Conference Proceedings

ICEMST

International Conference on Education in Mathematics, Science and Technology

06 - 10 May 2026
Cappadocia, Nevsehir / Türkiye

www.istes.org/event/icemst

Proceedings of International Conference on Education in Mathematics, Science and Technology

Volume: 13

© 2026 Published by ISTES

ISSN: 2833-7360

The proceedings is licensed under a Creative Commons Attribution-NonCommercialShareAlike 4.0 International License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. Authors alone are responsible for the contents of their papers. The Publisher, the ISTES, shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material. All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations regarding the submitted work. The submissions are subject to a double-blind peer review process by at least two reviewers with expertise in the relevant subject area. The review policy is available on the conference web page.

Conference: International Conference on Education in Mathematics, Science and Technology (ICEMST)

Conference Dates: 06 - 10 May 2026

Conference Location: Cappadocia, Nevsehir / Türkiye

Publication Date: December 30, 2026

Editors

Prof. Dr. M. Sami Ozturk, Necmettin Erbakan University, Turkiye
Dr. Mevlut Unal, Necmettin Erbakan University, Turkiye
M. Lutfi Ciddi, Necmettin Erbakan University, Turkiye

CONFERENCE PRESIDENTS

Dr. Mack SHELLEY - Iowa State University, United States
Dr. Wilfried ADMIRAAL - Leiden University, the Netherlands
Dr. I-Tsun CHIANG - National Taiwan Normal University, Taiwan
Dr. Wenxia WU - George Washington University, United States

SCIENTIFIC BOARD

Allan TARP - MATHeCADEMY, Denmark
Altay FIRAT - European University of Lefke, Cyprus
Andrea DEBELJUH - University Juraj Dobrila of Pula, Croatia
Brahim FERDI - Bechar University, Algeria
Branislav POPOVIĆ - University of Kragujevac, Serbia
Chalavadi SULOCHANA - Gulbarga University, India
Dariga NURKESHEVA - Nazarbayev University, Kazakhstan
Elizabeth ADAMSON - Edinburgh Napier University, United Kingdom
Farhad BALASH - Kharazmi University, Iran
Farouq ALMEQDADI - Emirates College for Advanced Education (ECAE), U.A.E.
Gordana SAVIC - University of Belgrade, Serbia
Haydar YUKSEK - Kafkas University, Turkey
Henry David KATNIYON - Federal College of Education, Pankshin, Plateau state, Nigeria
Hsin-Chih WU - National Taiwan Normal University, Taiwan
Janet VALDEZ - La Consolacion University Philippines, Philippines
Jessie BUSTILLOS - London Metropolitan University, United Kingdom
Kartika YULIANTI - Bina Nusantara University, Indonesia
Mesut AYDIN - Inonu University, Turkey
Milica PAVKOV HRVOJEVIĆ - University of Novi Sad, Serbia
Mohammad SARWAR - Scialert, Dubai, United Arab Emirates
Morteza BARIN - Farhangiyan University of Iran, Iran
Muteb ALQAHTANI - Rutgers University, United States
Nurten SARGIN - Necmettin Erbakan University, Turkey
Ognyan B. MANOLOV - European Polytechnic University, Bulgaria
Sanaa AL-DELAIMY - Mosul University, Iraq
Shynar BAIMAGANBETOVA - Nazarbayev University, Kazakhstan
Summer MOUALLEM - University of Central Lancashire, United Kingdom
Tamer USTUNER - Kahramanmaras Sutcu Imam University, Turkey
Tri Marhaeni PUDJI ASTUTI - Semarang State University, Indonesia
Yi-Teng HSU - National Taiwan Normal University, Taiwan

Prof. Dr Milan Zlatanović - University of Niš, Serbia

ORGANIZING COMMITTEE

Aliya MUSTAFINA - Nazarbayev University, Kazakhstan
Ann D. THOMPSON - Iowa State University, United States
Halil SNOPCE - South East European University, Macedonia
Hakan AKCAY - Yildiz Technical University, Turkey
Jacqueline T. MCDONNOUGH - Virginia Commonwealth University, United States
Mariusz JAWORSKI - Medical University of Warsaw, Poland
Mary M. CAPRARO - Texas A&M University, United States
Muhammad ZAYYAD - Al-Qasemi Academic College of Education, Israel
Murat BEYTUR - Kafkas University, Turkey
Mustafa PEHLIVAN - Necmettin Erbakan University, Turkey
Natela DOGHONADZE - International Black Sea University, Georgia
Natalija ACESKA - Ministry of Education and Science, Macedonia
Ossi AUTIO - University of Helsinki, Finland
O. Tayfur OZTURK - Necmettin Erbakan University, Turkey
Silvia MORARU - National High School Bucharest, Romania
Silviya KOSTOVA - The D.A. Tsenov Academy of Economics, Bulgaria

