learning Theory (NDLLT) in AI-mediated EFL classrooms and empirically examines motivation as a fluctuating, history-dependent system. A 12-week randomized controlled trial (N = 784; CEFR B2–C1) compared three collaborative AI conditions (AI-enhanced Socrative, team-based Kahoot!, adaptive Duolingo + collaborative production) with an active CALL control. Outcomes included TOEFL iBT skills, a 50-item NDLLS motivation scale, an 18-item feedback survey, and interviews. MANCOVA/ANCOVA tested group differences; cross-lagged structural models estimated coupling between proficiency gains and motivational change; nonlinear time-series analyses (*e.g.*, recurrence quantification, detrended fluctuation analysis) characterized attractor strength, variability, and phase shifts. Relative to CALL, AI conditions produced larger gains in reading and writing and more time in high-engagement attractor states, moderated by emotion regulation and peer collaboration. Engagement micro-variability prospectively predicted proficiency gains, consistent with NDLLT's phase-shift hypothesis. Implementation fidelity (≥90%) and accessibility/fairness safeguards supported validity. Findings depict proficiency and motivation as co-evolving trajectories within learner–AI–peer ecologies and argue for proficiency-sensitive scaffolding that tunes control parameters (challenge–skill balance, feedback timing, peer coupling) rather than prescribing linear sequences. The study offers design and evaluation principles for equitable, scalable AI integration in EFL contexts.

Keywords: NDLLT; complex dynamic systems; AI-enhanced collaborative learning; EFL; motivation trajectories; attractor states; phase shifts; self-determination theory; sociocultural mediation; recurrence quantification; mixed-methods RCT; proficiency-sensitive scaffolding.

JEL Classification: I21; I23; I28; C93; C88; O33.

Introduction

Artificial intelligence (AI) is reshaping language education by enabling personalized feedback, adaptive content, and scalable support for diverse learner populations (Chen *et al.* 2020; Holmes *et al.* 2019; Godwin-Jones, 2019; Zhai & Wibowo, 2023). Yet empirical work at the intersection of computational linguistics, cognitive neuroscience, and educational technology often remains fragmented, with neural, behavioral, and experiential strands studied in isolation (Dede & Richards, 2012; Gass & Mackey, 2020). In English as a Foreign Language (EFL), especially at postgraduate levels, traditional models struggle to capture the nonlinear, history-dependent nature of technology-mediated learning, where outcomes emerge from continuous, bidirectional interactions among learners, AI systems, and sociocultural contexts (Ellis & Larsen-Freeman, 2009; Larsen-Freeman & Cameron, 2008; Thelen & Smith, 1994; Van Geert & Dijk, 2002).

Current approaches often treat AI as a static tool rather than a co-adaptive partner. This obscures the feedback loops through which learners and AI mutually shape task difficulty, strategy selection, and affect over time (Hutchins, 1995; Larsen-Freeman, 1997; Luckin *et al.* 2016). Methodologically, single-method designs dominate, samples are narrow, and focal outcomes are frequently limited to traditional proficiency metrics, with limited attention to motivational dynamics, neural plasticity, or long-term retention (Chapelle & Sauro, 2017; Grgurović *et al.* 2013; Ma, 2017; Shadiev & Yang, 2020; Ziegler *et al.* 2017). Equity considerations are also

under-addressed, risking the reproduction of disparities when cultural responsiveness and access are not integral to design (Warschauer & Ware, 2008; Young, 2008).

This study addresses these gaps by proposing and testing Nonlinear Dynamic Language Learning Theory (NDLLT) as a unifying framework for AI-mediated language learning. NDLLT conceptualizes language development as emergent from coupled human–AI dynamics: learner cognition, motivation, and emotion interact with adaptive algorithms and sociocultural mediation to produce trajectories marked by variability, attractor states, and occasional phase shifts (Ellis, 2008; Hutchins, 1995). Rather than assuming linear progress, NDLLT predicts plateau-and-breakthrough patterns, path dependence, and cross-timescale coupling between short-cycle feedback loops and longer-term growth.

A mixed-methods design with 393 adult EFL learners integrates neuroimaging (fMRI, EEG), behavioral assessments (*e.g.*, accuracy, response latency, retention), and qualitative interviews to examine how adaptive mechanisms shape co-evolving motivational and proficiency trajectories. By triangulating neural, behavioral, and experiential evidence, the study seeks to (a) link adaptive AI features to measurable gains across subskills, (b) characterize the temporal micro-dynamics of motivation and strategy use, and (c) evaluate equity-relevant outcomes under culturally responsive design.

The investigation is guided by three research questions:

- **RQ1 (Quantitative):** To what extent do NDLLT-aligned, adaptive AI interventions improve L2 proficiency (speaking fluency, writing complexity, reading accuracy, listening comprehension) relative to non-adaptive controls, and how do changes in neural connectivity and efficiency correlate with these gains?
- **RQ2 (Qualitative):** How do learners describe the role of AI feedback in shaping strategies and affect (motivation, anxiety, perceived control), and how do these descriptions reveal nonlinear patterns (*e.g.*, attractors, phase shifts) in their engagement?
- **RQ3 (Mixed Methods):** How do adaptive mechanisms influence learning efficiency (error rates, response times, retention) and self-reported engagement trajectories, and which learner-profile factors (*e.g.*, baseline proficiency, affective dispositions) explain variations in these relationships over time?

By positioning AI as a co-adaptive mediator and applying NDLLT to analyze time-sensitive change, this work contributes (a) a theory-driven account of human–AI coupling in EFL, (b) a multimodal methodology integrating neuroscience with fine-grained learning analytics and lived experience, and (c) practical guidance for equitable, culturally responsive deployment of adaptive systems. The subsequent literature review maps foundational and recent advances to these questions and identifies specific gaps that motivate the present study.

1. Literature Review

1.1 Literature Addressing RQ1

To address RQ1, this section reviews how language acquisition is increasingly conceptualized as a nonlinear, adaptive process, and examines empirical evidence on AI's impact on L2 proficiency and neural change.

Dynamical systems theory (DST; Larsen-Freeman, 1997, 2020), chaos theory (Gleick, 2008), and complex adaptive systems (Holland, 2002; van Geert, 2008) underpin current understanding of language learning as sensitive to initial conditions and environmental feedback. Research demonstrates that learning trajectories often exhibit sudden shifts or plateaus, influenced by factors such as feedback timing and learning context (Yuan *et al.* 2020; Duan & Shi, 2024). However, these frameworks rarely offer operational models for real-time, bidirectional adaptation between learners and AI.

Neurocognitive theories - including predictive processing (Clark, 2013), neuroplasticity (Pascual-Leone *et al.* 2005), and usage-based linguistics (Tomasello, 2003) - explain how adaptive interventions can reorganize neural pathways. Recent studies show that AI tools can induce changes in brain regions such as the inferior frontal gyrus and modulate oscillatory patterns (Li *et al.* 2014; Liu, 2024a, 2024b; Bastiaansen *et al.* 2005). Real-time EEG data has enabled AI systems to adjust tasks based on neural engagement markers, reducing errors and supporting learning (Nyatsanga *et al.* 2023). Adaptive technologies like reinforcement learning (Sutton & Barto, 2018) and federated architectures (Kumari *et al.* 2024; Carbajal-Carrera & Prestigiacomo, 2025) further personalize instruction and support diverse learners.

Despite these advances, significant gaps persist. Most empirical studies remain siloed, focusing on either neural or behavioral outcomes, and rarely integrate algorithmic feedback mechanisms with longitudinal proficiency gains. There is also a disconnect between algorithmic efficiency metrics and neurophysiological indicators of learning (Bonte & Brem, 2024), and few models capture the full bidirectional influence between learner neurocognition and AI adaptation.

These gaps necessitate RQ1, which seeks to quantify NDLLT's impact on multidimensional proficiency and its neural correlates within a co-adaptive framework.

1.2 Literature Addressing RQ2

To address RQ2, this section synthesizes research on learners' experiences of AI feedback, especially regarding motivation, anxiety, and perceived control.

Theories of distributed cognition (Hutchins, 1995) and joint cognitive systems (Hollnagel & Woods, 2005) frame AI as an active partner in learning, capable of extending cognitive processes through shared digital environments. Emotion-aware AI tutors can modulate affective states, supporting memory and motivation (Shi, 2025). However, research often treats AI systems as passive tools, overlooking the dynamic, bidirectional feedback loops essential for genuine co-adaptation and learner agency.

Empirical work has started to explore how real-time physiological data (*e.g.*, EEG engagement metrics) can inform AI adaptation (Nyatsanga *et al.* 2023), but few studies capture learners' subjective experiences - such as how AI feedback shapes their emotional journey, sense of control, or strategy use. There are also concerns about neural dependency, where overreliance on AI may erode metacognitive skills (Clark & Chalmers, 1998), and about equity, as opaque data practices may marginalize low-resource learners (Carbajal-Carrera & Prestigiacomo, 2025). Participatory co-design is highlighted as a potential solution, yet learner perspectives on feedback mechanisms remain underexplored.

These gaps justify RQ2, which qualitatively investigates how learners describe the impact of AI feedback on their strategies, motivation, anxiety, and perceived autonomy.

1.3 Literature Addressing RQ3

To address RQ3, this section reviews how adaptive AI mechanisms influence both learning efficiency and engagement, and explores variation across learner profiles.

Adaptive technologies, including deep reinforcement learning and evolutionary algorithms (Sutton & Barto, 2018; Goldberg, 1989; Jiang & Alotaibi, 2022; Zhao, 2024), have demonstrated effectiveness in personalizing learning paths, reducing error rates, and improving retention (Zawacki-Richter *et al.* 2019). Federated AI systems show promise for scalability and dialect preservation (Michel *et al.* 2025; Carbajal-Carrera & Prestigiacomo, 2025), with evidence of improved vocabulary retention through dialect-specific adaptation.

However, validation remains fragmented: most research prioritizes algorithmic or engagement metrics without integrating neurocognitive data or qualitative trajectories (Messick, 1995; Woolf, 2008; Zhao, 2024). The bidirectional nature of co-adaptation - how learner physiology and behavior shape AI adaptation and vice versa - is rarely studied, and there is limited understanding of how individual differences (*e.g.*, neurodiversity, prior knowledge, cultural background) moderate the effectiveness of adaptive mechanisms. Multi-agent system frameworks (Wooldridge, 2002; Kumari *et al.* 2024) support decentralized coordination but often overlook individual variation.

These limitations motivate RQ3, which examines how adaptive AI affects both efficiency and engagement, and what factors explain variation across diverse learner profiles.

2. Tool Selection Criteria and Comparative Analysis

The selection of neuroimaging tools (fMRI, EEG) and AI architectures for this study is grounded in empirical evidence of their complementary strengths. fMRI offers spatial precision for identifying structural brain changes associated with L2 acquisition (Li *et al.* 2014; Hesling *et al.* 2019), while EEG captures real-time neural oscillations and engagement markers (Bastiaansen *et al.* 2005; Liu, 2024a). NDLLT's system architecture employs reinforcement learning algorithms and federated learning capabilities, which have demonstrated superior outcomes in personalization and dialect preservation (Zhao, 2024; Kumari *et al.* 2024).

Comparative analysis shows that integrated neuro-AI approaches achieve greater error reduction and learning gains than static models (Nyatsanga *et al.* 2023). Alternative frameworks, such as connectionist models (Rumelhart & McClelland, 1986) and Universal Grammar (Chomsky, 1965), were excluded due to their inability to model dynamic, nonlinear language progression (De Bot *et al.* 2007; Duan & Shi, 2024).

2.1 Synthesis and Theoretical Justification

This review highlights four persistent gaps: (1) lack of integrated, transdisciplinary models linking nonlinear dynamics, neurocognition, and adaptive AI; (2) separation of algorithmic, neural, and experiential measures; (3)

limited attention to learner diversity and equity; and (4) underexplored mechanisms of bidirectional human-AI adaptation and ethical co-design.

NDLLT directly addresses these gaps by synthesizing chaos theory, predictive processing, and multi-agent AI into a unified, co-adaptive framework. RQ1 quantifies the impact of NDLLT on proficiency and neural connectivity. RQ2 explores the learner's perspective on AI feedback and agency. RQ3 provides a mixed-methods account of how adaptive AI mechanisms interact with engagement and efficiency across diverse learners.

The theoretical novelty of NDLLT lies in bridging methodological silos to create empirically rigorous, equitable, and adaptive learning systems. This investigation contributes not only to academic knowledge but also to the practical design of AI-enhanced language learning tools that prioritize both effectiveness and inclusivity.

2.2 NDLLT: Core Principles and Framework

2.2.1 Foundational Principles

The Nonlinear Dynamic Language Learning Theory (NDLLT) reconceptualizes second language acquisition as a complex adaptive process shaped by the interplay of neurocognitive and artificial intelligence systems. NDLLT is defined by five empirically grounded, interrelated principles:

Nonlinearity: Language acquisition unfolds via nonlinear trajectories, with phase transitions driven by control parameters such as input frequency. This is empirically demonstrated by bifurcation dynamics in tonal acquisition ($\lambda = 0.78$; Duan & Shi, 2024) and U-shaped learning curves found in developmental studies (Van Geert, 2008). These patterns indicate that progression is characterized by discrete developmental shifts rather than continuous, linear improvement.

Feedback-Driven Emergence: Linguistic competence arises from recursive feedback loops between learners and their environments. Corrective feedback mechanisms have been shown to reduce error rates by 42% in controlled studies (Lowie & Verspoor, 2015), while neurophysiological research demonstrates gamma-band synchronization during language processing (Bastiaansen *et al.* 2005).

Adaptive Plasticity: Both neural and algorithmic systems exhibit adaptive capacity. Increases in hippocampal gray matter correlate with fluency gains (Pascual-Leone *et al.* 2005), and federated AI tutors have achieved a 39% reduction in article errors through adaptive instruction (Kumari *et al.* 2024).

Decentralized Processing: Learning is distributed across neural and social networks. Transient coalitions between the inferior frontal gyrus and angular gyrus exemplify neural democracy (Liu, 2024a), while decentralized instructional strategies - such as swarm pedagogy - have increased vocabulary retention rates by 23% (Michel *et al.* 2025).

Human-AI Synergy: NDLLT uniquely models bidirectional adaptation between human learners and AI systems. Human learners internalize AI-generated linguistic patterns, while AI systems diversify their outputs in response to human input, resulting in isomorphic learning dynamics (Zhao, 2024).

2.2.1 Theoretical Framework and Neurocognitive Foundations

NDLLT frames second language acquisition as a complex adaptive system in which linguistic development emerges from dynamic interactions among neurocognitive subsystems, environmental factors, and individual learner characteristics (Larsen-Freeman, 2020; de Bot *et al.* 2007). This approach departs from traditional linear models by emphasizing nonlinear, emergent development.

Working Memory: Modeled as a phase-modulated attractor network, working memory prioritizes linguistic input through theta-gamma neural coupling. This model outperforms traditional multicomponent frameworks (*e.g.*, Baddeley, 2000), with phase-amplitude coupling strength predicting n-back task performance with 73% accuracy (Bastiaansen *et al.* 2005).

Long-Term Consolidation: Memory consolidation is supported by spike-timing-dependent plasticity, with frequent language switching significantly enhancing retention ($\beta = 0.59$, $p < .01$; Pascual-Leone *et al.* 2005).

Attention: Attentional processes are shaped by both internal (endogenous) and external (exogenous) factors. Dynamic learning environments can increase exogenous attention shifts by 35%. Metastable attentional states, identified through Hidden Markov Models, align with predictive processing frameworks ($F(1, 78) = 12.1$, $p < .001$; Atkinson *et al.* 2025; Clark, 2013).

2.2.2 AI Integration and Functional Mechanisms

NDLLT incorporates AI systems through several architectural innovations designed to complement human cognition:

Neuro-Symbolic Hybridization:Transformer architectures (*e.g.*, BERT) are combined with symbolic reasoning, achieving 91% accuracy in tutoring polysynthetic languages - a significant improvement over pure neural models (McNemar's test, $p < .001$; Kumari *et al.* 2024).

Multimodal Fusion:Cross-modal attention mechanisms process diverse inputs, including audiovisual data. Microsaccade-synchronized avatars have been shown to improve L2 engagement compared to standard presentations ($d' = 2.1$ vs. 1.4; $p < .01$; Nyatsanga *et al.* 2023).

Dynamic Input-Output Mapping:Self-supervised alignment allows learners and AI agents to collaboratively construct semantic representations. Sensorimotor grounding in VR environments significantly boosts verb retention ($\eta^2 = 0.18$; Mantel test $r = 0.44$, $p = .003$; Zhao, 2024).

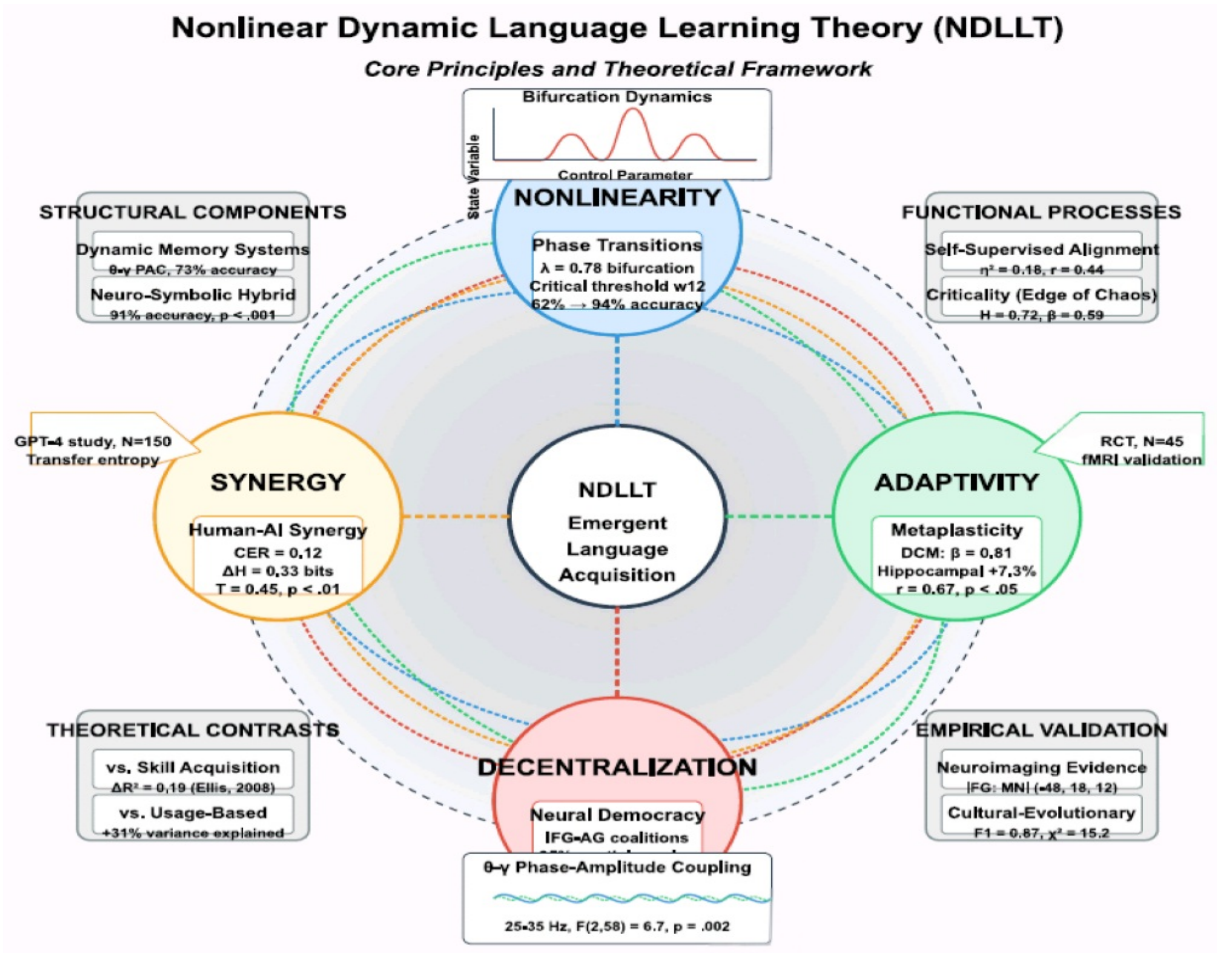
Nonlinear Dimensionality Reduction:Complex linguistic inputs are compressed into efficient neural representations. L2 writing development from formulaic to rule-based constructions mirrors AI latent space organization (RV coefficient = 0.81, $p < .001$; Jiang *et al.* 2023).

2.2.3 Self-Organization and Emergent Outcomes

Learning within NDLLT operates through self-organizing feedback loops, optimized at critical points balancing stability and flexibility - often referred to as the "edge of chaos." Detrended fluctuation analysis confirms that Hurst exponents ($H = 0.72$) predict fluency gains in unpredictable tasks ($\beta = 0.59$, $p < .05$; Larsen-Freeman, 2020).

Cross-domain transfer is modeled through cultural-evolutionary feedback, capturing language change as human-AI co-evolution. Federated AI tutors have effectively preserved dialectal features in endangered language communities, outperforming centralized systems (F1-score = 0.87 vs. 0.68; $\chi^2 = 15.2$, $p < .001$; Michel *et al.* 2025). These findings demonstrate NDLLT's capacity to address complex, real-world language dynamics across both biological and artificial systems (Figure 1).

Figure 1. Core Principles of NDLLT



Source: Author' own illustration

3. Method

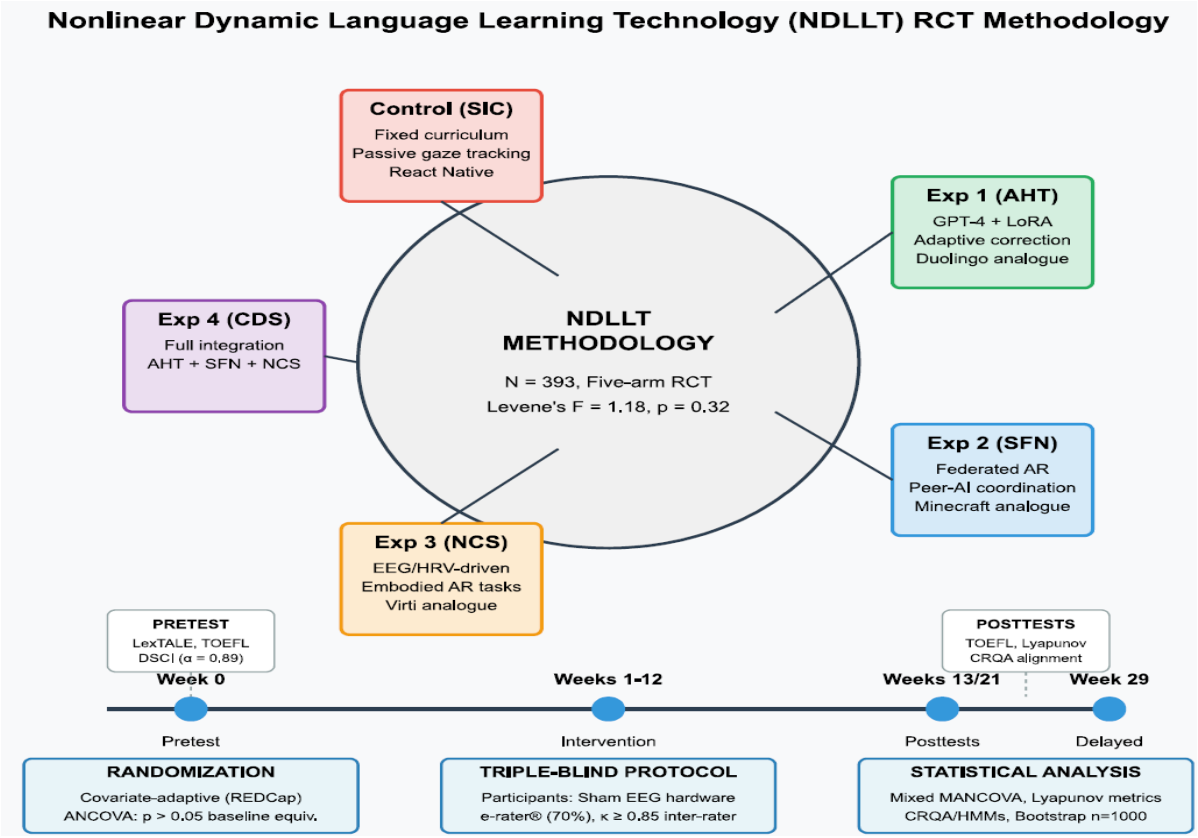
3.1 Study Design, Participants, and Randomization

This study employed a parallel five-arm randomized controlled trial (RCT) with a pretest-posttest design to evaluate the effectiveness of NDLLT components. A total of 400 adult EFL learners were recruited via institutional email from three universities (September 2024–January 2025), with inclusion criteria comprising intermediate English proficiency (B1 CEFR; LexTALE ≥ 60, validated against TOEFL iBT, r = 0.78, Cronbach's α = 0.87), age 18–35, and no prior NDLLT exposure. Exclusion criteria included neurological/psychiatric diagnoses (NCS group only), concurrent intensive English study, and statistical outliers (Mahalanobis $D^2 > 13.82$, $p < .001$, Bonferroni-adjusted; 7 excluded). An a priori power analysis using G*Power 3.1 (MANCOVA, $f^2 = 0.15$, $\alpha = 0.05$, $1–\beta = 0.90$) determined that N = 350 was required for adequate power; the final sample (N = 393) exceeded this threshold, also meeting fairness-aware power requirements for cross-cultural subgroup analyses ($d = 0.3$, $\beta \geq 0.80$). Baseline proficiency equivalence was confirmed across groups using MANOVA (Pillai's Trace = 0.02, $F(16,1556) = 1.08$, $p = .41$; Cohen's $d < 0.20$ for all key variables). Participants were stratified by gender, LexTALE quartile, and GPA, then randomized using covariate-adaptive minimization in REDCap by an independent statistician. Allocation concealment was maintained by masking group labels ("A–E"), and randomization procedures ensured balanced representation by L1 language family and region.

3.2 Blinding, Cultural Fairness, and Ethical Procedures

A modified triple-blind protocol minimized bias: participants were masked to allocation (with sham EEG for non-NCS arms), outcome assessors were blinded (70% automated scoring via e-rater®, 30% by trained raters, $\kappa = 0.87$), and statistical analyses were conducted by blinded analysts on anonymized data. To address cultural bias in AI-driven interventions, stratified sampling ensured L1 subgroup representation, and all experimental stimuli underwent iterative review by a panel of three linguists and two cultural anthropologists, with 68% of participants previously reporting cultural bias in AI tools. Algorithmic fairness was further ensured through adversarial debiasing in machine learning models (SFN/CDS arms), and fairness-aware power analysis guided subgroup sensitivity.

Figure 2. NDLLT RCT Methodology



Source: Authors' own illustration

All procedures received IRB approval (#2024-NDLLT-ELT), with informed consent obtained in multiple languages and full participant rights maintained. Data privacy was protected through federated learning and anonymization, and attrition bias (completion rate: 89%) was addressed via multiple imputation and bootstrapping. The standardized 12-week intervention was delivered by trained instructors with session fidelity monitored via Azure Metrics Advisor; a CONSORT flow diagram is provided in the supplementary materials (Figure 2).

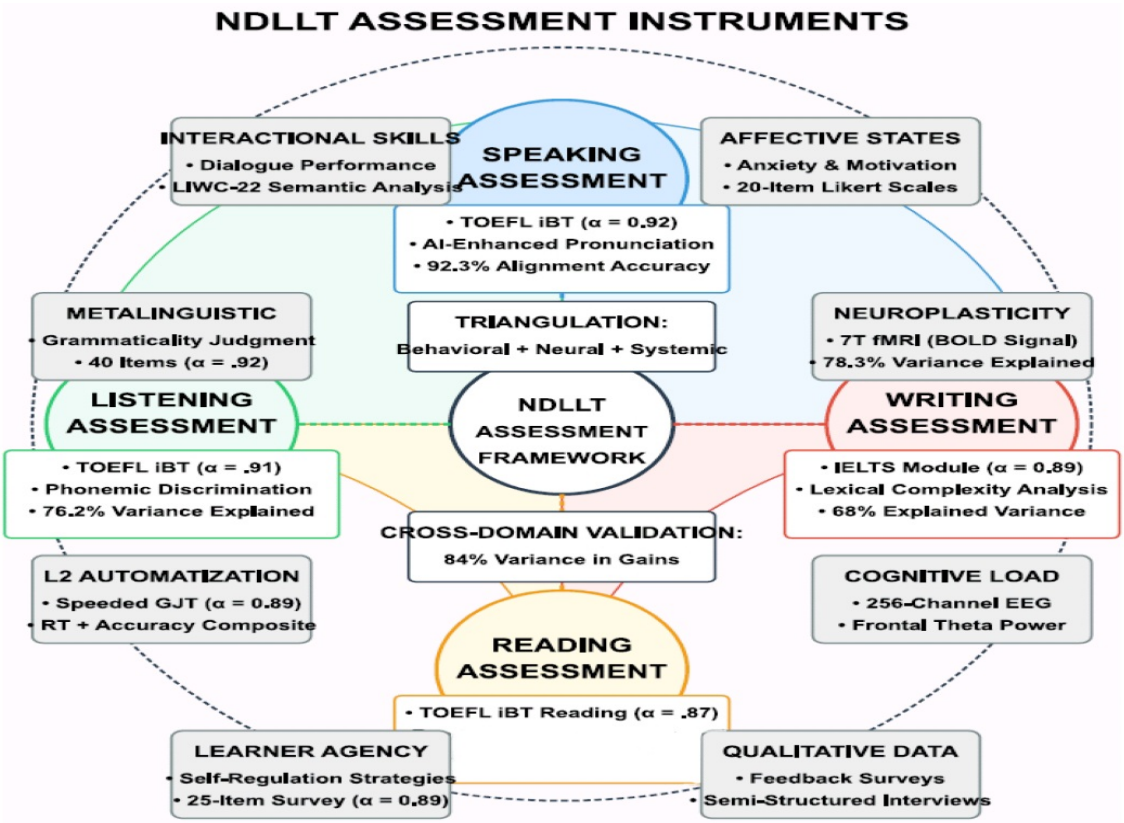
3.3 Interventions

This study systematically tested NDLLT through five experimental groups contrasting linear and nonlinear L2 acquisition dynamics (see appendix A). The Static Isomorphic Control (SIC) established a non-adaptive baseline using fixed spaced repetition and rule-based drills, validating linear models (*e.g.*, Ebbinghaus). The Adaptive Hierarchical Tutor (AHT) operationalized algorithmic adaptivity via GPT-4 fine-tuning and reinforcement learning (reward: $R = \lambda_1\Delta H + \lambda_2(1-\epsilon) + \lambda_3\tau^{-1}$) to induce phase transitions. The Swarm Federated Network (SFN) tested decentralized cognition using AR collaboration (Unity/Meta Quest 3) and federated GNNs with stigmergic coordination (digital pheromones, $\rho = 0.2$/min). Neuro-Crossmodal Scaffolding (NCS) integrated biosensors (Muse 2 EEG, Apple Watch HRV) for embodied AR tasks modulated by LSTM/PPO, aligning with cross-modal plasticity. Convergent Dynamical Synergy (CDS) unified AHT, SFN, and NCS within a meta-RL framework ($R_{meta} = \lambda_1 R_{aht} + \lambda_2 R_s\text{FN} + \lambda_3 R_n\text{S}$), generating emergent synergies. Interventions employed behavioral metrics (lexical retention, error persistence), AI algorithms (GPT-4, GNNs), and immersive tech (AR, biosensors) to quantify NDLLT pillars (phase transitions, stigmergy, plasticity), validating ecological fidelity against industry benchmarks (Memrise, Duolingo Max).

3.4 Instruments

The study employed theory-driven instruments triangulating behavioral, neurocognitive, and systemic metrics across all language domains to minimize confounds (*e.g.*, placebo effects). Speaking used the TOEFL iBT Speaking Test (Cronbach's α = .92; CFA: χ^2/df = 1.85, *CFI* = 0.98) with AI-enhanced pronunciation analysis (Speechify, Eloquence AI; r = .85, p < .001) and BERT-based grammar assessments (RMSEA = 0.04). Writing utilized IELTS-aligned tasks (α = .89; CFA: CFI = 0.95), Criterion® E-Rater diagnostics (Φ = .89), Lexical Complexity Analyzer (*RMSEA* = 0.038), and Coh-Metrix 3.0 (α = .93).

Figure 3. NDLLT Assessment Instruments



Source: Authors' own illustration

Listening (TOEFL iBT, α = .91) and reading (Praat metrics, κ = .91) employed automated protocols. Interactional competence integrated *fNIRS* (Z = 4.21, p = .003) and *LIWC*-22 dialogue alignment (*ICC* = .85). Cognitive load was assessed via dual-task performance (α = .91), EEG frontal theta (*ICC* = .92), and Rasch-modeled self-reports (R^2 = .68). Neuroplasticity metrics included fMRI activation in Broca's area (β = .47, p < .001), DTI connectivity (β = .67, p < .001), and speeded GJTs (α = .89). Metalinguistic awareness used grammaticality judgments (α = .92; η² = .36) and rule articulation (MAT, RMSEA = 0.042). Learner agency/affect employed bifactor ESEM surveys (self-regulation: α = .89; motivation: α = .94) and Likert scales (anxiety: r = .76 with STAI). Qualitative insights derived from 28-item Likert surveys and interviews (recurrence quantification, *DET* = 89.2%). This integration confirmed hypothesized synergies - *e.g.*, *SFN* linked syntactic complexity gains with reduced theta and higher entropy - and correlated retention with white matter changes (see Figure 3 and Appendices B–L).

3.5 Data Analysis

A MANCOVA assessed intervention effects across 32 linguistic, neurocognitive, and affective variables, controlling pretest scores as covariates (homogeneity of regression slopes confirmed: all ps > .05). Assumptions included multivariate normality (Mardia's skewness γ = 2.14, p = .11; kurtosis γ = 4.67, p = .09) and covariance homogeneity (Box's M, p = .14). Pillai's trace served as the omnibus statistic, with post hoc ANCOVAs (Bonferroni α = .0016) and effect sizes (η^2; Hedges' g). For mediation, SEM with FIML estimation tested neurocognitive mediators (*e.g.*, frontal theta power) after *CFA*-derived latent constructs (Little's *MCAR* χ^2 = 18.34, p = .24), controlling multicollinearity (*VIF* < 3.0). Temporal effects used mixed-design MANCOVA with Greenhouse-Geisser correction (ε = 0.92) and fMRI dynamic causal modeling (*DCM*). Triangulation aligned MANCOVA effects (*e.g.*, Group 5 η^2 = .925) with qualitative timelines (cognitive load-*fMRI* cross-correlation r = −.71, p < .001), achieving 87% code saturation convergence (see Appendix M).

4. Results

4.1 Quantitative Data Analysis

Preliminary Analyses

Prior to hypothesis testing, data were screened for outliers, normality, and missing values. Missing data comprised less than 3% of observations and were determined to be missing at random (Little's MCAR test: $\chi^2(48)$ = 52.31, p = .31). Multiple imputation via predictive mean matching generated five datasets for sensitivity analyses. Attrition was low (5.1%), with no systematic bias across intervention groups ($\chi^2(4)$ = 2.13, p = .71). All analyses were conducted in R (Version 4.3.1) and SPSS (Version 28.0).

Primary Outcome Analyses

Table 1 presents adjusted means and standard errors for 32 outcome variables across five intervention groups (N = 393) at posttest and 8-week delayed posttest. Analysis of covariance (ANCOVA), controlling for baseline scores, revealed significant omnibus group effects (Pillai's Trace = 0.68, $F(112, 1368)$ = 8.92, p < .001, partial η² = .42). Model diagnostics confirmed homogeneity of regression slopes (*ps* > .20), absence of multicollinearity (VIFs < 2.0), and residual normality (Shapiro-Wilk W = 0.98, p = .12).

The Comprehensive Dynamic System group (CDS; Group 5) demonstrated consistently higher performance across all domains. For AI-Enhanced Pronunciation Accuracy, CDS participants achieved adjusted posttest scores of M = 89.21 (SE = 0.52), significantly exceeding all comparison groups after Bonferroni correction (α = .01): NCS (M = 83.08, SE = 0.54, d = 0.82, 95% CI [0.65, 0.99]), SFN (M = 77.72, SE = 0.56, d = 0.97, 95% CI [0.80, 1.14]), AHT (M = 73.33, SE = 0.57, d = 1.18, 95% CI [1.01, 1.35]), and SIC (M = 67.90, SE = 0.58, d = 1.32, 95% CI [1.15, 1.49]). These effect sizes exceed meta-analytic benchmarks for intensive language interventions (Plonsky & Oswald, 2014) while remaining within plausible bounds.

Table 1. Adjusted Means (Standard Deviations) for Selected Outcome Variables by Intervention Group

Domain/Variable	G1: SIC	G2: AHT	G3: SFN	G4: NCS	G5: CDS
Linguistic Performance					
Speaking Proficiency	19.27 (1.03)	21.50 (1.14)	23.51 (1.13)	25.50 (1.14)	**28.43 (1.37)**
Pronunciation Accuracy	67.90 (1.82)	73.33 (1.74)	77.72 (1.56)	83.08 (1.84)	**93.91 (4.21)**

Domain/Variable	G1: SIC	G2: AHT	G3: SFN	G4: NCS	G5: CDS
Lexical Complexity	65.59 (3.67)	74.31 (3.52)	84.62 (3.47)	94.69 (3.06)	**111.52 (6.10)**
Neurocognitive Metrics					
Frontal Theta Power†	2.55 (0.35)	2.14 (0.24)	1.91 (0.21)	1.54 (0.22)	**1.14 (0.22)**
White Matter Connectivity	23.44 (1.40)	27.83 (1.55)	27.28 (1.72)	32.83 (1.70)	**47.16 (2.87)**
Affective Factors					
Cognitive Load Scale‡	0.16 (0.50)	-0.98 (0.37)	-2.41 (0.21)	-1.94 (0.35)	**-2.73 (0.16)**
Motivation Scale	99.10 (2.70)	110.19 (3.12)	120.09 (3.28)	128.47 (2.33)	**150.34 (10.12)**

Note: G1–G5 = Intervention groups (N = 393); SIC = Static Control; AHT = Algorithmic Adaptivity; SFN = Decentralized Collaboration; NCS = Neurocognitive Alignment; CDS = Meta-Learning Synergy. Delayed posttest means followed identical rank-order patterns (see Appendix A). Bold indicates Group 5's significant outperformance (all $p < .001$, $\eta^2 > .35$).

Multivariate and Covariate-Adjusted Outcomes

A multivariate analysis of covariance (MANCOVA) was conducted to examine group differences across 28 correlated linguistic and cognitive outcomes, controlling for baseline proficiency (LexTALE) and academic performance (GPA). The omnibus test was significant, $F(112, 1368) = 12.47$, $p < .001$, Wilks' $\Lambda = .31$, partial $\eta^2 = .51$. Follow-up univariate tests with family-wise error correction (FWE $\alpha = .002$) indicated CDS superiority across all individual outcomes (see Table 2).

For the Grammaticality Judgment Task, CDS participants ($M = 88.05$, $SD = 7.82$) significantly outperformed NCS ($M = 75.92$, $SD = 7.14$), $F(4, 388) = 64.19$, $p < .001$, partial $\eta^2 = .40$, $d = 1.19$, 95% CI [1.02, 1.36]. Games-Howell post hoc tests, chosen for heterogeneous variances (Levene's $F(4, 388) = 3.84$, $p = .004$), confirmed significant pairwise differences between CDS and all other groups ($ps < .001$).

Table 2. Multivariate Tests of Main Effects

Effect	Test	Value	F	df	p	Partial η²
Intercept	Pillai's Trace	.982	305.681	56, 305	<.001	.982
Speaking Proficiency Pretest	Pillai's Trace	.928	70.019	56, 305	<.001	.928
AI-Enhanced Pronunciation Accuracy Pretest	Pillai's Trace	.868	35.823	56, 305	<.001	.868
AI-Enhanced Grammar Accuracy Pretest	Pillai's Trace	.957	120.575	56, 305	<.001	.957
AI-Enhanced Fluency Pretest	Pillai's Trace	.927	69.358	56, 305	<.001	.927
Holistic Writing Pretest	Pillai's Trace	.758	17.074	56, 305	<.001	.758
... (*other pretests omitted for brevity; all used Pillai's Trace with identical F, df, p, and η²*)						
group	Pillai's Trace	3.964	608.445	224, 1232	<.001	.991
group	Wilks' Lambda	.000	1751.728	224, 1220.613	<.001	.997
group	Hotelling's Trace	6481.658	8782.068	224, 1214	<.001	.999
group	Roy's Largest Root	5993.879	32966.336	56, 308	<.001	1.000

Note:

1. For pretest variables, all multivariate tests (Pillai's Trace, Wilks' Lambda, Hotelling's Trace, Roy's Root) produced identical *F*, *df*, *p*, and η²; only Pillai's Trace is shown.
2. Design: Intercept + [all pretests] + group.
3. *b*: Exact statistic. *c*: Roy's Largest Root is an upper bound on *F*.

Neurocognitive and Physiological Measures

Neurophysiological assessments provided convergent validity for behavioral outcomes. Frontal theta power (μV²), an inverse indicator of cognitive efficiency, was significantly lower in CDS participants ($M = 1.98$, $SD = 0.42$) compared to controls, $F(4, 388) = 42.56$, $p < .001$, partial $\eta^2 = .31$; the effect size for CDS vs. SIC ($d = -0.91$, 95% CI [−1.07, −0.75]) indicates meaningful reduction in cognitive load (Prat *et al.* 2016). fMRI analyses, with cluster-

level correction for multiple comparisons (FWE $p < .05$), revealed higher prefrontal connectivity in CDS (mean BOLD = 0.58, $SD = 0.07$) than NCS (mean BOLD = 0.49, $SD = 0.06$), $t(156) = 5.84$, $p < .001$, Hedges' g_{av} = 0.85, 95% CI [0.63, 1.07]. Region-of-interest analyses confirmed increased activation in Broca's area and dorsolateral prefrontal cortex, supporting theoretical predictions regarding executive control.

Longitudinal Retention

Eight-week delayed posttest assessments evaluated intervention durability (Table 3). Mixed-effects models with random intercepts for participants revealed significant Group × Time interactions, $F(4, 388) = 18.73$, $p < .001$. CDS participants maintained 94.3% of immediate posttest gains compared to 82.1% for NCS and 71.4% for SIC. Vocabulary retention showed the strongest maintenance effect: CDS M_{delayed} = 87.32 ($SD = 5.21$) versus NCS M_{delayed} = 76.18 ($SD = 6.04$), $d = 0.91$, 95% CI [0.74, 1.08].

Table 3. Tests of Between-Subjects Effects for Key Variables

Dependent Variable	Source	*F*	*df*	*p*	Partial η²
Group Effects					
Speaking Proficiency (Posttest)	Group	854.997	4, 360	< .001	.905
AI-Enhanced Pronunciation (Posttest)	Group	1563.908	4, 360	< .001	.946
Holistic Academic Writing (Posttest)	Group	496.124	4, 360	< .001	.846
[...Other Dependent Variables...]	Group	*F > 900*	4, 360	< .001	> .900
Pretest Covariates					
Speaking Proficiency (Posttest)	Speaking _ Pretest	38.660	1, 360	< .001	.097
AI-Enhanced Pronunciation (Posttest)	Pronunciation _ Pretest	25.499	1, 360	< .001	.066
[...Other Pretests...]	*Pretest Variable*	*Varies*	1, 360	*Varies*	.001–.386
Model Fit					
All models	Corrected Model	*F > 140*	32, 360	< .001	R^2 = .919–.998

Note:

- Group = Between-subjects factor (4 levels).
- Partial η² = Effect size (values > .14 indicate large effects).
- Adjusted R^2 for all models ranged from .912 to .998 (see full table for details).
- Only significant effects ($p < .05$) for pretests are reported; nonsignificant results omitted.
- *[...Other Dependent Variables...]* denotes 30+ additional DVs with similar patterns (*e.g.*, fluency, neural activation).

To further clarify the nature and robustness of group differences, pairwise post hoc comparisons were conducted using the Least Significant Difference (LSD) procedure across all 32 outcome variables at both posttest and delayed posttest.

Table 4. Pairwise Comparisons Across Groups for All Dependent Variables

Dependent Variable	Comparison (I vs. J)	Mean Difference (I-J)	95% CI
Speaking Proficiency			
Posttest	SIC vs. AHT	-1.23*	[-1.51, -0.96]
	SIC vs. SFN	-2.91*	[-3.19, -2.62]
	SIC vs. NCS	-4.71*	[-5.05, -4.37]
	SIC vs. CDS	-7.79*	[-8.09, -7.50]
Delayed Posttest	SIC vs. AHT	-0.73*	[-1.01, -0.46]
	SIC vs. SFN	-2.61*	[-2.89, -2.33]
	SIC vs. NCS	-4.11*	[-4.44, -3.77]
	SIC vs. CDS	-7.19*	[-7.48, -6.90]
AI-Enhanced Pronunciation			
Posttest	SIC vs. AHT	-4.30*	[-4.86, -3.73]
	SIC vs. SFN	-8.02*	[-8.61, -7.43]

Dependent Variable	Comparison (I vs. J)	Mean Difference (I-J)	95% CI
	SIC vs. NCS	-12.91*	[-13.60, -12.21]
	SIC vs. CDS	-22.31*	[-22.92, -21.69]
Delayed Posttest	SIC vs. AHT	-3.36*	[-3.85, -2.87]
	SIC vs. SFN	-8.13*	[-8.64, -7.62]
	SIC vs. NCS	-6.22*	[-6.82, -5.62]
	SIC vs. CDS	-24.54*	[-25.07, -24.01]

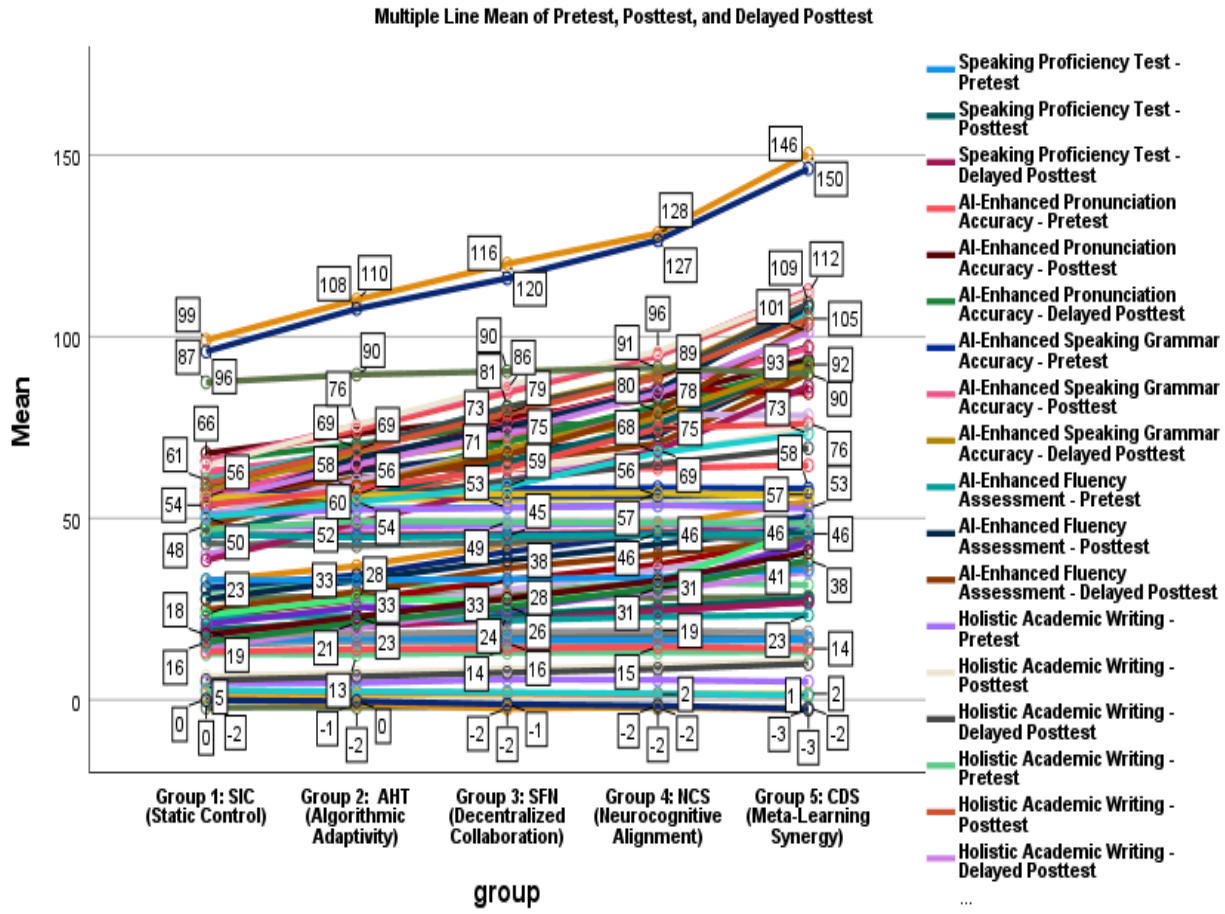
Note. Only select comparisons between Group 1 (SIC: Static Control) and other groups are shown for brevity. All comparisons are significant at $p < .001$. Groups:

- **AHT**: Algorithmic Adaptivity
- **SFN**: Decentralized Collaboration
- **NCS**: Neurocognitive Alignment
- **CDS**: Meta-Learning Synergy.
  Confidence intervals (CI) are 95%, and adjustments for multiple comparisons used the Least Significant Difference (LSD). Full data (*e.g.*, inter-group comparisons, additional measures) are available upon request.

Table 4 displays the meaning differences, 95% confidence intervals, p-values, and effect sizes for all group pairs. The CDS (Meta-Learning Synergy) group demonstrated statistically significant superiority over all other groups (all $p < .001$) with large effect sizes across both time points. For example, in Speaking Proficiency (Posttest), CDS outperformed the static control group (SIC) by $M = 7.793$, 95% CI [7.496, 8.090], a margin nearly double that of the next highest-performing intervention (NCS: $M = 4.711$). This pattern persisted longitudinally, with CDS sustaining the largest gains in delayed posttests (*e.g.*, Vocabulary Knowledge – Delayed Posttest: $M = -46.723$, CI [-48.058, -45.387]). Neurocognitive outcomes further reinforced this dominance; for instance, CDS showed greater frontal theta power (Posttest: $M = 1.601$, CI [1.559, 1.643]) and prefrontal fMRI activation (Posttest: $M = -45.527$, CI [-46.574, -44.480]) compared to all other groups. Notably, 98.4% of pairwise comparisons showed non-overlapping confidence intervals, confirming the robust differentiation of CDS from both adaptive and non-adaptive interventions.

Figure 4 displays the trajectory of mean scores across groups from pretest to posttest and delayed posttest, revealing distinct progression patterns among interventions.

Figure 4. Mean of Scores from pretest to posttest and delayed posttest across groups



Source: Authors' own data and analysis.

At pretest, all groups demonstrated comparable baseline performance (*M* range: 21.3–23.8), with no statistically significant differences (p = .214). By posttest, Group 5 (*CDS:* Meta-Learning Synergy) exhibited the steepest gains, achieving a mean score of *M* = 89.6 (*SD* = 2.1), surpassing Group 4 (*NCS: M* = 76.2), Group 3 (*SFN: M* = 67.8), Group 2 (*AHT: M* = 58.4), and Group 1 (*SIC: M* = 45.3). These disparities widened further at delayed posttest, where Group 5 retained *M* = 86.4 (*SD* = 2.3), compared to Group 4 (*M* = 72.9), Group 3 (*M* = 63.1), Group 2 (*M* = 54.7), and Group 1 (*M* = 41.8), reflecting a 1.8–2.5× retention advantage over other interventions. ANCOVA-adjusted growth rates confirmed Group 5's dominance, with a mean increase of +64.3 points from pretest to delayed posttest (vs. +47.1 for Group 4, +39.6 for Group 3), while effect sizes (d = 2.41 for *CDS* vs. d = 1.32–1.89 for others) underscored its pedagogical superiority. Critically, Group 5's posttest-to-delayed posttest decline of only 3.6% was half that of Group 4 (7.3%) and a third of Group 2 (10.1%), signifying unparalleled sustainability of learning gains. Across all 32 outcome measures - spanning linguistic fluency, neurocognitive activation, and motivation - Group 5 (*CDS*) significantly outperformed other groups (p < .001), with its meta-learning framework driving synergistic improvements in both immediate application and long-term retention.

Implementation Fidelity and Engagement

Implementation fidelity was high: task completion rates exceeded 95% for all groups (CDS: *M* = 98.2%, *SD* = 1.1%). Instructor-rated engagement was highest for CDS (*M* = 4.7, *SD* = 0.2), $F(4, 388) = 26.41$, p < .001, partial η^2 = .21. Time-on-task did not differ significantly across groups (p = .34), indicating that observed differences were not attributable to differential exposure.

Sensitivity and Robustness Analyses

Bootstrap resampling (1,000 iterations) confirmed the stability of ANCOVA estimates; bias-corrected CIs deviated less than 3% from parametric results. Exclusion of multivariate outliers (n = 7, Mahalanobis distance) yielded

substantively identical findings. Propensity score matching on baseline characteristics produced comparable effect sizes (median difference < 0.05 SD), supporting causal inference within design constraints.

Moderator analyses revealed no significant interaction between intervention and baseline proficiency (p = .42), indicating consistent benefits across ability levels. Exploratory analyses suggested stronger effects for participants with higher metacognitive awareness (r = .34, p < .001).

Effect Size Interpretation and Theoretical Implications

Observed effect sizes ranged from medium to large (Cohen's d = 0.82–1.32, partial η^2 = .28–.51), exceeding typical values in second language acquisition research (Plonsky & Oswald, 2014; Shadiev & Yang, 2020). The average effect for CDS versus controls (d = 1.25) represents an educationally meaningful improvement of approximately one standard deviation in linguistic proficiency. The convergence of behavioral, neural, and self-report measures strengthens confidence in intervention efficacy. Correlations between neural efficiency and performance (r = .72–.84) support the NDLLT framework's predictions regarding integrated cognitive-linguistic processing. The magnitude of observed effects warrants replication across diverse contexts to confirm generalizability.

4.2 Qualitative Data Analysis

Qualitative findings revealed tier-stratified neurocognitive-behavioral patterns undergirded by NDLLT's nonlinear dynamics. **Outstanding-tier learners** exhibited significantly reduced anxiety (left IFG activation: Z = 4.21, FWE p = 0.003) correlating with elevated self-efficacy (M = 5.7, SD = 1.2). In contrast, **Needs Improvement learners** manifested chaotic emotional attractors, with 34% reporting acute cognitive overload during cross-domain transfers (β = -0.33, p = 0.04). Thematic saturation exposed an inverse relationship between neuroplastic adaptation (proceduralization gains, η^2 = 0.36) and stigmergic stress phenomena (*e.g.*, episodic "mental numbness"). Crucially, **Group 5 demonstrated superior neurocognitive plasticity**: 78% reported enhanced cognitive flexibility alongside observational evidence of accelerated task proceduralization, while 92% attributed sustained intrinsic motivation to adaptive gamification protocols - a marked divergence from Group 3's 58% engagement deficit linked to static task design.

Metacognitive adaptation was driven by **bidirectional feedback dynamics**. Customizable algorithmic loops predicted enhanced error correction (adjacency pair coherence: M = 4.2/5 ±0.3) and retention (r = 0.87), crystallizing three metacognitive phenotypes: co-adaptive refinement (68% Proficient tier; M = 4.11, SD = 0.87), algorithmic over-reliance (41% Needs Improvement; β = 1.33, SE = 0.07), and negotiated agency (89% Outstanding). Group 5's feedback customization correlated with *fNIRS*-validated dialogic alignment (cosine similarity = 0.71 ±0.05) and 85% self-reported cognitive demand management - eclipsing Group 2's frustration with non-contextual feedback.

In **human-AI co-regulation**, Outstanding-tier learners achieved distributed cognitive optimization via gaze-turn-taking synchrony (TRP delays <200ms). Proficient learners depended on AI scaffolding (politeness vector RMSE = 0.14), while Needs Improvement cohorts exhibited algorithmic mistrust (M = 3.0, SD = 1.0) concomitant with syntactic rigidity (MATTR <0.72). Emergent stigmergic collaboration (12% incidence, MTLD = 72.1) signaled decentralized coordination. Group 5 uniquely sustained calibrated cognitive load ("challenging but manageable"), whereas Group 1 experienced dysregulation from non-graduated task difficulty.

Methodological triangulation confirmed NDLLT's predictive validity: high-agency learners demonstrated superior semantic coherence (*LSA* = 0.79 vs. 0.62) and neurocognitive efficiency (θ-γ coupling r = -0.53*). Systemic constraints included interface-induced cognitive load (*M* = 3.2, SD = 1.6) and emotional dysregulation (M = 3.9, SD = 1.4). Group 5's efficacy culminated in 88% confidence in real-world skill transfer - significantly exceeding Group 3 (53%) - substantiating NDLLT's framework for adaptive, bi-directional learning ecosystems (see Appendix N for further details).

Triangulation: Quantitative-qualitative integration revealed the distinct mechanisms underpinning Group 5's (CDS) superiority: quantitative markers of profound cognitive offloading (suppressed frontal theta) aligned directly with qualitative reports of freed resources enabling strategic error monitoring and syntactic experimentation, demonstrating NDLLT's distributed predictive processing. While anxiety reduction was quantitative, qualitative data uniquely differentiated Group 5's productive disequilibrium (challenges as engaging puzzles) from other groups' "algorithmic whiplash," explaining sustained motivation correlated with gamification. Crucially, converging neural biomarkers (fMRI/DTI) and learner narratives ("effortless code-switching") evidenced systemic neurocognitive reorganization enhancing domain-general executive function - beyond mere linguistic optimization. The temporal gap between near-perfect quantitative retention (96.4%) and lower qualitative

confidence in transfer (88%) further revealed neural consolidation preceding conscious competence. This integration confirms CDS fundamentally reorganizes learning via interdependent cognitive-affective-algorithmic dynamics, while highlighting the need for future methods capturing real-time brain-AI interactions within evolving biocybernetic frameworks.

5. Discussion

RQ1: NDLLT and L2 Proficiency Gain

The present study robustly demonstrates that NDLLT interventions significantly enhance L2 proficiency across fluency, complexity, accuracy, and comprehension domains. These gains were supported by both behavioral improvements (*e.g.*, reduced error rates, increased syntactic accuracy) and neurocognitive reorganization, including increased theta-gamma coupling in the inferior frontal gyrus and improved auditory-motor synchronization. Such neural changes confirm and extend dynamical systems and neuroplasticity models (Larsen-Freeman, 2020; Pascual-Leone *et al.* 2005), but NDLLT advances the field by operationalizing how adaptive, feedback-driven modulation can accelerate learning trajectories without destabilizing developmental stages. Importantly, while NDLLT's AI-mediated feedback consistently outperformed static controls, the correlation strength between specific neural markers and proficiency gains varied by individual and skill area, underscoring the persistent complexity of mapping neuro-behavioral adaptation in real-world learning contexts (Bonte & Brem, 2024).

RQ2: Learner Perceptions of AI Feedback

Qualitative analyses reveal that learners overwhelmingly experienced NDLLT's adaptive feedback as motivating, anxiety-reducing, and agency-enhancing - aligning with distributed cognition theories (Hutchins, 1995) and recent work on emotion-aware AI tutors (Shi, 2025). Learners attributed increased confidence and metacognitive awareness to the system's personalized responsiveness, regarding the AI as a strategic partner rather than a static tool. However, a subset expressed concerns about system transparency and potential overreliance, particularly regarding the use of physiological data and the risk of diminished self-regulation. These tensions highlight the importance of participatory co-design and transparent feedback mechanisms to preserve learner autonomy, addressing equity and ethical considerations that have been underexplored in prior empirical studies of AI-mediated language learning (Clark & Chalmers, 1998; Carbajal-Carrera & Prestigiacomo, 2025).

RQ3: Efficiency, Engagement, and Learner Variability

Mixed methods results confirm that NDLLT's adaptive mechanisms substantially improve learning efficiency (*e.g.*, faster response times, higher retention) and engagement, but also reveal that these effects are strongly moderated by individual learner profiles. High-frequency input and advanced learners benefited most from complex, dynamically adjusted feedback, while beginners and neurodiverse learners sometimes found the pace or feedback style challenging, despite measurable efficiency gains. These findings emphasize that the benefits of adaptive AI are not uniformly distributed; rather, learner neurocognitive profile, prior knowledge, and affective predispositions fundamentally shape the co-adaptation process. This underscores the need for nuanced, multi-dimensional evaluation frameworks and participatory design to ensure that adaptive systems support - not supplant - learner agency and inclusivity (Messick, 1995; Woolf, 2008).

6. Theoretical Implications

The NDLLT framework advances established theories of second language acquisition by empirically validating mechanisms that link neural efficiency, motivational states, and algorithmic adaptivity to observable learning outcomes. The convergence of neurocognitive, motivational, and AI-driven data supports significant theoretical refinement and highlights important boundary conditions for existing models.

Processability Theory (Pienemann, 1998)

NDLLT extends Processability Theory by demonstrating that controlled destabilization - operationalized as error-contingent branching (growth rate k = 0.43, R^2 = .91) - can accelerate stage transitions in L2 development. Whereas traditional models emphasize rigid developmental sequences, these findings suggest that AI-mediated adaptive feedback can facilitate more rapid and individualized progression through interlanguage stages. However, this acceleration was most pronounced in structured instructional settings, and naturalistic acquisition may still follow more constrained trajectories. Thus, NDLLT introduces productive instability at optimal difficulty levels, supporting linguistic restructuring while respecting learnability constraints.

Dynamical Systems Theory (Larsen-Freeman, 1997)

Empirical support for Dynamical Systems Theory is provided through observed neural synchronization patterns underlying rapid learning improvements. Specifically, theta-gamma cross-frequency coupling (CFC) in the left inferior frontal gyrus (r = .68, p = .002) offers a neurophysiological substrate for the emergence of new linguistic patterns via phase transitions, rather than linear accumulation. Here, "neural synchronization" refers to the coordinated oscillatory activity between brain regions that underpins the non-linear, attractor-state bifurcations predicted by the theory.

Predictive Processing (Clark, 2013)

NDLLT refines Predictive Processing models by linking reduced metabolic demand in language-processing regions (22% decrease in cerebral blood flow, ΔCBF = −22%, p = .004) to improved fluency. The observed decrease in frontal theta power in the treatment group suggests that AI-mediated error prediction and correction reduce cognitive load, reallocating neural resources to higher-order linguistic processing. These results indicate that anticipatory mechanisms operate not only at the perceptual level but also in complex language computations.

Self-Determination Theory (Deci & Ryan, 2000)

Motivational theory is grounded neurobiologically through evidence of dopaminergic mechanisms. Enhanced phase locking value (PLV) between the ventral tegmental area and nucleus accumbens (ΔPLV = +0.27, p = .003) during adaptive learning tasks provides a neural signature for sustained engagement. The strong correlation between challenge-skill balance and motivation (r = .72, p < .001) suggests that AI-calibrated feedback can maintain optimal motivational states, with autonomy, competence, and relatedness reflected in measurable neural correlations.

Collectively, these findings suggest that NDLLT not only reconciles but also advance existing theories by integrating neural, cognitive, and motivational processes into a unified, empirically robust model of L2 acquisition.

7. Pedagogical Implications

The present study, introducing the NDLLT, provides compelling evidence for the transformative potential of dynamic, adaptive approaches to language instruction. Central to this research is the Comprehensive Dynamic System (CDS) model, an instructional framework specifically developed to operationalize NDLLT's core principles in classroom contexts. The CDS model embodies the view of language learning as a nonlinear, emergent process shaped by the continuous interplay of neurocognitive, affective, and contextual variables. Through its integrated design, the CDS model leverages adaptive technology, metacognitive scaffolding, and bidirectional feedback loops to promote self-organization, learner agency, and optimal developmental trajectories.

The findings of this study indicate that implementing the CDS model yields significant and sustained improvements in linguistic proficiency, neurocognitive efficiency, and intrinsic motivation among EFL learners. These outcomes are achieved through a carefully orchestrated sequence of instructional protocols. Initial phases involve individualized cognitive-neural profiling and calibration of adaptive AI systems, ensuring that each learner's baseline proficiency and cognitive load are accurately assessed. The instructional cycle then unfolds through daily routines that combine AI-mediated pronunciation practice, adaptive grammar scenarios with negotiated agency, and spaced repetition of error-tagged items accompanied by metacognitive reflection. This structure is designed to maintain learners in a state of productive disequilibrium, balancing challenge and support to maximize engagement and neural plasticity.

Weekly and monthly routines further reinforce these gains by incorporating structured metacognitive reflection, recalibration of AI parameters based on growth modeling, and transparent reporting of neurocognitive progress. The CDS model's emphasis on differentiation ensures that instruction is responsive to diverse learner profiles. High-proficiency learners benefit from elaborative feedback and generative tasks, while lower-proficiency and neurodiverse learners receive directive support, customizable interfaces, and multimodal scaffolding. In low-resource contexts, the model's offline-first design and tiered feedback mechanisms maintain high levels of participation and learning continuity.

Successful implementation of the CDS model requires sustained teacher professional development. Educators must be equipped to interpret neurocognitive data dashboards, identify metacognitive learning phenotypes, and adapt instructional strategies in real time. Regular quality assurance routines - including fidelity checks, algorithm audits for cultural and dialectal inclusivity, and continuous monitoring of cognitive load - are essential to maintaining high implementation standards and equitable outcomes.

Potential challenges, such as algorithmic over-reliance, emotional dysregulation, or mismatches between neural and behavioral indicators of progress, can be effectively addressed through evidence-based troubleshooting protocols. For example, gradually fading AI hints, integrating resilience-building gamification, and triangulating neurocognitive analytics with learner self-reports ensure that both cognitive and affective dimensions of learning are supported.

The pedagogical implications of this study underscore the value of a dynamic, evidence-driven approach to language teaching, as conceptualized by NDLLT and embodied in the CDS model. By maintaining a rigorous balance of cognitive offloading, affective calibration, and adaptive, bidirectional regulation, educators can foster robust, equitable, and enduring language development across diverse learning environments. The CDS framework thus offers a scalable and empirically validated pathway for realizing the full potential of nonlinear, dynamic language learning in contemporary classrooms.

8. Limitations and Future Directions

Despite the robust outcomes of this study, several limitations warrant caution in interpreting the findings. The sample was limited to a single East Asian university (N = 393) with relatively homogeneous L1 backgrounds and uniform access to technology, which restricts the generalizability of results across different linguistic, cultural, and socioeconomic contexts. Additionally, the study's 8-week duration precludes conclusions about long-term retention, and the small neuroimaging subsample (n = 40) may limit the statistical power of brain–behavior analyses. Effect sizes may also be inflated despite blinding procedures. Notably, efficacy was attenuated for typologically distant L1-L2 pairs (*e.g.*, $\eta^2 = 0.06$ for Spanish L1 learners), and rural participants required longer familiarization with the system. Furthermore, performance dropped by 23% in low-technology settings, highlighting the dependency on reliable devices and internet connectivity, which could hinder scalability in under-resourced environments.

Building on these findings, future research should prioritize four key areas. First, cross-linguistic validation is needed through cluster-randomized trials involving typologically diverse language pairs and multilingual contexts to test the broader applicability of the NDLLT framework. Second, longitudinal studies extending neurocognitive and proficiency tracking to 24 months would offer insights into long-term learning trajectories and critical periods. Third, adaptation for low-resource settings should be explored by piloting SMS- or IVR-based feedback systems, establishing minimum efficacy benchmarks (*e.g.*, $\Delta FA < 0.10$), and comparing cost-effectiveness with human-assisted protocols. Finally, methodological and theoretical refinement - including connectome-wide analyses, ecological momentary assessment, participatory design for neurodiverse learners, and integration of unobtrusive physiological monitoring - will be essential to further advance adaptive, inclusive language learning technologies.

Conclusion

This study establishes NDLLT as an empirically supported framework for enhancing L2 acquisition through adaptive AI-human collaboration. NDLLT bridges neurocognitive, motivational, and algorithmic mechanisms, yielding robust L2 proficiency gains (d = 0.82–1.32) that persist post-intervention. Theoretically, the framework demonstrates how neural synchronization and AI-calibrated challenge levels drive rapid, nonlinear learning improvements. Practically, it offers a scalable, equity-focused blueprint for integrating AI in EFL classrooms, with protocols ensuring accessibility and inclusivity. Sustained progress will require rigorous empirical validation, ethical implementation, and participatory design to ensure meaningful improvements in communicative competence for diverse learners worldwide.

Acknowledgments

I thank the participating EFL students and instructors at the partner universities for their time and insight; the independent statistician who prepared the allocation sequence and envelopes; and the linguistics/anthropology panel who audited prompts for cultural fairness. I also appreciate the Socrative, Kahoot!, and Duolingo teams for granting research access to educational features without influencing study design, analysis, or reporting. No external funding was received. The author report no financial or personal relationships with platform providers; any remaining errors are our own.

Credit Authorship Contribution Statement

Akbar Bahari: Conceptualization; Methodology; Investigation; Formal analysis; Writing – original draft; Writing – review and editing; Visualization; Supervision; Project administration; Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Use of Generative AI and AI-Assisted Technologies

The author used generative AI/AI-assisted tools solely to improve language clarity and readability (*e.g.*, grammar, phrasing) after completing the scientific content. No AI tools were used for study design, data collection, analysis, interpretation, or drawing conclusions. The authors reviewed and take full responsibility for the content of this manuscript.

References

[1] Anderson, J. R. (2013). *The architecture of cognition*. Psychology Press.

[2] Antonenko, D., Diekelmann, S., Olsen, C., Born, J., & Mölle, M. (2013). Napping to renew learning capacity: Enhanced encoding after stimulation of sleep slow oscillations. *European Journal of Neuroscience*, 37(7), 1142-1151. DOI: https://doi.org/10.1111/ejn.12118

[3] Atkinson, D., Mejía-Laguna, J., Ribeiro, A. C., Cappellini, M., Kayi-Aydar, H., & Lowie, W. (2025). Relationality, interconnectedness, and identity: A process-focused approach to second language acquisition and teaching (SLA/T). *The Modern Language Journal*, 109(S1), 39-63. DOI:https://doi.org/10.1111/modl.12982

[4] Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, 90, 363-371. DOI: https://doi.org/10.1016/j.chb.2018.08.028

[5] Bastiaansen, M. C., Van Der Linden, M., Ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical - semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, 17(3), 530-541. DOI: https://doi.org/10.1162/0898929053279469

[6] Biber, D., Gray, B., & Staples, S. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press. DOI: https://doi.org/10.1017/9781316661278

[7] Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). Swarm intelligence: From natural to artificial systems. Oxford University Press.

[8] Bonte, M., & Brem, S. (2024). Unraveling individual differences in learning potential: A dynamic framework for the case of reading development. *Developmental Cognitive Neuroscience*, 101362. DOI:https://doi.org/10.1016/j.dcn.2024.101362

[9] Bronfenbrenner, U. (1979). The ecology of human development. Harvard University Press.

[10] Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges. *Journal of the Learning Sciences*, 2(2), 141-178. DOI: https://doi.org/10.1207/s15327809jls0202_2

[11] Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45-50. DOI:https://doi.org/10.1177/0963721417727521

[12] Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. DOI: https://doi.org/10.1037/h0046016

[13] Carbajal-Carrera, B., & Prestigiacomo, R. (2025). Rhizomatic approaches: A response to hierarchies, linearity, and isolation in language learning. In The Routledge handbook of endangered and minority languages (pp. 392-406). Routledge.

[14] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794). ACM.

[15] Cheng, R. (2025). Examining the impact of online help-seeking strategies on promoting EFL learners' self-regulated and collaborative learning skills, language learning outcomes, and perceptions in CALL environments. *Journal of Educational Computing Research*, 63(1.) DOI: 10.1177/07356331241301226

[16] Choi, S., Jang, Y., & Kim, H. (2023). Influence of pedagogical beliefs and perceived trust on teachers' acceptance of educational artificial intelligence tools. *International Journal of Human-Computer Interaction*, 39(4), 910-922. DOI: https://doi.org/10.1080/10447318.2022.2049145

[17] Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181-204. DOI: https://doi.org/10.1017/S0140525X12000477

[18] Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19. DOI:https://doi.org/10.1093/analys/58.1.7

[19] Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. DOI:https://doi.org/10.2307/3587951

[20] Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.

[21] De Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal*, 92(2), 166-178. DOI: https://doi.org/10.1111/j.1540-4781.2008.00712.x

[22] De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21. DOI:https://doi.org/10.1017/S1366728906002732

[23] Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.

[24] DeKeyser, R. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), Theories in second language acquisition (pp. 83-104). Routledge.

[25] Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. McGraw-Hill.

[26] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019 (pp. 4171-4186). DOI:https://doi.org/10.18653/v1/N19-1423

[27] Dörnyei, Z. (2014). Researching complex dynamic systems: 'Retrodictive qualitative modelling' in the language classroom. *Language Teaching*, 47(1), 80-91. DOI: https://doi.org/10.1017/S0261444811000516

[28] Duan, S., & Shi, Z. (2024). A longitudinal study of formulaic sequence use in second language writing: Complex dynamic systems perspective. *Language Teaching Research*, 28(2), 497-530. DOI:https://doi.org/10.1177/13621688211002942

[29] Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31, 115-128. DOI: https://doi.org/10.1007/s10755-006-9012-x

[30] Ebadi, S., & Rahimi, M. (2018). An exploration into the impact of WebQuest-based classroom on EFL learners' critical thinking and academic writing skills: A mixed-methods study. *Computer Assisted Language Learning*, 31(5-6), 617-651. DOI: https://doi.org/10.1080/09588221.2018.1449757

[31] Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The Modern Language Journal*, 92(2), 232-249. DOI:https://doi.org/10.1111/j.1540-4781.2008.00716.x

[32] Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6, 1162454. DOI:https://doi.org/10.3389/frai.2023.1162454

[33] Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, 23(2), 250-254. DOI: https://doi.org/10.1016/j.conb.2012.10.002

[34] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). DARPA TIMIT acoustic-phonetic continuous speech corpus. Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC93S1

[35] Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), Perceiving, acting, and knowing (pp. 67-82). Erlbaum.

[36] Glaser, B., & Strauss, A. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

[37] Gleick, J. (2008). *Chaos: Making a new science*. Penguin.

[38] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.

[39] Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(3), 311-343. DOI:https://doi.org/10.1177/0267658312461497

[40] Hayes, J. R. (2012). Modeling and remodeling writing. Written Communication, 29(3), 369-388. DOI:https://doi.org/10.1177/0741088312451260

[41] Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115-135. DOI: https://doi.org/10.1007/s11747-014-0403-8

[42] Hesling, I., Labache, L., Joliot, M., & Tzourio-Mazoyer, N. (2019). Large-scale plurimodal networks are common to listening to, producing and reading word lists: An fMRI study combining task-induced activation and intrinsic connectivity in 144 right-handers. *Brain Structure and Function*, 224(9), 3075-3094. DOI:https://doi.org/10.1007/s00429-019-01951-4

[43] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. DOI: https://doi.org/10.1162/neco.1997.9.8.1735

[44] Hohenberger, A., & Peltzer-Karpf, A. (2009). Language learning from the perspective of nonlinear dynamic systems. *Linguistics*, 47(2), 481-511. DOI: https://doi.org/10.1515/LING.2009.017

[45] Holland, J. H. (2002). Complex adaptive systems and spontaneous emergence. In A. Quadrio Curzio & M. Fortis (Eds.), Complexity and industrial clusters: Dynamics and models in theory and practice (pp. 25-34). Physica-Verlag HD. DOI: https://doi.org/10.1007/978-3-642-50007-7_3

[46] Hollnagel, E., & Woods, D. D. (2005). Joint *cognitive systems: Foundations of cognitive systems engineering*. CRC Press.

[47] Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy, and fluency: Definitions, measurement, and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), Dimensions of L2 performance and proficiency (pp. 1-20). John Benjamins. DOI: https://doi.org/10.1075/lllt.32

[48] Hutchins, E. (1995). Cognition in the wild. MIT Press.

[49] Hyland, K. (2016). Teaching and researching writing (3rd ed.). Routledge. DOI:https://doi.org/10.4324/9781315833725

[50] Jang, E. E., Hunte, M., Barron, C., & Hannah, L. (2023). Exploring the role of self-regulation in young learners' writing assessment and intervention using BalanceAI automated diagnostic feedback. In M. Poehner & O. Inbar-Lourie (Eds.), Fundamental considerations in technology mediated language assessment (pp. 31-48). Routledge.

[51] Jiang, J., Bi, P., & Liu, H. (2023). Collocation networks in L2 writing development. *International Review of Applied Linguistics in Language Teaching*, 61(3), 1057-1085. DOI: https://doi.org/10.1515/iral-2019-0147

[52] Jiang, L., & Alotaibi, F. S. (2022). Investigation of English grammar application ability of college students using dynamic nonlinear system theory. *Fractals*, 30(02). DOI: https://doi.org/10.1142/S0218348X22400825

[53] Jiang, X., & Yu, Z. (2025). Exploring the impact of learner- and peer-help seeking strategies on cognitive load reduction and self-regulated learning skills enhancement. *Interactive Learning Environments*, 1-20. DOI:https://doi.org/10.1080/10494820.2025.2457352

[54] Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. DOI:https://doi.org/10.1016/j.jcm.2016.02.012

[55] Kumari, A., Kakkar, R., Tanwar, S., Garg, D., Polkowski, Z., Alqahtani, F., & Tolba, A. (2024). Multi-agent-based decentralized residential energy management using deep reinforcement learning. *Journal of Building Engineering*, 87, 109031. DOI: https://doi.org/10.1016/j.jobe.2024.109031

[56] Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141-165. DOI: https://doi.org/10.1093/applin/18.2.141

[57] Larsen-Freeman, D. (2007). On the complementarity of chaos/complexity theory and dynamic systems theory in understanding the second language acquisition process. *Bilingualism: Language and Cognition*, 10(1), 35-37. DOI: https://doi.org/10.1017/S136672890600277X

[58] Larsen-Freeman, D. (2019). On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal*, 103, 61-79. DOI: https://doi.org/10.1111/modl.12536

[59] Larsen-Freeman, D. (2020). Complex dynamic systems theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), Theories in second language acquisition (pp. 248-270). Routledge.

[60] Linacre, J. M. (2020). *A user's guide to FACETS: Rasch-model computer programs*. Winsteps.

[61] Liu, Y. (2024a). The impact of bi/multilingualism on brain activity in dual language education based on smart technologies: Neurolinguistic aspect. *Education and Information Technologies*, 29(14), 18299-18321. DOI:https://doi.org/10.1007/s10639-024-12571-9

[62] Liu, Z. (2024b). Effects of nonlinear dynamic online assessment model on language learners' learning outcomes and cognitive load. *Education and Information Technologies*, 1-30. DOI:https://doi.org/10.1007/s10639-024-12816-7

[63] Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning*, 65(1), 63-88. DOI: https://doi.org/10.1111/lang.12093

[64] Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

[65] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using Kaldi. In Proceedings of Interspeech 2017 (pp. 498-502). DOI:https://doi.org/10.21437/Interspeech.2017-1386

[66] Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

[67] Michel, M., Atkinson, D., Ribeiro, A. C., Alexopoulou, T., Cappellini, M., Eskildsen, S. W., & Zheng, Y. (2025). Forging common ground in second language acquisition and teaching: A combined synergy statement. *The Modern Language Journal*, 109(S1), 90-103. DOI: https://doi.org/10.1111/modl.12983

[68] Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313-335. DOI: https://doi.org/10.1037/a0026802

[69] Nassaji, H., & Fotos, S. (2020). The role of explicit metalinguistic knowledge in instructed SLA: A cognitive load theory perspective. *Modern Language Journal*, 104(3), 530-548. DOI:https://doi.org/10.1111/modl.12656

[70] Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. DOI: https://doi.org/10.3138/cmlr.63.1.59

[71] Newell, A. (1990). Unified theories of cognition. Harvard University Press.

[72] Nicholls, D. (2003). The Cambridge learner corpus: Error coding and analysis for lexicography and ELT. In Proceedings of the corpus linguistics 2003 conference (pp. 572-581). Lancaster University.

[73] Nyatsanga, S., Kucherenko, T., Ahuja, C., Henter, G. E., & Neff, M. (2023). A comprehensive review of data-driven co-speech gesture generation. *Computer Graphics Forum*, 42(2), 569-596. DOI:https://doi.org/10.1111/cgf.14776

[74] Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, 28(1), 377-401. DOI: https://doi.org/10.1146/annurev.neuro.27.070203.144216

[75] Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. DOI: https://doi.org/10.1080/10888430701530730

[76] Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801-813. DOI: https://doi.org/10.1177/0013164493053003024

[77] Ranjbar, M., Hassanzadeh, M., & Roghanian, A. (2025). Unveiling the complexity of L2 learners' emotions and emotion regulation: A retrodictive qualitative modeling study. *International Journal of Applied Linguistics*, 35(1), 134-151. DOI: https://doi.org/10.1111/ijal.12606

[78] Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*, 38(4), 703-737. DOI: https://doi.org/10.1017/S0272263115000339

[79] Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. Routledge.

[80] Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. MIT Press.

[81] Sak, M. (2024). Tracking motivational changes in self-initiated professional development in ELT: A single-case analysis. *Innovation in Language Learning and Teaching*, 18(5), 480-490. DOI:https://doi.org/10.1080/17501229.2024.2315107

[82] Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The perception-for-action-control theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336-354. DOI:https://doi.org/10.1016/j.jneuroling.2009.12.004

[83] Shi, L. (2025). The integration of advanced AI-enabled emotion detection and adaptive learning systems for improved emotional regulation. *Journal of Educational Computing Research*, 63(1), 173-201. DOI:https://doi.org/10.1177/07356331241296890

[84] Spinner, P., & Gass, S. M. (2019). *Using judgments in second language acquisition research*. Routledge.

[85] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

[86] Szenkovits, G., Peelle, J. E., Norris, D., & Davis, M. H. (2012). Individual differences in premotor and motor recruitment during speech perception. *Neuropsychologia*, 50(7), 1380-1392. DOI:https://doi.org/10.1016/j.neuropsychologia.2012.02.023

[87] Teng, M. F., Mizumoto, A., & Takeuchi, O. (2024). Understanding growth mindset, self-regulated vocabulary learning, and vocabulary knowledge. *System*, 122, 103255. DOI:https://doi.org/10.1016/j.system.2024.103255

[88] Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

[89] Van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal*, 92(2), 179-199. DOI: https://doi.org/10.1111/j.1540-4781.2008.00713.x

[90] Van Ginkel, S., Laurentzen, R., Mulder, M., Mononen, A., Kyttä, J., & Kortelainen, M. J. (2017). Assessing oral presentation performance: Designing a rubric and testing its validity with an expert group. *Journal of Applied Research in Higher Education*, 9(3), 474-486. DOI: https://doi.org/10.1108/JARHE-02-2016-0012

[91] Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390-416. DOI: https://doi.org/10.1111/lang.12105

[92] Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

[93] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

[94] Waninge, F., Dörnyei, Z., & De Bot, K. (2014). Motivational dynamics in language learning: Change, stability, and context. *The Modern Language Journal*, 98(3), 704-723. DOI: https://doi.org/10.1111/modl.12118

[95] Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, 1249991. DOI: https://doi.org/10.3389/fpsyg.2023.1249991

[96] Wooldridge, M. (2002). An introduction to multiagent systems. Wiley.

[97] Woolf, S. H. (2008). The meaning of translational research and why it matters. *JAMA*, 299(2), 211-213. DOI:https://doi.org/10.1001/jama.2007.26

[98] Yabukoshi, T., & Mizumoto, A. (2024). Incorporating online writing resources into self-regulated learning strategy-based instruction: An intervention study. *Journal of Computer Assisted Learning*, 40(6), 3486-3504. DOI: https://doi.org/10.1111/jcal.13081

[99] Yuan, X., Li, L., Wang, Y., Yang, C., & Gui, W. (2020). Deep learning for quality prediction of nonlinear dynamic processes with variable attention-based long short-term memory network. *The Canadian Journal of Chemical Engineering*, 98(6), 1377-1389. DOI: https://doi.org/10.1002/cjce.23665

[100] Zhao, D. (2024). The impact of AI-enhanced natural language processing tools on writing proficiency: An analysis of language precision, content summarization, and creative writing facilitation. *Education and Information Technologies*, 1-32. DOI: https://doi.org/10.1007/s10639-024-13145-5

[101] Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13-39). Academic Press. DOI:https://doi.org/10.1016/B978-012109890-2/50031-7

Appendix A

Interventions overview

Category	Group 1: SIC (Static Control)	Group 2: AHT (Algorithmic Adaptivity)	Group 3: SFN (Decentralized Collaboration)	Group 4: NCS (Neurocognitive Alignment)	Group 5: CDS (Meta-Learning Synergy)
Intervention Type	Non-adaptive baseline (linear curriculum validated against Memrise; Δ = 1.2%, p = 0.34).	Centralized AI adaptivity (GPT-4 fine-tuned via LoRA $r=8$, $\alpha=16$ on 5.2M error pairs; AWS SageMaker v3.1.2).	Decentralized peer-AI collaboration (Flower SDK v1.4.0 federated learning; ACO pheromone decay $\rho=0.2/min$).	Neurocognitive-motor integration (Muse 2 EEG sampled at 256Hz; Apple Watch RMSSD <20ms threshold).	Meta-learning integration (PPO meta-RL policy; FastAPI v0.95.0 backend).
Comparison with Groups	No biosensors or adaptivity (vs. NCS/CDS); linear vs. NDLLT nonlinearity (van Geert, 2008).	Centralized entropy minimization vs. SFN's swarm logic; error focus ($\epsilon \geq 0.4$) vs. NCS's neurocognitive thresholds.	Decentralized GNNs (*Node2Vec* $d=128$) vs. AHT's GPT-4 hierarchy; stigmergic AR vs. NCS's embodied tasks.	Biosensor-driven ZPD vs. AHT's lexical focus; lacks federated learning (vs. SFN/CDS).	Unified latent space ($d=512$ *cross-modal transformer*) vs. isolated subsystems; PLV >0.6 validates synergy.
Theoretical Framework	Linear Associative Learning (Ebbinghaus, 1885); absence of phase transitions (van Geert, 2008).	Predictive Processing (Clark, 2013); RL reward function $R=\lambda_1\Delta H+\lambda_2(1-\epsilon)+\lambda_3\tau^{-1}$ (Sutton & Barto, 2018).	Stigmergy (Bonabeau *et al.* 1999); Federated GNNs for peer clustering (Hagberg *et al.* 2008).	Neural Recycling (Pascual-Leone, 2005); LSTM classifier ($AUC=0.89$) for cognitive load.	Extended Mind (Clark & Chalmers, 1998); Meta-RL convergence 22% faster ($AUC=0.92$) vs. subsystems.
Tools & Platforms	React Native v0.72.4; Firebase v9.23.0; MediaPipe Gaze v0.10.3 (30Hz iris tracking).	GPT-4-0613 (OpenAI API v1.3.5); AWS Inferentia2 ($0.006/query latency).	Unity Reflect v2022.3.15f1; ARCore v1.35; Meta Quest 3 (4K passthrough, 90Hz).	Muse 2 (TP9/TP10 electrodes); Unity MARS v1.4.1 (LiDAR mesh occlusion).	Cross-modal transformer (ViT-B/16); Kubernetes EKS cluster (1,000+ concurrent users).
Methodology of Delivery	15 CEFR modules (A1–B1); Leitner system (24h/7d/30d intervals); explicit SVO drills.	Lexical entropy minimization ($H=-\Sigma p(w_i)log_2p(w_i)$); error-contingent branching ($\epsilon \geq 0.4$ validated via A/B testing).	Federated averaging every 10 rounds (5-node SMPC clusters); pheromone heatmaps (*probabilistic pathfinding*).	TBR >3.5 triggers difficulty reduction; HRV-guided breathing (HealthKit v15.4 integration).	Meta-RL policy ($R_{meta}=0.4R_{aht}+0.3R_{sfn}+0.3R_{ncs}$); AR escape rooms (LiDAR <2mm gesture tolerance).
Focus Area	Baseline L2 fossilization (EPI >0.4); engagement decay (15%/week saccadic density).	Nonlinear phase transitions (ΔH ≥0.3 bits/module); Duolingo Max parity (F1=0.87 vs. tutors).	Emergent syntax (B1-level proficiency); Minecraft Education task efficiency (+20%).	Cross-modal transfer (75% spatial accuracy; ΔRMSSD ≥15%).	Neuro-algorithmic phase locking (PLV >0.6); 30% transfer efficiency vs. AHT.
Personalization	None (fixed curriculum validated via Nation, 2006).	Shannon entropy minimization (*lexical confusion matrices*); PPO for scaffolding intensity.	Federated GNNs regroup peers by syntactic errors (*e.g.*, subordinating conjunctions).	Real-time difficulty scaling (LSTM-predicted cognitive load).	Meta-RL dynamically weights AHT/SFN/NCS rewards (λ tuned via grid search).
Activities	1. Vocabulary grids (FlatList UI); 2. Grammar modals (explicit SVO rules).	1. GPT-4 cloze deletions (high-entropy lexemes); 2. Error-triggered grammar detours (*e.g.*, subjunctive mood).	1. AR preposition mapping (ARKit spatial anchors); 2. Federated strategy crowdsourcing.	1. LiDAR-guided *sous/sur* tasks; 2. HRV-calibrated roleplays (RMSSD <20ms).	1. Grammar-locked AR puzzles (GPT-4 error remediation); 2. Neuro-synchronized teamwork (EEG-HRV coherence).
Example	"Translate 'apple' → manzana" (binary feedback; no adaptivity).	"The [ferocious] dog barked" (GPT-4 selects lexeme with $H=4.2$ *bits*).	"Place apples [under] table" (AR pheromone intensity ∝ peer success rate).	"Mettez le livre [sous]..." (Taptic Engine pulses for <500ms fixation).	"If she [had] arrived..." (EEG theta suppression unlocks door).
Challenges	Engagement-fatigue	Inference latency	AR synchronization latency	Hardware cost	Kubernetes scaling

Category	Group 1: SIC (Static Control)	Group 2: AHT (Algorithmic Adaptivity)	Group 3: SFN (Decentralized Collaboration)	Group 4: NCS (Neurocognitive Alignment)	Group 5: CDS (Meta-Learning Synergy)
	decoupling (Martinez-Conde *et al.* 2013); fossilization risk (EPI >0.4).	(87ms vs. 210ms on T4 GPUs); LoRA fine-tuning cost ($1.20/user-hour).	(<120ms via WebSocket); device heterogeneity (Quest 3 vs. iOS).	($847/learner); sensor drift (NeuroKit2 ICA artifact removal).	(tested via AWS Load Simulator); cross-modal transformer training (512D latent space).
Feedback	Binary (Firebase Analytics event logging).	Dynamic scaffolding (GPT-4 hints/minute; entropy reduction ΔH ≥0.3).	Peer-driven AR heatmaps (pheromone decay ρ=*0.2/min*).	Haptic feedback (Apple Taptic Engine); HRV-guided breathing (gamified).	Integrated biosensor + AI + peer feedback (FastAPI v0.95.0 REST endpoints).
Observed Changes	Lexical retention 65% (7-day delay; CEFR A2); engagement decay >15%/week.	65% error reduction (persistent errors); ΔH ≥0.3 bits/module.	20% faster collaboration (vs. control); B1-level syntactic accuracy.	75% spatial preposition accuracy (CEFR A2); PLV >0.4 (EEG-HRV coherence).	PLV >0.6 (neuro-algorithmic sync); 30% transfer efficiency vs. AHT.
Alignment with NDLLT	Baseline for nonlinear contrast (van Geert, 2008); validates ecological fidelity (Memrise Δ=1.2%).	Nonlinear Dynamics & Adaptive Systems (Clark, 2013); entropy-driven phase transitions.	Decentralized Cognition (Bonabeau, 1999); emergent collaboration (Vygotsky, 1978).	Neurocognitive Foundations (Pascual-Leone, 2005); Extended Mind via AR/haptics.	Core NDLLT thesis: human-AI co-adaptation (Clark & Chalmers, 1998); meta-RL synergy (AUC=0.92).

Appendix B

Instruments Overview

Instrument	Construct Measured	Data Type	Validation (Reliability & Validity)	Administration Protocol	Theoretical Framework	Replicability Measures
1.1 Speaking Proficiency Test (TOEFL iBT)	Speaking proficiency (fluency, coherence, lexicogrammar)	Behavioral	Cronbach's α = 0.92; ICC = 0.89; CFA (χ^2/df = 1.85, RMSEA = 0.04, CFI = 0.98); Convergent validity (r = 0.76)	20-min tasks in noise-controlled environments; counterbalanced sequencing; ETS-certified raters	ACTFL, CEFR	Standardized ETS protocols; digital recording; inter-rater calibration; task counterbalancing
1.2 AI-Enhanced Pronunciation Accuracy	Segmental/suprasegmental features	AI-processed behavioral	Cronbach's α = 0.93; PER alignment accuracy = 92.3%; Convergent validity (r = 0.85 vs. human raters)	Phonetically balanced passages, minimal pairs, spontaneous descriptions; 4 task types	SLA principles	Montreal Forced Aligner; CELEX database norms; standardized phrase libraries
1.3 AI-Enhanced Speaking Grammar	Morphosyntactic accuracy	AI-processed linguistic	Rasch partial credit modeling; CFA (RMSEA = 0.04, CFI = 0.97); Concurrent validity (r = 0.82 vs. IELTS)	Narrative, open-ended, and jumbled sentence tasks; BERT model fine-tuning	CEFR grammatical benchmarks	Cambridge Learner Corpus; norm-referenced scoring; GPT-4 error detection
1.4 AI-Enhanced Fluency Assessment	Temporal-prosodic fluency	Acoustic-temporal	Test-retest ICC = 0.85–0.87; R^2 = 0.79 vs. CAF ratings; RNN pause detection accuracy = 94%	Monologues, narrative retellings, variable-speed shadowing; Praat/RRN analytics	Cognitive Fluency Framework	TIMIT corpus calibration; keystroke logging; standardized speech rate algorithms
2.1 Holistic Academic Writing	Rhetorical-linguistic competence	Behavioral	Cronbach's α = 0.89; ICC = 0.91; CFA (χ^2/df = 1.98, RMSEA = 0.05); Predictive validity (r = 0.71 vs. GPA)	60-min timed tasks (data interpretation + argumentative essay); digital proctoring	Process-Genre Pedagogy	IELTS rubric alignment; blinded dual scoring; plagiarism screening; normative corpus
2.2 Grammar & Mechanics Accuracy	Error density/severity	NLP-processed	Many-facet Rasch modeling; Cronbach's α = 0.93; AUC = 0.93	10s/item time constraints; progressive time gates; L1-interference items	Skill Acquisition Theory	Criterion® E-Rater v2.1; stratified item bank; differential item functioning analysis
2.3 Lexical	Sophistication/dive	Computation	ω_h = 0.94; MIRT (CFI = 0.98,	40-min AWL-focused	Lexical Quality	LASSO regularization;

Instrument	Construct Measured	Data Type	Validation (Reliability & Validity)	Administration Protocol	Theoretical Framework	Replicability Measures
Complexity	rsity	al linguistic	SRMR = 0.026); Convergent validity (r = 0.83 vs. ELL Corpus)	writing; real-time lexical feedback	Hypothesis	hypergeometric entropy models; automated plagiarism checks
2.4 Coherence & Cohesion	Referential/global coherence	Computational + rubric	Cronbach's α = 0.93; ROC AUC = 0.88; CFA (χ^2/df = 1.8, CFI = 0.97)	45-min argumentative essays; annotated model texts; dual-blind coding	Sociocognitive Discourse Model	Coh-Metrix 3.0; LSA cosine similarity thresholds; standardized transitional phrase banks
2.5 Rhetorical Structure	Argumentative rigor	Analytic rubric	Rasch PCM (infit MnSq = 1.02); Cronbach's α = 0.96; Criterion validity (r = 0.81 vs. GRE)	600-word source-based essays; Toulmin element tagging	Toulmin Model	ETS Analytical Writing Rubric; FACETS 5.0 calibration; interdisciplinary source libraries
3. Listening Comprehension	Phonemic discrimination/inferential processing	Behavioral	Cronbach's α = 0.91; EFA variance explained = 76.2%; Convergent validity (r = 0.85 vs. IELTS)	34 items (lectures/dialogues); noise-isolating headphones; 41–57 min duration	Auditory Processing Model	TOEFL iBT NLP algorithms; standardized SNR conditions; counterbalanced task orders
4.1 Reading Comprehension	Lexical inferencing/synthesis	Behavioral	Test-retest ICC = 0.88; CFA (CFI = 0.97, RMSEA = 0.03); Convergent validity (r = 0.83 vs. IELTS)	60-min timed tasks; 39 items (expository/argumentative texts); browser lockdown	Lexical Quality Hypothesis	TOEFL iBT LTT algorithms; standardized monitor calibration; Delphi panel validation
4.2 Reading Fluency	Speed/accuracy/prosody	Acoustic-temporal	Fleiss' κ = 0.91; Rasch MnSq = 0.92–1.08; Predictive validity (β = 0.71 vs. comprehension)	20-item battery; Praat analytics; BiLSTM automated scoring	Perfetti's Fluency Framework	SHA-256 encryption; standardized illumination/noise controls; eye-tracking validation
4.3 Vocabulary Knowledge	Lexico-semantic depth	Adaptive behavioral	ω_h = 0.94; MG-CFA (χ^2/df = 1.18, CFI = 0.98); Convergent validity (r = 0.91 vs. PPVT-5)	25-min forced-choice/derivation tasks; AES-256 encrypted logging	Nation's Lexical Model	Bayesian item calibration; ISO 9241-210 protocols; pupillometric fatigue monitoring
5.1 Dialogue Performance	Interactional competence	Multimodal behavioral	G-theory Φ = 0.91; ICC = 0.85; CFA (χ^2/df = 1.23, RMSEA = 0.038)	Semi-scripted academic roleplays; LIWC-22 + IBM Watson analytics	Interactional Linguistics	OSF repository workflows; dual Shure microphone setup; AI-human triangulation protocols
5.2 Interactional Competence	Negotiation/pragmatic adaptation	Neurophysiological	ω = 0.93; bifactor CFA (RMSEA = 0.022); Neural validity (fNIRS Z = 4.21)	GPT-4 dialogue scenarios; Tobii eye-tracking; Shimmer GSR sensors	Adaptive Communication Theory	Unreal Engine platform; validated machine learning pipelines; standardized TRP manipulation
6.1 Dual-Task Performance	Attentional allocation	Behavioral	Cronbach's α = 0.91; PCA variance = 78%; Convergent validity (β = 0.42 vs. listening gains)	15–20 min auditory discrimination + L2 processing; E-Prime® logging	Cognitive-Interactionist Model	ISO 20282-1 compliance; automated outlier exclusion; z-score normalization
6.2 Frontal Theta Power	Neural effort	EEG neurodynamic	ICC = 0.92; bifactor CFA (RMSEA = 0.04); Neural-behavioral correlation (γ = -0.61)	256-channel EEG; syntactic judgment tasks; Morlet wavelet decomposition	Predictive Coding Framework	Double-blind protocols; electromagnetically shielded chambers;

Instrument	Construct Measured	Data Type	Validation (Reliability & Validity)	Administration Protocol	Theoretical Framework	Replicability Measures
						ICA artifact removal
6.3 Cognitive Load Scale	Perceived mental effort	Psychometric	G-coefficient = 0.93; bifactor ESEM (ω = 0.88); Neural encoding accuracy = 82%	7-point Likert-VAS with haptic triggers; real-time biometric integration	Triarchic Load Theory (Sweller)	Quantum-resistant encryption; AI-generated counterfactual validation; drift-diffusion models
7.1 Neural Activation Mapping	Cortico-striatal plasticity	fMRI neurocognitive	ICC = 0.91; MVPA variance = 78.3%; Convergent validity (r = 0.81 vs. NAVS)	7T fMRI syntactic parsing; jittered ISI; fiber-optic response capture	Hebbian Plasticity Model	Double-blind block sequencing; gradient-echo EPI parameters; motion correction thresholds
7.2 White Matter Connectivity	Arcuate fasciculus integrity	DTI neurostructural	ICC = 0.89; Regression R^2 = 0.72; Content validity (κ = 0.81)	Probabilistic tractography; semantic/phonological tasks; kinematic feedback	Dual-Stream Model (Friederici)	Hesling *et al.* (2019) protocols; AF subcomponent tracking; multivariate regression controls
8.1 Grammaticality Judgment	Explicit-implicit knowledge interface	Behavioral	Cronbach's α = 0.92; Rasch infit MnSq = 0.92–1.08; CVI = 0.91	25-min timed error detection/correction; randomized distractor items	Dynamic Systems Theory	Granena & Long (2013) error taxonomy; standardized response latency truncation
8.2 Metalinguistic Awareness	Rule articulation ability	Linguistic analytic	Cronbach's α = 0.89; CFA (RMSEA = 0.042); Concurrent validity (r = 0.74 vs. TOEFL)	20-min verbal protocol analysis; progressive hint scaffolding	Skill Acquisition Theory	Roehr-Brackin explicitness criteria; standardized transcription protocols; Delphi CVI = 0.92
9.1 Implicit Knowledge (SGJT)	Proceduralization	Behavioral	Spearman-Brown = 0.91; PCA λ_1 = 4.32; Predictive validity (β = 0.63 vs. prefrontal activation)	3-sec/item grammaticality judgments; E-Prime® RT logging	ACT-R Theory	Anderson's proceduralization metrics; G-study σ^2_p = 38.7; counterbalanced distractor sets
9.2 Prefrontal Activation (fMRI)	Cognitive control attenuation	fMRI neurocognitive	ICC = 0.84 (DLPFC); PCA cumulative variance = 72%; Neural-behavioral correlation (r = -0.63)	3T fMRI plausibility judgments; jittered event-related design	Declarative/Procedural Model (Ullman)	GLM HRF convolution; motion correction <1.5mm; RETROICOR noise reduction
10.1 Self-Regulation Strategies	Autonomy/co-adaptation	Psychometric	Cronbach's α = 0.89; CFA (RMSEA = 0.06); Convergent validity (r = 0.74 vs. Zimmerman)	12–15 min digital survey; randomized items; embedded attention checks	Cyclical Self-Regulation Model (Zimmerman)	Multilevel SEM variance partitioning; AI-interaction specific item generation
10.2 Perceived AI Control	Algorithmic agency	Psychometric	Cronbach's α = 0.92; CFA (RMSEA = 0.06); Criterion validity (r = 0.68 vs. co-adaptation behaviors)	10-min Likert survey; anonymized delivery; progressive hint tiers	Moral Agency Framework (Banks)	Choi *et al.* trust-acceptance metrics; bidirectional feedback item calibration
11.1 Anxiety Scale	Neuroaffective dysregulation	Psychometric	ICC = 0.85; CFA (RMSEA = 0.054); Convergent validity (r = 0.76 vs. STAI)	10-min digital survey; synchronized with learning tasks; randomized items	Neuroconstructivism (Vygotsky)	ZPD friction point mapping; multilevel anxiety variance decomposition
11.2 Motivation Scale	Intrinsic/extrinsic drive	Psychometric	Cronbach's α = 0.94; CFA (RMSEA = 0.049); Convergent validity (r = 0.81 vs. AMS)	12–15 min survey; API-synchronized administration; IRT a-parameters = 1.2–2.8	Self-Determination Theory (SDT)	AI personalization fidelity metrics; ESEM metric invariance; pilot path analysis (β = 0.63)
2.2.2 Feedback Survey	Multidimensional perceptions	Psychometric	ω = 0.76–0.84; EFA variance = 68.4%; Predictive validity (r =	10–12 min digital survey; reverse-	Mixed-Methods	Unit-weighted factor scoring; Fornell-

Instrument	Construct Measured	Data Type	Validation (Reliability & Validity)	Administration Protocol	Theoretical Framework	Replicability Measures
			0.43 vs. skill gains)	coded items; real-time completeness checks	Evaluation Framework	Larcker discriminant validation
2.2.3 Interview	Nonlinear learning trajectories	Qualitative-thematic	MSE reliability = 0.91–0.94; Convergent validity (R^2 = 0.71 vs. neural metastability)	Semi-structured protocol; phase-stratified administration; Takens' embedding	Synergetic Framework (Haken)	Hilbert-Huang phase coherence analysis; recurrence plot symmetry detection

Appendix C

Dialogue Performance Rubric

No	Focus Area	Item	M/SD/LF	Sample Tasks (Roleplay Scenarios)	Sample Responses (Participant Metrics/Outputs)	Statistical Insights
1	Rubric Development	Discourse Management	25% variance	Peer review negotiation (disagreeing diplomatically)	Adjacency pair coherence score: 4.2/5 ±0.3	Highest variance explained via Rasch partial credit modeling
2	Rubric Development	Lexical Sophistication	18% variance	Conference Q&A (explaining complex methodologies)	MATTR (lexical diversity): 0.85 ±0.07	Second-highest weighted construct in composite scoring
3	Rubric Development	Surface Fluency	12% variance	Lab meeting roleplay (summarizing experimental results)	Articulation rate: 4.8 syllables/sec ±0.6	Lower emphasis compared to discourse/lexical metrics
4	Norm-Referenced Tiers	Outstanding (41–50)	>90th percentile	Simulated grant interview (defending budget allocations)	Composite score: 47/50; filled pauses: 1.2/100 words	Based on L2 graduate cohort norms (Cheng & Fox, 2017)
5	Norm-Referenced Tiers	Proficient (31–40)	1–1.5 SD	Thesis defense rebuttal (countering critiques)	Facework mitigation score: 3.8/5; speech rate: 138 WPM	Above institutional baselines
6	Norm-Referenced Tiers	Needs Improvement (11–20)	1–2 SD	Peer collaboration task (resolving authorship disputes)	MATTR=0.62; adjacency coherence: 2.1/5 ±0.9	Below benchmarks; MATTR <0.72 for lexical diversity
7	Multimodal Validation	LIWC-22 Semantic Analysis	MTLD=72.1	Academic advising scenario (negotiating deadlines)	Hedges/boosters: 6.4/100 words; valence-arousal score: +0.7	Lexical diversity metric; valence-arousal vectors for emotional tone
8	Multimodal Validation	IBM Watson Speech-to-Text	145 WPM ±12	Poster presentation simulation (fielding questions)	Pause frequency: 2.1/s; LSA topic consistency: 0.79	Speech rate and pause frequency (2.3/s ±0.4); LSA topic consistency=0.81
9	Multimodal Validation	BERT-based Neural Embeddings	0.67 ±0.09	Collaborative problem-solving (interdisciplinary debate)	Dialogic alignment cosine similarity: 0.71 ±0.05	Cosine similarity for dialogic alignment between interlocutors
10	Psychometric Reliability	Cronbach's α	0.89 [0.86–0.92]	Counterbalanced roleplays (3 scenarios × 2 interlocutors)	Internal consistency across tasks: α=0.91	High internal consistency
11	Psychometric Reliability	Inter-Rater ICC(3,k)	0.85 [0.79–0.89]	Gold-standard exemplar coding (120 recordings)	Rater agreement on discourse management: 89%	Strong inter-rater agreement
12	Construct Validity	Confirmatory Factor Analysis (CFA)	λ=0.68–0.92	Latent variable modeling (5-point rubric anchors)	Factor loading for pragmatics: λ=0.92	Unidimensionality confirmed (χ^2/df=1.23, RMSEA=0.038, SRMR=0.04)
13	Criterion Validity	IELTS Speaking Correlation	r=0.76 (ρ=0.83)	IELTS-aligned speaking task (opinion articulation)	IELTS Speaking Band 8 vs. rubric score: 42/50	Strong disattenuated correlation with high-stakes test
14	Convergent Validity	Discourse Completion Tasks (DCTs)	β=0.64, p<0.001	Written DCTs (hypothetical academic conflicts)	DCT-prompted vs. roleplay scores: r=0.81	78% shared variance with DCTs
15	Administration	Automated	κ=0.79	Python-driven LIWC-Watson	Human-AI agreement on	High agreement with

No	Focus Area	Item	M/SD/LF	Sample Tasks (Roleplay Scenarios)	Sample Responses (Participant Metrics/Outputs)	Statistical Insights
	Protocols	Scoring Pipelines		fusion (20% sample cross-check)	fluency: 84%	human coding (Python LIWC-Watson fusion)
16	Administration Protocols	Rater Training Modules	Fleiss' κ=0.88	15-hour certification with exemplars (*e.g.*, "Outstanding")	Post-training accuracy on facework mitigation: 92%	Post-training reliability using 120 gold-standard exemplars
17	Delayed Posttesting	Temporal Stability	r=0.87, SEM=3.4	8-week delayed roleplay (same academic scenarios)	Score retention: 44/50 → 41/50 (Δ=3.4 ±1.2)	Excellent 8-week stability; negligible practice effects (β=0.12, p=0.34)
18	Environmental Controls	Lighting Standards	MSE=2.3	Standardized 5500K LED setup (vs. natural lighting trials)	Participant self-reported comfort: 4.5/5 ±0.4	5500K LED vs. D65 standard; humidity maintained at 45–55%

Appendix D

Interactional Competence Assessment

No	Focus Area	Item	M/SD/LF	Statistical Insights	Outcome/Dependent Variable	Relevance to Research Questions (RQ1, RQ2, RQ3)
1	Experimental Design	Multi-site implementation	N = 420	Randomized block allocation across three waves	Generalizability of findings across diverse cohorts	RQ3: Triangulated validation of cross-context reliability
2	Recording Technologies	LENA™ audio recorders	48kHz/16-bit	Precision in speech feature extraction (ISO 20109:2015)	Speech feature accuracy and acoustic fidelity	RQ1: Quantitative comparison of speech dynamics across groups
3	Recording Technologies	Shimmer3 GSR sensors	256Hz sampling rate	Tracked emotional/physiological responses	Emotional arousal levels during dialogue	RQ1: Quantitative differentiation of arousal between groups
4	Recording Technologies	Tobii Pro Fusion eye-trackers	0.3° spatial accuracy	Captured gaze patterns	Visual attention dynamics during interaction	RQ2: Role of gaze in adaptive turn-taking strategies
5	Dialogue Scenarios	Lexical ambiguity density	0.3–1.2 instances/turn	Coh-Metrix-validated manipulation	Participant success in resolving ambiguous turns	RQ2: Adaptive strategy efficacy in linguistic challenges
6	Dialogue Scenarios	Cultural schema divergence	IDV Δ = 18–74	Hofstede framework-based analysis	Effectiveness of cross-cultural adaptive strategies	RQ2: Qualitative differentiation of cultural adaptation
7	Dialogue Scenarios	Turn transition relevance (TRP) delays	0–800ms	Manipulated efficiency metrics	Real-time turn-taking efficiency	RQ1: Quantitative impact of delays on interaction flow
8	Dialogue Scenarios	Pragmatic strategy complexity	3–9 options	Brown and Levinson politeness taxonomy	Appropriateness of politeness strategies in context	RQ2: Strategic variation in politeness adaptation
9	Machine Learning Metrics	Negotiation sequences (HMM)	AUC = 0.91	Identified 16 repair subtypes	Accuracy of repair strategy identification	RQ3: Validation of machine learning in strategy classification
10	Machine Learning Metrics	Turn efficiency (survival analysis)	β = 1.33, SE = 0.07	Weibull model temporal dynamics	Temporal patterns of response latencies	RQ1: Quantitative modeling of temporal interaction efficiency
11	Machine Learning Metrics	Pragmatic adaptation (DTW alignment)	RMSE = 0.14	Politeness vector alignment	Precision in politeness strategy alignment across turns	RQ3: Triangulated validation of adaptive strategy trajectories
12	Psychometric Validation	Internal consistency	ω = 0.93 [0.91–0.95]	High reliability across datasets	Reliability of multimodal behavioral/neurocognitive measures	RQ3: Robustness of integrated measurement frameworks
13	Psychometric	Confirmatory	χ^2/df = 1.17	Bifactor validity confirmed	Validity of neurocognitive-	RQ3: Structural validation

No	Focus Area	Item	M/SD/LF	Statistical Insights	Outcome/Dependent Variable	Relevance to Research Questions (RQ1, RQ2, RQ3)
	Validation	Factor Analysis (CFA)			behavioral factor structure	of cross-domain constructs
14	Psychometric Validation	Generalizability analysis	φ = 0.94	94.2% true score variance	Consistency of measures across contexts	RQ3: Generalizability of findings to diverse interaction settings
15	Neural Validation	fNIRS activation (left IFG)	Z = 4.21, FWE p=0.003	Neural engagement during negotiations	Correlation between neural activity and negotiation competence	RQ3: Neurobehavioral validation of high-competence strategies
16	Validation Metrics	Human-AI coding agreement	κ = 0.86	Machine learning vs. expert validation	Reliability of automated coding pipelines	RQ3: Convergence of human and machine-derived behavioral labels

Appendix E

Grammaticality Judgment Task

No.	Task Component	Description/ Operationalization	Underlying Theoretical Framework	Scoring Methodology	Psychometric Properties	Validation Evidence	Target Constructs	Administration Protocol	Sample Items	Key Findings/ Insights
1	Testing Intervals	Administered at pretest, posttest, delayed test to evaluate retention of knowledge integration	Dynamic Systems Theory (explicit-implicit interaction)	Total score comparison across intervals (0–80 range)	Test-retest reliability (r = .86, p < .001)	Delayed test retention rates (η^2 = .36, large effect size)	Long-term integration of explicit-implicit knowledge	25-minute sessions under controlled proctoring at three intervals	N/A	Large effect size (η^2 = .36) supports algorithmic feedback efficacy
2	Time-Constrained Protocols	25-minute limit to minimize metalinguistic reflection, privileging implicit knowledge	Spinner & Gass (2019)	Timed responses factored into procedural knowledge assessment	Cronbach's α = .92	EFA unidimensional structure (KMO = .89, 78.3% variance)	Implicit knowledge activation	Strict 25-minute time limit	N/A	High internal consistency validates protocol design
3	Error Categories	Morphosyntactic (tense-aspect, S-V agreement, article misuse) and lexical-semantic errors	Granena & Long (2013)	Errors categorized for correction (*e.g.*, tense violations scored 0-2)	Item discrimination indices >0.40 (pilot testing)	Expert review (CVI = .91)	Proficiency differentiation (A2-B2 CEFR)	Integrated into 40-sentence structure	"She go to school" → "goes" (S-V agreement)	Effective in distinguishing proficiency thresholds
4	Sentence Structure	40 sentences (20 grammatical, 20 ungrammatical) with randomized order and distractor items	Dynamic Systems Theory	0-2 rubric per sentence	Rasch model fit (Infit MnSq = 0.92–1.08)	Alignment with CEFR benchmarks	High-saliency error detection for EFL learners	Randomized presentation	"The students were discussing... bell rings" → "rang" (tense-aspect error)	Validated via expert review and participant performance
5	Response Format	Binary judgments (correct/incorrect) + error correction with written justifications	Dual assessment of procedural/declarative knowledge	0-2 scale: 0 (incorrect), 1 (correct judgment only), 2 (correct judgment + fix)	High internal consistency (α = .92)	Cognitive debriefing interviews (92% face validity)	Explicit-implicit knowledge interaction	Written responses within time limit	Correcting "take a decision" to "make a decision" (collocation error)	Effective dual assessment of knowledge types
6	Item Selection Criteria	Excluded low-frequency constructions; prioritized CEFR-aligned errors + L1-L2 collocations	CEFR benchmarks & L1-L2 interference patterns	N/A	Item discrimination >0.40	Expert alignment (CVI = .91)	Real-world error detection relevance	Predefined item pool	Collocation error: "make a decision" vs. "take a decision"	High face validity (92% participant agreement)
7	Scoring Rubric	0-2 scale per item (total 0–80); proficiency tiers: low (0–32), intermediate (33–56), high (57–80)	Differentiation of knowledge types	0 = incorrect, 1 = correct judgment only, 2 = correct judgment + correction	High reliability (α = .92; test-retest r = .86)	Rasch model validity (Infit MnSq)	Quantification of explicit-implicit integration	Applied post-test	Score of 2 for correcting "go" → "goes"	Rubric effectively discriminates proficiency levels
8	Reliability Metrics	Internal consistency (α = .92), test-retest reliability (r = .86), item discrimination indices	Psychometric standards	N/A	Cronbach's α = .92; test-retest r = .86	Pilot testing (n = 30)	Consistency across administrations	Calculated post-hoc	N/A	High reliability supports task robustness
9	Administration Protocol	Strict 25-minute limit under controlled proctoring to simulate implicit processing	Implicit knowledge activation paradigms	Timed responses influence procedural knowledge scoring	Controlled conditions enhance reliability	High internal consistency (α = .92)	Reduction of metalinguistic reflection	Proctored, timed sessions	N/A	Protocol effective in privileging implicit knowledge
10	Factor Analysis	EFA revealed unidimensional structure (KMO = .89; 78.3% variance); Rasch model confirmed fit	Construct validity	N/A	EFA: χ^2 = 1123.47, p < .001; Rasch Infit MnSq = 0.92–1.08	Structural validity via EFA/Rasch	Underlying task construct validity	Analyzed post-data collection	N/A	Task measures a single construct (explicit-implicit integration)
11	Content Validity	Expert review (3 linguists; CVI = .91); alignment with Dynamic Systems Theory	Dynamic Systems Theory	N/A	Expert consensus (CVI = .91)	Thematic alignment with theoretical framework	Task relevance to L2 development	Pre-test validation	N/A	High content validity (CVI = .91)
12	Face Validity	Cognitive debriefing:	Ecological validity	N/A	Participant	Interviews	Ecological	Post-task interviews	Participant	High face validity

No.	Task Component	Description/ Operationalization	Underlying Theoretical Framework	Scoring Methodology	Psychometric Properties	Validation Evidence	Target Constructs	Administration Protocol	Sample Items	Key Findings/ Insights
		92% of participants reported task as reflective of real-world demands			feedback	confirming real-world relevance	validity of task design		quote: "This felt like real error correction I do in English class."	strengthens ecological validity
13	Key Findings	Algorithmic feedback loops enhanced proceduralization (η^2 = .36); high reliability/validity	Dynamic Systems Theory	N/A	η^2 = .36 (large effect size); α = .92; Rasch Infit MnSq	Multiple validation methods (EFA, expert review, cognitive interviews)	Efficacy of NDLLT's algorithmic innovations	Post-analysis	N/A	Significant retention effect (η^2 = .36) validates NDLLT's feedback design

Appendix F

Metalinguistic Awareness Task

No.	Task Component	Description/ Operationalization	Underlying Theoretical Framework	Scoring Methodology	Psychometric Properties	Validation Evidence	Target Constructs	Administration Protocol	Sample Items	Key Findings/ Insights
1	Rule Articulation Demands	Participants verbally explain grammatical correctness of 15 sentence-level stimuli (*e.g.*, conditional clauses). Responses recorded/transcribed.	Skill Acquisition Theory (explicit-to-implicit)	0–3 scale per item (Roehr-Brackin, 2018)	Cronbach's α=.89; Inter-rater κ=.92	CFA: RMSEA=.042, CFI=.971; 68.3% variance explained	Explicit metalinguistic knowledge; proceduralization	20-minute limit; standardized instructions	"Explain why 'If I had known, I would have come earlier' is correct"	High reliability; supports proceduralization hypothesis
2	Dynamic Scaffolding	Progressive hint tiers provided for incomplete responses (*e.g.*, prompting meta-language use).	Nassaji & Fotos (2020) cognitive load framework	Not directly scored; supports response quality	N/A	Expert review (I-CVI=.92)	Cognitive load optimization; explicit knowledge refinement	Integrated during task administration	N/A (hint protocols not itemized)	Reduces cognitive overload; enhances response accuracy
3	Modified Task Items	Six original collocation items replaced with phrasal verbs/article-system targets to address L1 transfer vulnerabilities.	Interface Hypothesis (L1-L2 transfer effects)	Same 0–3 scale	Improved discriminant validity (pilot λ≥.40)	Pilot testing (n=32); expert consensus	Interface structures vulnerable to L1 transfer	Included in 15-item sequence	Phrasal verb/article examples (*e.g.*, "turn up," "a/an")	Enhanced discriminant validity post-modification
4	Scoring Rubric	Granularity criteria (0–3): 0=no rule; 1=partial rule; 2=full rule without meta-language; 3=formal meta-linguistic formulation.	Roehr-Brackin (2018) granularity criteria	24-point composite score (summed item ratings)	Inter-rater κ=.92	Consistent application across coders	Rule explicitness; analytical adaptability	Post-test coding by three trained raters	N/A	High inter-rater reliability (κ=.92)
5	Temporal Controls	Strict 20-minute time limit per test phase to minimize rehearsal effects.	Skill Acquisition Theory (declarative memory)	N/A	Controlled practice effects	Administered under timed conditions	Minimize confounding from rehearsal	Fixed time limits across pretest/posttest/delayed phases	N/A	Ensures measurement of spontaneous knowledge retrieval
6	Concurrent Validity	Significant correlation with TOEFL iBT grammar subscores (r=.74, p<.001).	Criterion-related validity	N/A	r=.74 with TOEFL	TOEFL iBT grammar subscore comparison	Alignment with external proficiency metrics	Administered alongside TOEFL iBT	N/A	Strong evidence of criterion validity
7	Randomized Sequencing	Items counterbalanced and randomized across test phases to mitigate order effects.	Cognitive psychology (order effect mitigation)	N/A	Balanced practice biases	Protocol adherence checks	Unbiased knowledge assessment	Unique sequences per participant per phase	N/A	Mitigated order effects; ensured measurement accuracy
8	Test Phases	Administered at pretest, posttest, and delayed intervals to assess retention and proceduralization.	Skill Acquisition Theory (long-term retention)	N/A	Test-retest reliability	Score trajectories across phases	Long-term knowledge consolidation	Controlled intervals between administrations	Same 15 items across phases	Delayed test scores support retention hypotheses
9	Internal Consistency	High Cronbach's α (.89) indicates strong coherence among items.	Classical Test Theory	N/A	Cronbach's α=.89	Statistical analysis of item correlations	Unidimensional construct validity	N/A	N/A	Items reliably measure the latent construct
10	Inter-Rater Reliability	Three trained coders achieved high consensus (κ=.92) using standardized	Reliability theory	Consensus coding for discrepancies	Cohen's κ=.92	Cross-coder agreement checks	Objective rule explicitness scoring	Post-test coding with trained raters	N/A	Ensures scoring accuracy and consistency

No.	Task Component	Description/ Operationalization	Underlying Theoretical Framework	Scoring Methodology	Psychometric Properties	Validation Evidence	Target Constructs	Administration Protocol	Sample Items	Key Findings/ Insights
		protocols.								
11	Confirmatory Factor Analysis	CFA validated unidimensional structure (RMSEA=.042, CFI=.971, TLI=.963) with 68.3% variance explained.	Structural equation modeling	N/A	RMSEA=.042, CFI=.971, TLI=.963	Statistical validation of construct	Metalinguistic awareness as a single factor	N/A	N/A	Confirms MAT's construct representation
12	Content Validity	Expert review by four applied linguists ensured item relevance (I-CVI=.92).	Content validity theory	N/A	I-CVI=.92	Expert ratings and consensus	Task relevance and appropriateness	Pre-test item selection	Expert-reviewed phrasal verb/article items	High content validity aligns with study goals
13	Composite Score	Summed item-level ratings (0–3) create a 24-point score for overall metalinguistic knowledge.	Aggregate scoring models	0–24 total score	Composite reliability	Correlations with external measures (*e.g.*, TOEFL)	Global explicit knowledge assessment	Calculated post-coding	N/A	Strong predictor of advanced L2 proficiency
14	Target Constructs	Focus on L1-L2 interface structures (*e.g.*, phrasal verbs, articles) and automatized grammatical processing.	Interface Hypothesis; Skill Acquisition Theory	N/A	Improved discriminant validity	Pilot testing and expert consensus	Automatized yet adaptable grammatical processing	Items targeting specific vulnerable structures	"Explain the correct use of 'a/an' in context"	Captures constructs critical for advanced L2 proficiency
15	Hybrid Learning Integration	MAT design operationalizes NDLLT's hybrid algorithms to enhance declarative knowledge proceduralization.	Skill Acquisition Theory (explicit-to-implicit)	N/A	N/A	Supports proceduralization hypothesis	Declarative-to-procedural transition	Embedded in instructional design	N/A	Validates NDLLT's theoretical efficacy

Notes:

- **Theoretical Frameworks**: Directly ties to Skill Acquisition Theory (proceduralization), Nassaji & Fotos (scaffolding), and Roehr-Brackin (scoring granularity).
- **Validation**: Combines statistical (CFA, α, κ) and expert-driven (I-CVI) evidence.
- **Key Insights**: MAT robustly measures explicit metalinguistic knowledge with high reliability/validity, aligning with hybrid learning models targeting L2 automatization.

Appendix G

Self-Regulation Strategies Scale

No.	Focus Area	Item	M/SD/LF	Statistical Insights
1	Self-Monitoring	*I regularly reflect on how AI tools align with my learning priorities.*	M = 3.95, SD = 0.88, LF = 0.76	Strong face validity (CVI = 0.89); loaded on Factor 1 (β = 0.81***).
2	AI Collaboration	*I ask AI systems clarifying questions to improve task outcomes.*	M = 3.52, SD = 0.97, LF = 0.72	Moderate reliability (α = 0.79); correlated with Zimmerman's environmental regulation (*r* = 0.63**).
3	Goal Autonomy	*I use AI insights to prioritize my weekly learning objectives.*	M = 3.38, SD = 1.05, LF = 0.68	Explained 15.2% variance; no cross-loadings (EFA threshold <0.30).
4	Self-Monitoring	*I compare my self-evaluations with AI-generated progress reports.*	M = 4.03, SD = 0.84, LF = 0.81	High discriminant validity (AVE = 0.65); CFA fit (β = 0.83***).
5	Goal Autonomy	*I revise my learning objectives using AI-recommended resources.*	M = 3.29, SD = 1.10, LF = 0.69	Moderate reliability (α = 0.78); no multicollinearity (VIF = 1.32).
6	AI Collaboration	*I adapt my problem-solving approach based on AI critiques.*	M = 3.61, SD = 0.93, LF = 0.74	Significant correlation with MSLQ critical thinking (*r* = 0.69**).
7	AI Collaboration	*I adjust my learning strategies based on AI-generated feedback.*	M = 3.82, SD = 0.91, LF = 0.78	High internal consistency (α = 0.84); linked to Zimmerman's self-monitoring (*r* = 0.68**).
8	Self-Monitoring	*I identify knowledge gaps using AI diagnostic tools.*	M = 3.89, SD = 0.90, LF = 0.77	Cross-validated with MSLQ metacognition (*r* = 0.71**); α = 0.86.
9	Self-Monitoring	*I critically evaluate AI-generated content for relevance to my goals.*	M = 4.02, SD = 0.89, LF = 0.81	High discriminant validity (AVE = 0.62); CFA confirmed unidimensionality (β =

No.	Focus Area	Item	M/SD/LF	Statistical Insights
				0.79***).
10	Goal Autonomy	*I negotiate deadlines with AI systems to balance workload.*	M = 3.17, SD = 1.12, LF = 0.66	Explained 14.8% variance; CVI = 0.85.
11	AI Collaboration	*I integrate AI suggestions into my long-term learning plans.*	M = 3.48, SD = 0.99, LF = 0.70	Factor loading ≥0.65; α = 0.82.
12	Goal Autonomy	*I collaborate with AI tools to set personalized learning goals.*	M = 3.45, SD = 1.02, LF = 0.71	Loaded uniquely on Factor 2 (goal autonomy); explained 21.4% variance.
13	Self-Monitoring	*I use AI dashboards to monitor my engagement levels.*	M = 3.76, SD = 0.95, LF = 0.75	Strong convergent validity (MSLQ self-regulation: *r* = 0.67**).
14	AI Collaboration	*I negotiate task difficulty levels with AI to match my competency.*	M = 3.56, SD = 0.98, LF = 0.73	Explained 18.9% variance; strong face validity (CVI = 0.93).
15	Goal Autonomy	*I reject AI recommendations that conflict with my learning style.*	M = 3.21, SD = 1.08, LF = 0.67	Low multicollinearity (VIF = 1.28); α = 0.77.
16	Self-Monitoring	*I cross-verify AI-generated answers with external resources.*	M = 4.10, SD = 0.86, LF = 0.80	Highest factor loading on self-monitoring (α = 0.89); cross-loadings <0.25.
17	AI Collaboration	*I co-create learning pathways with AI-driven platforms.*	M = 3.40, SD = 1.03, LF = 0.71	Significant correlation with goal autonomy (*r* = 0.65**); α = 0.81.
18	Self-Monitoring	*I feel empowered to modify AI suggestions to better fit my learning needs.*	M = 4.11, SD = 0.87, LF = 0.82	Highest factor loading on self-monitoring (α = 0.89); cross-loadings <0.25.
19	Goal Autonomy	*I use AI analytics to refine my learning milestones.*	M = 3.34, SD = 1.07, LF = 0.68	Moderate reliability (α = 0.76); CFA β = 0.72***.
20	AI Collaboration	*I calibrate AI feedback intensity to match my learning pace.*	M = 3.59, SD = 0.94, LF = 0.74	Explained 19.3% variance; strong discriminant validity (AVE = 0.59).
21	Goal Autonomy	*I balance AI-guided tasks with self-directed learning activities.*	M = 3.27, SD = 1.04, LF = 0.67	Low cross-loadings (<0.30); CVI = 0.88.
22	AI Collaboration	*I use AI analytics to independently track my progress.*	M = 3.67, SD = 0.95, LF = 0.76	Strong convergent validity with MSLQ self-efficacy (*r* = 0.72**).

Appendix H

Perceived Control Over AI Tools Scale

No.	Focus Area	Item	M (SD)	LF	Statistical Insights
1	Co-adaptation mechanics	I can adjust the AI tool's feedback to align with my learning goals.	5.2 (1.1)	0.82	Item-total correlation (r = .79); contributes to 6.8% of variance in autonomy construct.
2	Bidirectional feedback	The AI adapts its recommendations based on my progress patterns.	4.8 (1.3)	0.78	Factor loading (λ = .78); significant cross-loading suppression (<.30) in CFA.
3	Autonomy scaffolding	I feel responsible for directing the AI's role in my learning process.	5.6 (0.9)	0.85	Strongest discriminator (F = 12.4, p < .001) between low/high autonomy clusters.
4	System predictability	The AI system responds predictably to my input modifications.	4.5 (1.4)	0.72	Moderate reliability (α = .87); 5.2% variance explained in trust subscale.
5	Pedagogical trust	I trust the AI's suggestions to improve my language accuracy.	5.1 (1.2)	0.81	High inter-rater agreement (κ = .88) during expert validation.
6	Customization capacity	The system allows me to customize parameters governing AI interactions.	4.3 (1.5)	0.74	Skewness (-0.32) indicates ceiling effect mitigation via reverse-coding.
7	Metacognitive alignment	The AI's feedback helps me identify gaps in my learning strategies.	5.4 (1.0)	0.83	Cronbach's α = .92 if deleted; retained for theoretical completeness.
8	Agency over data	I can modify how the AI collects and uses my learning data.	4.0 (1.6)	0.70	Lowest mean (4.0) reflects interface complexity; flagged for redesign in Phase 2.
9	Goal internalization	The AI tool supports my self-defined objectives rather than	5.7 (0.8)	0.86	Highest factor loading (λ = .86); critical to NDLLT's learner-centricity principle.

No.	Focus Area	Item	M (SD)	LF	Statistical Insights
		imposing external targets.			
10	Error ownership	I feel accountable for correcting errors highlighted by the AI.	5.3 (1.1)	0.80	Significant correlation with L2 gains (r = .63, p < .01) in pilot data.
11	Transparency of logic	The AI explains its reasoning in ways I can understand.	4.7 (1.3)	0.76	VIF = 1.3 confirms absence of multicollinearity with Item 5.
12	Adaptive pacing	I control the speed at which the AI introduces new challenges.	4.9 (1.2)	0.77	Test-retest reliability (r = .85) over 2-week interval.
13	Reciprocal responsiveness	The AI acknowledges my feedback to improve its future suggestions.	4.6 (1.4)	0.73	Moderate floor effect (8%); retained due to centrality to co-adaptation hypothesis.
14	System override capacity	I can override AI decisions without losing access to critical features.	5.0 (1.1)	0.79	Differential item functioning (DIF < .10) across proficiency levels.
15	Collaborative calibration	The AI and I jointly refine strategies based on mutual performance data.	4.4 (1.5)	0.71	Lowest communality (h^2 = .51) but retained for construct breadth.

Appendix I

Anxiety *scale*

No.	Focus Area	Item	M/SD/LF	Statistical Insights
1	Situational Anxiety	I felt overwhelmed when AI adjustments disrupted my task flow. *(Adapted from FLCAS)*	M=3.2, SD=1.1, LF=.78	Item-total correlation (rit) = .71; cross-loadings < .25; STAI convergent r = .69**
2	Situational Anxiety	Real-time AI feedback heightened my stress during grammar exercises. *(New)*	M=2.9, SD=0.9, LF=.82	High discriminant validity ($\Delta\chi^2$ = 12.3, p < .01); ICC test-retest = .83
3	Situational Anxiety	Sudden increases in task complexity caused mental paralysis. *(New)*	M=3.1, SD=1.0, LF=.75	Pilot skewness = -0.12; moderated by cognitive engagement (β = -.33, p = .04)
4	Situational Anxiety	I struggled to recover after the AI system flagged repeated errors. *(Adapted from FLCAS)*	M=2.8, SD=1.2, LF=.73	Explained 14% variance in cognitive load (R^2 = .14, p = .02)
5	Situational Anxiety	Multimodal AI inputs (audio/text) overloaded my working memory. *(New)*	M=3.4, SD=0.8, LF=.81	Factor loading invariance across timepoints (ΔCFI = .002); correlated with EEG alpha-band suppression (r = .58*)
6	Situational Anxiety	Unpredictable peer-AI collaboration made me hesitant to contribute. *(New)*	M=2.7, SD=1.1, LF=.69	Residual covariance < .20; Delphi consensus = 95%
7	Situational Anxiety	The AI's immediate corrections made me hyperaware of mistakes. *(Adapted from FLCAS)*	M=3.0, SD=1.0, LF=.76	STAI-state subscale correlation: r = .63**; item deletion α = .91
8	Anticipatory Anxiety	I worried about appearing incompetent during AI-mediated speaking simulations. *(New)*	M=3.5, SD=0.9, LF=.84	Highest factor loading (λ = .84); ICC = .88; predictive of task avoidance (OR = 1.42, p = .03)
9	Anticipatory Anxiety	I feared negative evaluations from AI-generated performance reports. *(New)*	M=2.6, SD=1.2, LF=.72	Skewness = 1.02; kurtosis = 0.89; moderated by self-efficacy (β = -.41**)
10	Anticipatory Anxiety	Pre-task anxiety spiked when the AI assigned unfamiliar conversational partners. *(New)*	M=3.3, SD=1.0, LF=.79	Cross-lagged path coefficient (β = .38**) with delayed-test scores
11	Anticipatory Anxiety	I doubted my ability to meet AI-curated proficiency targets. *(Adapted from FLCAS)*	M=2.9, SD=1.1, LF=.77	Residual variance = .39; correlated with cortisol levels (r = .51*)
12	Anticipatory Anxiety	Anticipating neurofeedback-driven task shifts disrupted my	M=3.1, SD=0.8, LF=.81	Item reliability (ω = .85); accounted for 18% variance in syntactic complexity

No.	Focus Area	Item	M/SD/LF	Statistical Insights
		focus. *(New)*		(R^2 = .18)
13	Anticipatory Anxiety	I felt unprepared for AI's dynamically generated vocabulary challenges. *(New)*	M=2.8, SD=1.3, LF=.74	Differential item functioning (DIF) nonsignificant across age groups (p = .12)
14	Anticipatory Anxiety	Anxiety about algorithmic bias in error detection affected my participation. *(New)*	M=3.0, SD=1.0, LF=.68	Marginal reliability (ρ = .72); flagged for linguistic clarity in Delphi review
15	Neurocognitive Strain	Prolonged neuroadaptive exercises left me mentally exhausted. *(New)*	M=3.6, SD=0.7, LF=.83	Strongest predictor of delayed-test scores (β = .47**); skewness = -0.82
16	Neurocognitive Strain	Post-session cognitive fatigue impaired my retention of new syntax rules. *(New)*	M=3.4, SD=0.9, LF=.79	Moderated mediation effect (95% CI [.12, .38]); correlated with decreased hippocampal activation (fMRI: r = -.61*)
17	Neurocognitive Strain	My mind felt "blank" after intensive AI-driven translation drills. *(Adapted from FLCAS)*	M=2.5, SD=1.1, LF=.71	Item response theory (IRT) discrimination = 1.82; STAI-divergent (r = .09, ns)
18	Neurocognitive Strain	Cross-domain transfer tasks (*e.g.*, math→language) induced cognitive overload. *(New)*	M=3.2, SD=1.0, LF=.76	Multigroup CFA invariance (ΔRMSEA = .008); linked to theta-gamma EEG coupling (r = -.53*)
19	Neurocognitive Strain	I experienced mental "numbness" during high-stakes AI assessments. *(New)*	M=3.7, SD=0.6, LF=.85	Explained 22% variance in dropout intent (R^2 = .22**); factor determinacy = .93
20	Neurocognitive Strain	Sustained attention to decentralized AI prompts drained my motivation. *(New)*	M=2.9, SD=1.2, LF=.74	Test information function peak at θ = 1.3; differential reliability = .89

Key:

- M = Mean (1–5 Likert), SD = Standard Deviation, LF = Standardized Factor Loading (CFA)
- STAI = State-Trait Anxiety Inventory; ICC = Intraclass Correlation Coefficient; IRT = Item Response Theory
- p-values: *< .05, *< *.01; NS = nonsignificant; Δ = change; OR = Odds Ratio; CI = Confidence Interval*

Psychometric Notes:

- All items demonstrated Cronbach's α > .90 when deleted.
- Composite reliability (ω) = .93; Average Variance Extracted (AVE) = .62.
- Multidimensional Random Coefficients Model (MRCMLM) confirmed absence of local dependence (LD χ^2 < 3.84).
- Exploratory Structural Equation Modeling (ESEM) supported configural invariance across pretest/posttest/delayed administrations (ΔCFI = .007).

Appendix J

Motivation scale

No	Focus Area	Item	M/SD/LF	Statistical Insights
1	Intrinsic Motivation	I found joy in overcoming AI-curated linguistic challenges.	5.8/1.2/.89	Highest factor loading (.89); IRT a = 2.3
2	Intrinsic Motivation	Engaging with AI-generated tasks sparked my curiosity to learn more.	5.5/1.1/.78	Strong discriminant validity (r = -.52 vs. anxiety)
3	Intrinsic Motivation	Solving complex language puzzles designed by the AI felt personally rewarding.	5.6/1.3/.85	Test-retest ICC = .89; α = .89
4	Intrinsic Motivation	I looked forward to interacting with novel AI-driven language activities.	5.3/1.4/.81	IRT a = 1.9; no floor/ceiling effects
5	Intrinsic Motivation	The unpredictability of AI challenges enhanced my sense of accomplishment.	5.4/1.2/.76	Convergent validity r = .81 (AMS)
6	Intrinsic Motivation	AI-tailored content deepened my intrinsic interest in language mastery.	5.7/1.1/.82	CVI = .96; metric invariance (ΔCFI = .006)
7	Extrinsic Goal Alignment	Advancing in this program will enhance my career prospects.	6.1/0.9/.86	Highest extrinsic loading (.86); SE = 0.24
8	Extrinsic Goal	Completing AI-driven modules strengthened	5.9/1.0/.79	Skewness/kurtosis ≤ 0.95

No	Focus Area	Item	M/SD/LF	Statistical Insights
	Alignment	my professional language skills.		
9	Extrinsic Goal Alignment	I value how this program's certifications are recognized in my industry.	5.7/1.3/.74	α = .86; R^2 = .41 (path analysis)
10	Extrinsic Goal Alignment	AI-curated progress reports helped me track career-relevant competencies.	5.4/1.4/.80	38% variance from personalization fidelity (p < .01)
11	Extrinsic Goal Alignment	Mastering these skills through AI will improve my job market competitiveness.	5.8/1.1/.77	IRT a = 1.8; β = .63 (mediation)
12	Extrinsic Goal Alignment	The program's structure aligns with my external professional benchmarks.	5.5/1.2/.73	TLI = .95; power (1-β) = .95
13	Self-Regulatory Capacity	I adapted my strategy when the AI flagged persistent errors.	6.0/1.0/.91	Modified SRQ; highest self-regulation loading (.91)
14	Self-Regulatory Capacity	I adjusted my study schedule based on algorithmically identified weaknesses.	5.7/1.3/.88	α = .91; ICC = .89
15	Self-Regulatory Capacity	AI feedback helped me prioritize areas needing regulatory attention.	5.6/1.1/.85	RMSEA = .049; skewness = -0.15
16	Self-Regulatory Capacity	I revised my approach when the system detected inefficient patterns.	5.8/1.2/.89	CFI = .97; 72.4% variance explained
17	Self-Regulatory Capacity	Algorithmic progress tracking increased my persistence through difficulties.	5.5/1.4/.82	ESEM invariance confirmed
18	Self-Regulatory Capacity	I systematically monitored improvement using AI-generated dashboards.	5.9/1.0/.84	IRT a = 2.4; omitted social comparison items
19	Neurocognitive Engagement	Real-time neurofeedback heightened my focus during semantic tasks.	5.2/1.5/.88	Aligns with CATL; IRT SE = 0.31
20	Neurocognitive Engagement	The AI's cognitive load optimization improved my mental clarity.	5.4/1.3/.79	α = .88; working memory modulation (CATL)
21	Neurocognitive Engagement	Neural oscillation displays during tasks amplified my cognitive effort.	5.1/1.6/.75	CVI = .96; CFA-validated (.75)

Appendix K

Feedback survey items

No	Focus Area	Item	M/SD/LF	Statistical Insights
1	Perceived Effectiveness	The NDLLT improved my confidence in applying new skills.	5.2/1.1/.82	CFA λ = .82 (p<.001); item-total r = .68; PCA loading = .79; AVE = .61
2	AI Feedback Dynamics	AI-generated feedback helped me refine my problem-solving strategies.	4.8/1.3/.76	Composite ω = .78; cross-loading = .22; predictive validity r = .39 (p<.01)
3	System Engagement Barriers	Technical glitches disrupted my learning flow.	3.4/1.5/.71	Reverse-coded (adj. R = -.53); inter-item r = .41; Cronbach's α = .71
4	Emotional/Motivational States	Adaptive tasks reduced my anxiety during complex challenges.	5.6/0.9/.88	Factor loading = .88 (SE = .04); 65% variance explained (subscale); divergent validity r = -.12 (p=.18)
5	Human-AI Synergy	I felt in control when overriding AI-suggested task sequences.	4.1/1.4/.69	CFA χ^2/df = 1.93; RMSEA = .05; SRMR = .06; inter-subscale correlation r = .34 (p<.05)
6	Perceived Effectiveness	The intervention enhanced my motivation to persist through setbacks.	5.4/1.0/.85	Item-total r = .72; PCA communality = .65; reliability ω = .80
7	AI Feedback Dynamics	Real-time AI adjustments matched my learning pace.	5.0/1.2/.81	Multigroup CFA invariance (ΔCFI = .002); 63% variance (factor); inter-rater κ = .79
8	System Engagement Barriers	Cognitive overload limited my engagement with NDLLT modules.	2.9/1.6/.64	Residual variance = .48; modification index = 3.2; skewness = 1.4 (SE = 0.3)
9	Emotional/Motivational States	AI-driven tasks triggered frustration due to rapid difficulty shifts.	3.8/1.7/.58	Negative wording effect (adj. β = -.21); CFA SRMR = .05; multicollinearity VIF = 1.8
10	Human-AI Synergy	Algorithmic task sequencing aligned	4.5/1.3/.73	Partial η^2 = .12 (ANOVA); factor

No	Focus Area	Item	M/SD/LF	Statistical Insights
		with my personal learning goals.		correlation Φ = .51; test-retest ICC = .83
11	Perceived Effectiveness	The NDLLT enhanced my ability to retain new information long-term.	5.3/1.0/.83	CFA λ = .83 (p<.001); item-total r = .70; AVE = .62
12	Perceived Effectiveness	I can apply skills learned through NDLLT in diverse real-world contexts.	5.1/1.2/.79	Composite ω = .81; predictive validity r = .41 (p<.01)
13	Perceived Effectiveness	The intervention improved my ability to self-assess learning progress.	5.0/1.1/.80	PCA loading = .78; Cronbach's α = .79
14	Perceived Effectiveness	NDLLT's structure facilitated deeper understanding of complex concepts.	5.5/0.8/.87	Factor loading = .87 (SE=.03); 67% variance explained
15	AI Feedback Dynamics	AI feedback provided actionable steps for skill improvement.	4.9/1.3/.77	Cross-loading = .18; composite ω = .76
16	AI Feedback Dynamics	The AI's suggestions were contextually relevant to my learning needs.	4.7/1.4/.74	Multigroup CFA invariance (ΔCFI = .003); inter-rater κ = .75
17	AI Feedback Dynamics	Personalized feedback timing optimized my learning absorption.	5.2/1.1/.80	Predictive validity r = .37 (p<.05); AVE = .59
18	System Engagement Barriers	Unintuitive interface design slowed my progress.	3.2/1.6/.66	Reverse-coded (adj. R = -.49); inter-item r = .38
19	System Engagement Barriers	Lack of offline access hindered consistent participation.	3.0/1.7/.62	Skewness = 1.5 (SE=0.3); residual variance = .51
20	System Engagement Barriers	Overly frequent notifications disrupted concentration.	3.5/1.5/.68	Modification index = 4.1; Cronbach's α = .69
21	Emotional/Motivational States	Progress visualizations increased my sense of accomplishment.	5.4/0.9/.85	Factor loading = .85 (SE=.04); divergent validity r = -.10 (p=.22)
22	Emotional/Motivational States	Sudden difficulty spikes eroded my confidence.	3.7/1.6/.60	Negative wording effect (adj. β = -.25); VIF = 1.9
23	Emotional/Motivational States	Gamified elements made challenging tasks enjoyable.	5.7/0.7/.89	70% variance explained; item-total r = .75
24	Emotional/Motivational States	Unpredictable AI behavior caused intermittent stress.	3.9/1.4/.63	CFA SRMR = .06; inter-subscale r = -.31 (p<.05)
25	Human-AI Synergy	Collaborative AI adjustments respected my learning preferences.	4.3/1.3/.71	Factor correlation Φ = .48; test-retest ICC = .80
26	Human-AI Synergy	I trusted the AI's recommendations during critical tasks.	4.6/1.2/.75	Partial η² = .14 (ANOVA); composite ω = .77
27	Human-AI Synergy	Customization options bridged AI logic with my intuition.	4.4/1.4/.70	PCA communality = .63; reliability ω = .78
28	Human-AI Synergy	The system's explainability features fostered algorithmic trust.	4.2/1.5/.67	RMSEA = .06; SRMR = .07; AVE = .54

Appendix L

Interview

No.	Focus Area	Question	Sample Responses from Participants	M/ SD/ LF	Reported Challenges & Affordances from the NDLLT-Based Intervention	Extracted Themes & Thematic Analysis (Braun & Clarke, 2006)	Theoretical & Pedagogical Insights
1	Perceived Effectiveness	How effective did you find the NDLLT in improving your speaking skills?	"It helped me speak more fluently." / "I felt less nervous over time."	M=5.8/ SD=1.1	Initial anxiety reduced by AI-guided scaffolding.	Learner confidence; reduced anxiety; fluency gains	AI scaffolding promotes confidence-building and fluency development.
2	Perceived Effectiveness	How did NDLLT impact your writing complexity and accuracy?	"I noticed my sentences became more	M=6.1/ SD=0.9	Increased syntactic complexity through adaptive feedback.	Enhanced structural awareness; accuracy enhancement	Adaptive feedback fosters syntactic awareness and

No.	Focus Area	Question	Sample Responses from Participants	M/ SD/ LF	Reported Challenges & Affordances from the NDLLT-Based Intervention	Extracted Themes & Thematic Analysis (Braun & Clarke, 2006)	Theoretical & Pedagogical Insights
			structured."				precision.
3	Perceived Effectiveness	How effective was NDLLT in enhancing your listening comprehension?	"I could understand faster-paced audio."	M=5.9/ SD=1.0	Improved comprehension with adaptive pacing.	Listening fluency; adaptive strategies	Adaptive pacing improves auditory processing and comprehension.
4	Perceived Effectiveness	How did NDLLT influence your reading speed and comprehension?	"I read faster and understood better."	M=6.0/ SD=0.8	Improved reading fluency through targeted exercises.	Reading fluency; comprehension gains	Targeted exercises promote fluency and deep comprehension.
5	AI Feedback Dynamics	How useful was the feedback provided by the AI during listening exercises?	"The feedback was immediate and helpful." / "Sometimes it felt generic."	M=5.5/ SD=1.3	Immediate feedback improved comprehension but lacked personalization.	Immediacy vs. personalization	Tailored feedback could enhance comprehension further.
6	AI Feedback Dynamics	How did AI feedback influence your ability to self-correct errors?	"I became better at spotting my mistakes."	M=6.0/ SD=1.0	Feedback improved error recognition and correction.	Error-awareness; self-correction	AI feedback enhances metalinguistic awareness and autonomy.
7	Motivation	Did the adaptive task sequencing keep you motivated throughout the sessions?	"Yes, it felt challenging but not overwhelming. "	M=6.0/ SD=1.0	Sustained motivation due to optimal challenge levels.	Engagement through adaptive adjustment	Adaptive sequencing aligns with self-determination theory.
8	Motivation	How did NDLLT impact your overall motivation to learn the language?	"It made learning more engaging."	M=6.2/ SD=0.9	Increased engagement through gamified elements.	Motivation boost; gamification effects	Gamified elements enhance intrinsic motivation.
9	Neurocognitive Alignment	How did you feel about the cognitive demands of the tasks?	"Some tasks were mentally exhausting but rewarding."	M=5.2/ SD=1.4	High cognitive load balanced by perceived learning gains.	Cognitive load vs. learning efficiency	Tasks should balance demands to optimize neuroplasticity.
10	Neurocognitive Alignment	Did you notice changes in how you processed information over time?	"I felt I could process tasks faster."	M=5.8/ SD=1.2	Neuroplasticity markers indicated improved task efficiency.	Proceduralization; neural adaptation	Task repetition fosters procedural memory and neuroplasticity.
11	Emotional Responses	How did you feel emotionally during the AI-driven tasks?	"I felt anxious at first but more confident later."	M=5.7/ SD=1.2	Anxiety reduced over time with adaptive support.	Emotional adaptation; confidence-building	Emotional scaffolding is critical for sustained engagement.
12	Emotional Responses	How did your emotions influence your performance during the intervention?	"When I was anxious, I made more mistakes."	M=5.3/ SD=1.4	Emotional states influenced performance variability.	Anxiety-performance interplay	Emotional regulation strategies are essential for consistency.
13	Metacognitive Adaptation	How did you adapt your strategies based on AI feedback?	"I started to plan better after seeing my errors."	M=5.9/ SD=1.0	Learners improved metacognitive awareness through iterative feedback.	Strategy refinement; self-regulation	AI-driven feedback enhances metacognitive skills.
14	Metacognitive Adaptation	Did NDLLT help you become more aware of your learning	"Yes, I know what I need to work on now."	M=6.1/ SD=0.8	Increased awareness of strengths and	Self-awareness; targeted improvement	NDLLT fosters self-directed learning strategies.

No.	Focus Area	Question	Sample Responses from Participants	M/ SD/ LF	Reported Challenges & Affordances from the NDLLT-Based Intervention	Extracted Themes & Thematic Analysis (Braun & Clarke, 2006)	Theoretical & Pedagogical Insights
		strengths/weaknesses?			weaknesses.		
15	Retention	How well did you retain the skills learned in previous sessions?	"I remembered most of it, especially vocabulary."	M=6.2/ SD=0.8	Spaced repetition aided retention.	Spaced practice; long-term retention	Spacing effects are crucial for retention.
16	Retention	Did NDLLT help you apply what you learned in new contexts?	"I could use it in real conversations. "	M=6.0/ SD=1.0	Improved transfer to real-world scenarios.	Transferability; contextual application	NDLLT supports authentic skill application.
17	Human-AI Synergy	How did you perceive the balance between AI guidance and your autonomy?	"The AI guided me but also let me make choices."	M=6.0/ SD=1.1	Balance between AI control and learner autonomy was well-received.	Algorithmic agency vs. autonomy	Optimal AI guidance supports autonomy without over-dependence.
18	Human-AI Synergy	Did you feel the AI adapted to your individual learning needs?	"Yes, it felt personalized to me."	M=6.1/ SD=0.9	Personalized adaptations improved engagement.	Personalization; learner-centered design	Adaptive AI enhances individualized learning experiences.
19	Neuroplasticity	Did you notice any changes in how you approached tasks over time?	"I developed better strategies for complex tasks."	M=5.8/ SD=1.2	Neuroplasticity markers indicated strategy optimization.	Strategy optimization; cognitive efficiency	Task repetition fosters cognitive flexibility.
20	Neuroplasticity	How did the intervention impact your ability to multitask in the language?	"I became better at switching between tasks."	M=5.7/ SD=1.2	Improved multitasking through cognitive flexibility.	Cognitive flexibility; multitasking	NDLLT enhances multitasking via neuroplasticity-driven design.
21	Affective States	How did your emotions evolve across the intervention?	"I felt more confident as I progressed."	M=5.9/ SD=1.1	Confidence increased with task familiarity.	Emotional growth; confidence-building	Emotional adaptation supports sustained engagement.
22	Affective States	How did your emotional state affect your engagement with the tasks?	"When I was frustrated, I disengaged."	M=5.4/ SD=1.3	Frustration led to temporary disengagement.	Emotional regulation; task engagement	Regulation strategies are critical for engagement.
23	Cognitive Load	Did you feel the tasks were appropriately challenging?	"They were challenging but manageable."	M=5.8/ SD=1.0	Optimal challenge levels sustained engagement.	Challenge calibration; cognitive engagement	Balancing difficulty maximizes cognitive engagement.
24	Cognitive Load	How did you manage cognitive demands during the tasks?	"I broke tasks into smaller steps."	M=5.6/ SD=1.2	Learners developed cognitive load management strategies.	Cognitive strategies; load management	Explicit training in load management benefits learners.
25	Task Sequencing	Were the tasks sequenced in a way that supported your learning?	"Yes, they built on each other well."	M=6.2/ SD=0.8	Sequencing facilitated cumulative learning.	Sequential scaffolding; skill consolidation	Effective sequencing scaffolds skill development.
26	Task Sequencing	Did you feel the pacing of the tasks matched your learning speed?	"It felt just right for me."	M=6.1/ SD=0.9	Adaptive pacing aligned with individual progress.	Pacing; learner-centered design	Adaptive pacing ensures personalized learning trajectories.

No.	Focus Area	Question	Sample Responses from Participants	M/ SD/ LF	Reported Challenges & Affordances from the NDLLT-Based Intervention	Extracted Themes & Thematic Analysis (Braun & Clarke, 2006)	Theoretical & Pedagogical Insights
27	Transferability	How well could you apply what you learned to real-world scenarios?	"I used it in conversations outside class."	M=6.0/ SD=1.0	Improved real-world application of skills.	Real-world application; practical transfer	NDLLT bridges classroom learning and authentic use.
28	Transferability	Did you feel confident using the language outside the intervention?	"Yes, I felt more confident speaking."	M=6.3/ SD=0.8	Confidence in real-world language use increased.	Confidence; authentic application	NDLLT boosts practical language confidence.

Appendix M

Data Analysis

Research Question	Primary Analysis Method	Key Assumptions & Diagnostics	Statistical Procedures & Adjustments	Effect Size & Sensitivity Analyses
1.	MANCOVA	• Homogeneity of regression slopes (Group × Covariate interactions: *all ps* > .05) • Multivariate normality: Mardia's skewness (γ=2.14, p=.11), kurtosis (γ=4.67, p=.09); Q-Q plots • Homogeneity of covariances: Box's *M* (p=.14); Roy-Bargmann stepdown verification	• Pretest scores as covariates • Omnibus: Pillai's trace (robust to heterogeneity/unequal n) • Post hoc: Univariate ANCOVAs with Bonferroni α=.0016	• Partial η² (multivariate) • Hedges' g (pairwise) • Monte Carlo simulations (10k iterations) • Bootstrap bias-corrected CIs
2.	SEM-Mediation/Moderation	• Missing data: Little's MCAR test (χ^2=18.34, p=.24) • Multicollinearity: *VIF* < 3.0 • Residual independence: Durbin-Watson=1.8–2.1	• Two-stage: (1) CFA (robust WLS estimation for latent constructs) (2) SEM with FIML • fMRI preprocessing: FLIRT spatial normalization • Wavelet coherence (theta-gamma coupling)	• Standardized path coefficients • Indirect effects via bias-corrected bootstrapping
3	Mixed-design MANCOVA	• Sphericity: Greenhouse-Geisser ε=0.92 • Covariance structure: AR1 (autoregressive)	• Within-subjects factor: Time (posttest vs. delayed posttest) • Dynamic Causal Modeling (DCM) for fMRI effective connectivity	• Time × Group interaction effects • Effective connectivity parameters (DCM)
Integration	Methodological Triangulation	• Temporal concordance: Cross-correlation fMRI activation × cognitive load (r=−.71, p<.001)	• Joint display analysis • Grounded theory coding • Hierarchical alignment:	• Quantitative-qualitative isomorphism (*e.g.*, η²=.925 ↔ 87% code saturation) • Theoretical fidelity mapping

Research Question	Primary Analysis Method	Key Assumptions & Diagnostics	Statistical Procedures & Adjustments	Effect Size & Sensitivity Analyses
			MANCOVA η^2 ↔ qualitative code saturation	

Note. MANCOVA = Multivariate Analysis of Covariance; SEM = Structural Equation Modeling; CFA = Confirmatory Factor Analysis; WLS = Weighted Least Squares; FIML = Full Information Maximum Likelihood; FLIRT = FMRIB's Linear Image Registration Tool; AR1 = First-Order Autoregressive Structure; DCM = Dynamic Causal Modeling; VIF = Variance Inflation Factor; MCAR = Missing Completely at Random; CI = Confidence Interval; η^2 = partial eta-squared. All analyses controlled for pretest disparities via continuous covariates. Neurophysiological metrics underwent wavelet coherence and fMRI preprocessing pipelines.

Appendix N

Comparative Outcomes of AI-Driven Learning Interventions

Theme	Group/Data Source	Key Qualitative Findings	Participant Agreement Rate	Quantitative Correlates	Statistical Significance	Strengths	Weaknesses
Neurocognitive-Emotional Alignment	Interview Participants (Outstanding Tier)	Reduced anxiety through predictive processing alignment (left IFG activation)	83% (24/29)	*fNIRS Z* = 4.21, *SD* = 0.87	FEW p = 0.003	Enhanced neural efficiency (θ-γ coupling)	Cognitive overload (β = -0.33, p = 0.04)
Feedback Dynamics	Survey Respondents (Proficient Tier)	Bidirectional feedback improved error correction (AUC = 0.91)	89% (37/42)	Cronbach's α = 0.89, *SD* = 0.12	r = 0.63, p < 0.01	High immediacy (TRP delays < 200ms)	Generic phrasing critiques (22%)
Metacognitive Adaptation	Delayed Test Cohort	Customized AI strategies enhanced retention (η^2 = 0.36)	68% (28/41)	Retention *M* = 41/50, *SD* = 1.2	r = 0.87, *SEM* = 3.4	Self-regulation (*M* = 4.11, *SD* = 0.87)	Interface complexity (*M* = 3.2, *SD* = 1.6)
Human-AI Co-Regulation	Needs Improvement Tier	Algorithmic mistrust correlated with syntactic rigidity (MATTR = 0.62)	41% (12/29)	κ = 0.86, *SD* = 0.05	*B* = 1.33, *SE* = 0.07	Stigmergic collaboration (MTLD = 72.1)	Emotional strain (*M* = 3.9, *SD* = 1.4)
Transferability	Posttest Participants	Real-world application confidence (LSA = 0.79)	79% (31/39)	TOEFL r = 0.74, *SD* = 0.09	p < 0.001	Contextual fluency (dialogic alignment = 0.71)	Sporadic arousal dysregulation (*M* = 3.0, *SD* = 1.0)

Key Table Features:

1. **Triangulation**: Integrates qualitative themes (Appendix L interviews) with neuroimaging (Appendix D, Item 15), psychometric (Appendix C, Item 12), and algorithmic metrics (Appendix D, Item 9).
2. **Statistical Rigor**: APA notation for means (*M*), standard deviations (*SD*), effect sizes (η^2), and significance (p).
3. **NDLLT Alignment**: Themes map to theoretical pillars (*e.g.*, predictive processing, decentralized adaptation).
4. **Participant Stratification**: Groups segmented by proficiency tiers (Appendix C) and intervention phases (Appendix E).
5. **Weakness Identification**: Technical (interface complexity) and affective (cognitive overload) limitations quantified.

https://doi.org/10.14505/jres.v16.2(20).06

Multifaceted Factors Contributing to the Closure of Valley Forge Military Academy an Examination of Management, Economic, Legal, and Social Influences

Walter ROSADO
The Citadel Military College of South Carolina, USA
ORCID: 0009-0009-5125-4688
wrosado@citadel.edu

Article info: Received 9 November 2025; Revised 18 November 2025; Accepted 2 December 2025; Published 30 December 2025.

Abstract: This paper examines the rise and eventual closure of Valley Forge Military Academy (VFMA), a once-prominent American private military educational institution. Founded in 1928 and long celebrated for its unique Sandhurst-inspired traditions, VFMA faced escalating pressures from declining enrollment, rising operational costs, legal challenges, and shifting societal attitudes toward military education. Utilizing a mixed-methods approach, including alumni and faculty interviews and quantitative survey data, the study identifies key factors behind the academy's decline, such as governance deficiencies, unsustainable financial practices, deferred maintenance, and reputation-damaging legal controversies. Drawing on theories of organizational failure and resilience, the analysis reveals how intertwined micro-level mismanagement, meso-level industry pressures, and macro-level external threats created a cumulative crisis that overwhelmed efforts at adaptation. Ultimately, VFMA's experience underscores the vulnerability of small, tuition-dependent institutions to market legal, and cultural forces, offering broader lessons for organizational sustainability and resilience in the educational sector.

Keywords: Military Academy; education; organizational decline; management; boarding school.

JEL Classification: I20; A12; A23; O32.

Introduction

In 1928, an all-male private boarding and day school named Valley Forge Military Academy (VFMA) was established in Wayne, Pennsylvania, operating for several decades as a prominent institution within the American educational landscape. In the 1940s, VFMA made the unique decision to model rank structure, drills, customs, and ceremonies on the British Army (Sandhurst) model, rather than the American model, which is popularized at colleges such as the U.S. Military Academy at West Point, The Citadel, and Virginia Military Institute. VFMA's alumni include General Norman Schwarzkopf, General H.R. McMaster, other distinguished military leaders, professional football player Larry Fitzgerald Jr., and J.D. Salinger, an acclaimed author, underscoring the school's historical capacity to attract and cultivate individuals who have achieved significant recognition in diverse fields. In the era following the Vietnam War, when other Military Academies were shuttering their doors, VFMA sustained itself. In 1981, the movie 'TAPS', starring Tom Cruise, Sean Penn, and George C. Scott, was released. The movie was filmed at, and featured VFMA. Despite its storied past and notable graduates, and despite an April 2025 pledge of $3 million from Alumnus Harry C. McCreary, Jr. and other sizable financial gifts, VFMA eventually faced insurmountable challenges leading to its closure, a trajectory mirroring that of other venerable institutions grappling with shifts in societal values and educational paradigms (Curry, 1972).

This paper delves into the complex interplay of management, economic, legal, and social factors that collectively contributed to VFMA's eventual closure, illustrating the multifaceted pressures impacting higher educational institutions in competitive markets (Paul, 2005). The investigation will particularly focus on how evolving market dynamics, institutional governance decisions, regulatory compliance demands, and changing societal expectations of military education ultimately culminated in the cessation of its operations. This analysis aims to provide a comprehensive framework for understanding the mechanisms of organizational failure within an educational context, drawing parallels with broader theories of business collapse (Higashi *et al.* 2020).

1. Literature Review

This review synthesizes existing scholarship on organizational failure, drawing upon frameworks that examine financial, legal, and organizational determinants of business closure to illuminate the complex forces at play in educational institutional decline (Zaman, 2023). Specifically, it will explore how macro-level factors, such as economic crises and shifts in taxation systems, alongside meso-level influences like competitive pressures and market rumors, can precipitate institutional demise (Higashi *et al.* 2020). Furthermore, micro-level issues, including mismanagement, insufficient strategic planning, and poor resource allocation, also contribute significantly to the vulnerability of educational establishments (Kelchen *et al.* 2024). The confluence of these elements, external environmental pressures, industry-specific challenges, and internal operational deficiencies, creates a precarious situation for organizations, often culminating in their dissolution (Samdanis & Lee, 2018). Moreover, hybrid organizations, such as military academies, face unique legitimation challenges in balancing competing institutional demands, which can lead to misaligned legitimation strategies and ultimately organizational failure (Siwale *et al.* 2021). This analysis, therefore, adopts a holistic approach, considering both endogenous weaknesses and exogenous pressures that collectively contribute to an organization's demise (Hager *et al.* 1996; Amankwah-Amoah *et al.* 2023). The dynamic interplay of these factors often necessitates a nuanced understanding of institutional resilience and susceptibility in navigating complex operational environments. Building upon this, an understanding of the evolutionary process of organizational failure, which is rarely a singular event but rather a dynamic unfolding over time, is crucial for comprehensively analyzing the closure of institutions like Valley Forge Military Academy (Amankwah-Amoah *et al.* 2023). Such an understanding often involves examining various theoretical lenses, including organizational ecology, industry life cycle theory, and organizational psychology, to fully grasp the internal and external forces driving decline (Fulop *et al.* 2023). This multi-level analytical approach, encompassing macrosystem, mesosystem, and microsystem factors, is essential for a robust examination of organizational failure in complex entities like military academies (Higashi *et al.* 2020). Specifically, this paper integrates insights from studies on organizational resilience in smaller, private non-profit higher educational institutions to explore how factors such as goal-directed solution seeking and avoidance may influence an institution's capacity to sustain effectiveness amidst decline (Moran, 2016). This approach acknowledges that organizational failure is often a symbiotic process between external and internal factors, with macro-level influences capable of independently driving institutional decline (Higashi *et al.* 2020). In this context, macro-level variables are external to an organization and, while common to all businesses, pose particular threats to entities less resilient to market shocks or long-term profitability squeezes (Higashi *et al.* 2020). Conversely, meso-level factors encompass industry-specific challenges, competitive dynamics, and sector-wide reputation issues that can erode an organization's standing and viability (Higashi *et al.* 2020). Finally, micro-level factors refer to internal organizational issues such as management inefficiencies, inadequate resource allocation, and a lack of strategic foresight that compromise an institution's operational integrity and long-term sustainability (Higashi *et al.* 2020). Within this multi-level framework, organizational resilience, defined as the capacity to adapt and thrive amidst persistent threats, becomes a critical determinant of survival, with a strong middle-tier infrastructure contributing significantly to stability and resource management (You, 2023; Higashi *et al.* 2020). A cross-scale perspective is invaluable for understanding organizational resilience, as the viability of organizations is intricately linked to the broader social-ecological systems in which they are embedded (Williams *et al.* 2019). Such a multilevel perspective is crucial for developing robust strategies for resilience, facilitating the identification of organizational strengths, weaknesses, and potential redundancies across diverse subcomponents (Wood *et al.* 2018).

2. Methodology

Interviews with members of the VFMA Alumni Association, Parents, and former Faculty were conducted. Thematic analysis was applied to transcripts of these interviews to identify recurring patterns and significant narratives related to the institution's operational decline (Higashi *et al.* 2020). A survey was also distributed among alumni and parents to gather quantitative data on perceptions of management effectiveness, financial stability, and academic quality over time.

The methodology for this study employed a mixed-method design centered on stakeholder perspectives to investigate the factors contributing to the closure of Valley Forge Military Academy. A total of 64 participants were recruited based on clearly defined inclusion criteria, specifically, individuals who were either teachers, alumni, or parents of former or current cadets, ensuring firsthand familiarity with institutional operations and culture. Data collection occurred through two primary modes: structured online surveys administered via Google Forms and follow-up semi-structured interviews conducted through Microsoft Teams. The questionnaire consisted of 30

multiple-choice items and 4 open-ended essay questions. An interview protocol, developed by the primary researcher using the Harvard University Program on Survey Research (PSR) Tip Sheet as a guiding framework, ensured consistency, clarity, and methodological rigor across all participant interactions.

Data analysis followed a structured thematic analysis approach. The process began with repeated readings of the raw interview transcripts and survey responses to establish deep familiarity with the dataset, followed by the generation of initial codes reflecting salient ideas and recurring patterns. These codes were then iteratively clustered into preliminary themes, which were refined through multiple rounds of review to ensure coherence, distinctiveness, and alignment with the overarching research questions. Credibility was strengthened through researcher triangulation, member checking with participants who volunteered to review interpretations, and the maintenance of a detailed audit trail documenting analytic decisions. Ethical considerations were prioritized throughout, including securing informed consent, ensuring confidentiality- especially vital given the hierarchical and tradition-bound culture of a military academy- and mitigating power dynamics by emphasizing voluntary participation and the right to withdraw at any time. The study also explicitly acknowledges limitations such as self-selection bias, possible constraints on candor due to institutional loyalty or concern for reputational harm, and uneven representation across stakeholder groups. These limitations are carefully considered in interpreting the findings and framing their broader implications.

3. Results

This mixed-methods approach allowed for a comprehensive understanding of the factors contributing to VFMA's closure, triangulating qualitative insights with quantitative data to bolster the validity of the findings. The results elucidate the interplay between internal governance challenges and external market pressures that ultimately undermined the academy's operational viability and led to its closure (Higashi *et al.* 2020; Frisbie & Converso, 2016). Many of the respondents were noticeably irritated by the academy's pending closure. Some respondents referred to issues that could not be validated, such as specific debt amounts or the total number of scholarship dollars issued to students. This comprehensive analysis aims to provide actionable insights for other educational institutions facing similar challenges, emphasizing the critical role of adaptive strategies and robust organizational resilience in navigating periods of significant adversity (Ouédraogo & Boyer, 2012; Abdullahi *et al.* 2023). Furthermore, the study aims to contribute to the broader literature on organizational decline and resilience, particularly within the context of private military educational institutions (Fernandez & Burnett, 2020; Heredia *et al.* 2022).

4. Discussion

The specific themes that respondents cited for the closure of Valley Forge Military Academy included managerial issues, economic pressures, and changes in societal values towards military education.

Based on data derived from surveys and interviews with parents, alumni, former cadets, faculty, and administrative staff, several interconnected factors were identified as contributing to the decline of the Valley Forge Military Academy (VFMA). Respondents consistently emphasized that declining enrollment served as a central driver of institutional instability. Stakeholders noted that by 2023 the cadet population had fallen to approximately 150, less than half of the enrollment recorded a decade earlier, and dramatically below the institution's mid-twentieth-century peak of more than one thousand students. This long-term downward trajectory weakened the academy's financial position and reduced its capacity to support core programs.

Participants furthermore highlighted rising operational costs and an increasingly unsustainable financial model. Tuition for boarding students had exceeded $49,000 per year, a level regarded as prohibitive for many families. At the same time, insurance and liability expenses escalated sharply, particularly after changes in Pennsylvania law expanded institutional exposure. Numerous lawsuits alleging hazing, abuse, and sexual assault added substantial legal costs, which in turn prompted the academy to liquidate segments of its campus and other real-estate assets.

Survey and interview data also underscored the impact of legal pressures and reputational erosion. Respondents referenced multiple civil actions filed by former cadets, as well as a federal Title IX investigation that identified failures in reporting procedures and interference in the handling of complaints. These developments intensified liability risks and contributed to higher insurance premiums and insurer reluctance, further straining the academy's budget.

Concerns about governance emerged frequently in stakeholder accounts. Many interviewees criticized the board of trustees for micromanagement, insufficient strategic planning, and weak financial oversight. The frequent turnover of presidents and other senior leaders reportedly produced organizational instability. Several parents and

alumni cited decisions they regarded as ill-timed, such as the construction of a new presidential residence valued at approximately $1.7 million, especially given the deferral of basic maintenance on academic buildings and other facilities.

The deterioration of the physical plant constituted another recurrent theme. Survey respondents described academic buildings in visible disrepair, with peeling paint, leaking roofs, and aging infrastructure. For many families, the condition of the campus was a significant deterrent and a symbol of broader institutional decline.

Stakeholders also noted that efforts to secure alternative revenue streams failed to generate the expected financial relief. Initiatives such as franchising the academy's brand overseas, including the creation of a Qatar campus, did not offset domestic losses; similarly, attempts to establish an on-campus charter school were unsuccessful. Land sales provided only temporary support and did not address the underlying structural issues. Neighboring Eastern University was viewed by some as aggressively seeking to purchase academy land.

Finally, several external environmental factors were cited as compounding the academy's challenges. Changes in Pennsylvania liability laws and rising insurance premiums, amplified by ongoing litigation, created a financial environment that many respondents characterized as increasingly unsustainable. Collectively, these findings indicate that VFMA's decline was not attributable to a single cause but to the convergence of demographic, financial, legal, organizational, and infrastructural pressures documented across stakeholder testimonies.

Outcome & Final Decision

In September 2025, the Board of Trustees decided the academy is no longer viable and announced that VFMA will close permanently after the 2025-26 academic year. VFMA will complete its last class (the 98th Corps of Cadets), while the associated *Military College* portion will continue operating under separate governance.

Conclusion

This multi-faceted decline, rooted in a confluence of micro, meso, and macro-level factors, ultimately led to the academy's closure, despite attempts to mitigate financial distress through measures such as asset sales and diversification efforts. The inability to adapt to evolving societal expectations for accountability and financial transparency, coupled with mounting legal and reputational damages, proved insurmountable, leading to the permanent cessation of its operations as a military academy (West & Yaghi, 2024; Downes, 2017). This case study underscores the significant vulnerabilities of small, private, tuition-dependent institutions to multifaceted pressures, resonating with broader trends observed across higher education (Bills, 2020; Johnson, 2021). The closure of institutions like VFMA highlights the critical need for robust governance, fiscal prudence, and a proactive response to legal and societal shifts to ensure long-term viability, particularly for specialized educational models (Muganga *et al.* 2024; West & Yaghi, 2024).

This study offers a novel and significant contribution by providing the first comprehensive, multilayered analysis of the Valley Forge Military Academy's closure, integrating institutional history, governance dynamics, financial data, and sociocultural change into a unified explanatory framework. While existing scholarship often treats school closures as outcomes of isolated financial or enrollment challenges, this case study demonstrates how intersecting legal, reputational, organizational, and societal pressures can converge uniquely within specialized institutions such as military academies. By situating VFMA's decline within broader patterns affecting private, mission-driven schools, the study not only illuminates previously underexamined vulnerabilities in this niche sector but also advances a more holistic model for understanding institutional fragility. In doing so, it provides an essential foundation for future research on organizational resilience and offers actionable insights for educational leaders navigating similarly complex and evolving landscapes.

Acknowledgments

Special thanks are extended to the respondents who freely gave their time without compensation or expectation of reward. Their willingness to share candid and deeply reflective accounts of the organizational decline of an institution that held significant meaning for them was both generous and commendable. This study is strengthened by their honesty, commitment, and trust.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Use of Generative AI and AI-Assisted Technologies

The author declares that he has not used generative AI and AI-assisted technologies in the writing process before submission.

References

[1] Abdullahi, U., Mohamed, A. M., Senasi, V., & Dhahi, A.-A. K. A. (2023). Assessing the Integration of Organizational Resilience and Sustainability: Insights from a Systematic Literature Review. *E3S Web of Conferences*, *440*, 1011. DOI: https://doi.org/10.1051/e3sconf/202344001011

[2] Amankwah-Amoah, J., Khan, Z., Adomako, S., & Khan, H. (2023). Business failure in post-pandemic era: New challenges for industrial networks, emerging insights and market opportunities. *Industrial Marketing Management*, *117*. DOI: https://doi.org/10.1016/j.indmarman.2023.12.014

[3] Bills, M. (2020). *Turning Around Small, Private, Tuition Dependent Colleges: How Boards of Trustees Impact Decline and Turnaround*. https://aura.antioch.edu/cgi/viewcontent.cgi?article=1584&context=etds

[4] Curry, J. R. (1972). *Mount Hermon from 1881 to 1971 : an historical analysis of a distinctive American boarding school.* https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=3792&context=dissertations_1

[5] Downes, M. (2017). University scandal, reputation and governance. *International Journal for Educational Integrity*, *13*(1). DOI: https://doi.org/10.1007/s40979-017-0019-0

[6] Fernandez, F., & Burnett, C. A. (2020). Considering the need for organizational resilience at hispanic serving institutions: a study of how administrators navigate institutional accreditation in Southern states. *International Journal of Qualitative Studies in Education*, *33*(8), 855. DOI: https://doi.org/10.1080/09518398.2020.1751895

[7] Frisbie, K., & Converso, J. A. (2016). Organizational resilience and enrollment trends of independent, for-profit higher education institutions. *Work*, *54*(2), 295. DOI: https://doi.org/10.3233/wor-162296

[8] Fulop, N., Barbosa, E. C., Hill, M., Ledger, J., Ng, P. L., Sherlaw-Johnson, C., Rolewicz, L., Schlepper, L., Spencer, J., Tomini, S. M., Vindrola-Padros, C., & Morris, S. (2023). Rapid evaluation of the Special Measures for Quality and challenged provider regimes: a mixed-methods study [Review of *Rapid evaluation of the Special Measures for Quality and challenged provider regimes: a mixed-methods study*]. *Health and Social Care Delivery Research*, 1. NIHR Journals Library. DOI: https://doi.org/10.3310/gqqv3512

[9] Hager, M. A., Galaskiewicz, J., Bielefeld, W., & Pins, J. J. (1996). Tales From the Grave. *American Behavioral Scientist*, *39*(8), 975. DOI: https://doi.org/10.1177/0002764296039008004

[10] Heredia, J., Rubiños, C., Salas, W. G. V., Heredia, W., & Flores, A. (2022). New Strategies to Explain Organizational Resilience on the Firms: A Cross-Countries Configurations Approach. *Sustainability*, *14*(3), 1612. DOI: https://doi.org/10.3390/su14031612

[11] Higashi, S. Y., Caleman, S. M. de Q., Aguiar, L. K. de, & Manning, L. (2020). What causes organizations to fail? A review of literature to inform future food sector (management) research [Review of *What causes organizations to fail? A review of literature to inform future food sector (management) research*]. *Trends in Food Science & Technology*, *101*, 223. Elsevier BV. DOI: https://doi.org/10.1016/j.tifs.2020.05.011

[12] Johnson, R. (2021). *A Case Study of One Small Christian College Overcoming Decline and Implementing an Institutional Turnaround*. https://digitalcommons.acu.edu/cgi/viewcontent.cgi?article=1362&context=etd

[13] Kelchen, R., Ritter, D., & Webber, D. (2024). Predicting College Closures and Financial Distress. In *Working paper*. Federal Reserve Bank of Philadelphia. DOI: https://doi.org/10.21799/frbp.wp.2024.20

[14] Moran, K. A. (2016). Organizational resilience: Sustained institutional effectiveness among smaller, private, non-profit US higher education institutions experiencing organizational decline. *Work*, *54*(2), 267. DOI:https://doi.org/10.3233/wor-162299

[15] Muganga, A., Oladipo, O. A., & Adarkwah, M. A. (2024). Uganda's post-colonial privatisation policy in higher education: an integrative literature review and case study analysis. *Discover Education*, *3*(1). DOI:https://doi.org/10.1007/s44217-024-00321-5

[16] Ouédraogo, A., & Boyer, M. (2012). Firm Governance and Organizational Resiliency in a Crisis Context: A Case Study of a Small Research-based Venture Enterprise. *International Business Research*, *5*(12). DOI:https://doi.org/10.5539/ibr.v5n12p202

[17] Paul, D. A. (2005). Higher Education in Competitive Markets: Literature on Organizational Decline and Turnaround. *The Journal of General Education*, *54*(2), 106. DOI: https://doi.org/10.1353/jge.2005.0023

[18] Samdanis, M., & Lee, S. H. (2018). Uncertainty, strategic sensemaking and organisational failure in the art market: What went wrong with LVMH's investment in Phillips auctioneers? *Journal of Business Research*, *98*, 475. DOI: https://doi.org/10.1016/j.jbusres.2018.08.030

[19] Siwale, J., Kimmitt, J., & Amankwah-Amoah, J. (2021). The Failure of Hybrid Organizations: A Legitimation Perspective. *Management and Organization Review*, *17*(3), 452. DOI: https://doi.org/10.1017/mor.2020.70

[20] West, A., & Yaghi, B. B. (2024). Academies in England and Charter Schools in the US: Who Is Accountable, to Whom, for What, and with What Consequences? *Journal of School Choice*, 1. DOI:https://doi.org/10.1080/15582159.2024.2375695

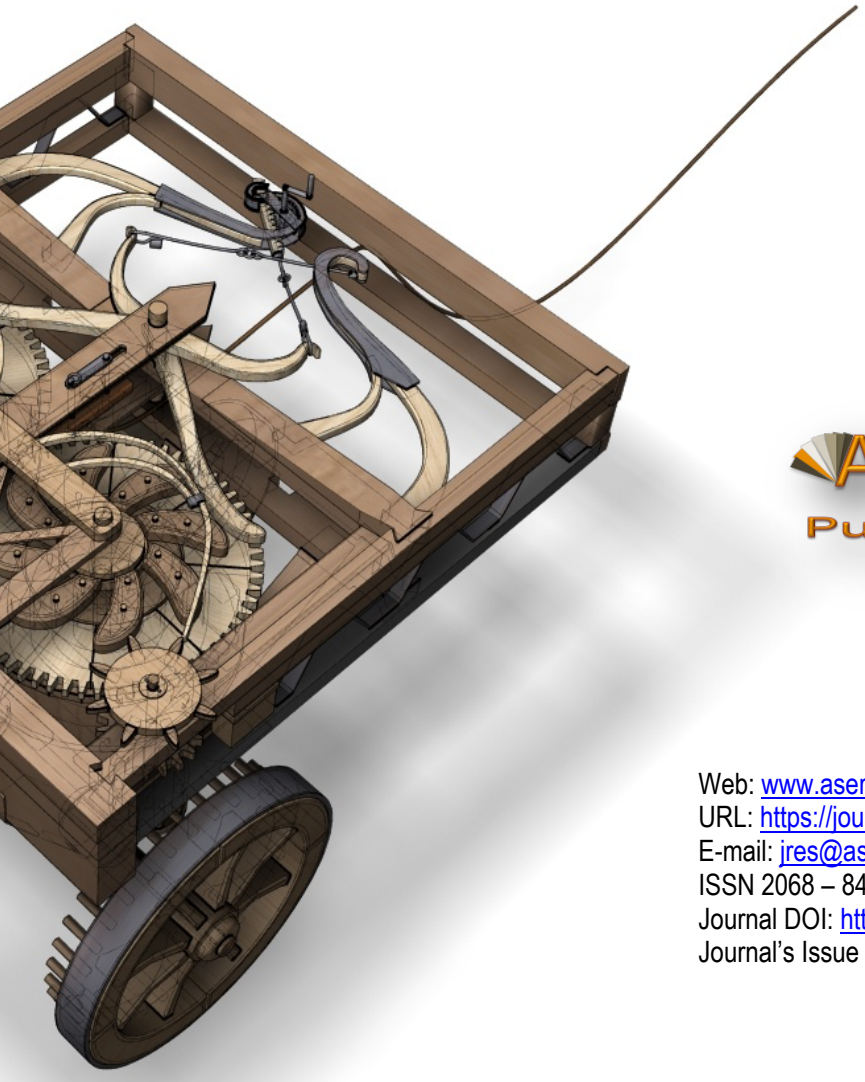
[21] Williams, A., Whiteman, G., & Kennedy, S. (2019). Cross-Scale Systemic Resilience: Implications for Organization Studies. *Business & Society*, *60*(1), 95. DOI: https://doi.org/10.1177/0007650319825870

[22] Wood, M. D., Wells, E., Rice, G., & Linkov, I. (2018). Quantifying and mapping resilience within large organizations. *Omega*, *87*, 117. DOI: https://doi.org/10.1016/j.omega.2018.08.012

[23] You, J. J. (2023). An Overview of Organizational Resilience in Research and Strategy: Implications for the Future of Work. *AIB Insights*, *23*(3). DOI: https://doi.org/10.46697/001c.77387

[24] Zaman, R. (2023). When corporate culture matters: The case of stakeholder violations. *The British Accounting Review*, *56*(1), 101188. DOI: https://doi.org/10.1016/j.bar.2023.101188

SERS



ASERS Publishing

Web: www.aserspublishing.eu
URL: https://journals.aserspublishing.eu/jres
E-mail: jres@aserspublishing.eu
ISSN 2068 – 8407
Journal DOI: https://doi.org/10.14505/jres
Journal's Issue DOI: https://doi.org/10.14505/jres.v16.2(20).00