Prof. Dr Milan Zlatanović - University of Niš, Serbia

PARTICIPATING COUNTRIES

Philippines	Kyrgyz Republic	Indonesia
Oman	Türkiye	China
Peru	Ukraine	Albania
Greece	Georgia	Bulgaria
United States of America (USA)	Saudi Arabia	Malaysia
Algeria	United Arab Emirates	Croatia
Morocco	Kazakhstan	Ecuador
Slovakia	Czech Republic	Portugal
Canada	Hungary	Germany

INDEXES

The publications affiliated with ISTES Organization are indexed or listed by all or some of the following sources:

SUPPORTED BY

ISTES EVENTS



International Conference on Studies in Education and Social Sciences

05 - 09 November 2026 - Istanbul, Türkiye



International Conference on Studies in Engineering, Science, and Technology

05 - 09 November 2026 - Istanbul, Türkiye



International Conference on Social, Educational, and Human sciences

04 - 07 December 2026 - Istanbul, Türkiye



International Conference on Engineering, Technology, and Life sciences

04 - 07 December 2026 - Istanbul, Türkiye



International Conference on Behavioral, Educational, and Social Sciences

04 - 07 December 2026 - Miami, USA



International Conference on Business, Engineering, Science, and Technology

04 - 07 December 2026 - Miami, USA



International Conference on Business, Art, Social sciences, and Education

04 - 07 December 2026 - Paris, France



International Conference on Technology, Applied Sciences, and Engineering

04 - 07 December 2026 - Paris, France



International Conference on Humanities, Social and Education Sciences

08 - 11 April 2027 - Los Angeles, USA



International Conference on Life Sciences, Engineering and Technology

08 - 11 April 2027 - Los Angeles, USA



International Conference on Research in Education and Science

28 April - 02 May 2027 - Marmaris, Muğla / Türkiye

ISTES EVENTS

International Conference on Education in Mathematics, Science and Technology
28 April - 02 May 2027 - Marmaris, Muğla / Türkiye

International Conference on Humanities, Education, and Social Sciences
16 - 19 March 2027 - Singapore

International Conference on Technology, Engineering, and Life Sciences
16 - 19 March 2027 - Singapore

International Conference on Social science, Education science, and Humanities
26 - 29 July 2027 - Wahsington, USA

International Conference on Engineering, Life sciences, and Technology
26 - 29 July 2027 - Wahsington, USA

International Conference on Humanities and Social Sciences
26 - 29 July 2027 - Vienna, Austria

International Conference on Life Sciences and Engineering
26 - 29 July 2027 - Vienna, Austria

International Conference on Humanities, Education, and Social sciences
26 - 29 July 2027 - Singapore

International Conference on Health, Engineering, and Life sciences
26 - 29 July 2027 - Singapore

International Conference on Engineering, Science and Technology
21 - 24 October 2027 - Washington D.C., USA

International Conference on Social and Education Sciences
21 - 24 October 2027 - Washington D.C., USA

Contents

When AI Replaces Thinking: Cognitive Struggle and Shortcut Behavior in Mathematics Learning.....1
Using Gapped Notes to Enhance Students' Learning and Interaction in the Calculus Classroom.....16
Prompt Engineering as Metacognitive Scaffolding: A Self-Regulated Learning Theory Framework for AI-Mediated Education.....30
An Economic Mathematical Curricular Complexity Analysis of Data Science Bachelor's Degree Programs at U.S. Public Universities.....49
Critical Thinking in the Age of Generative AI: A Neo-Piagetian Cognitive Development Framework.....63
A Mathematical Curricular Complexity Analysis of U.S. Primary and Lower Secondary School Mathematics Curricula.....85
Multimedia Presentation Modalities and Mathematical Problem Solving: An Experimental Study Within Mayer's Cognitive Theory of Multimedia Learning.....100
AI-Enhanced Pedagogical Transformation in Teacher Education: A Systematic Framework.....114
Application of Machine Learning Techniques to Support Professional Track Selection in Medical Education.129
A Culturally Responsive Android Learning Media to Promote Differentiated Deep Learning and Cultural Literacy.....139
Artificial Intelligence Integration into AS9100-Based Quality Management Systems: A Literature Review in the Context of Defense Technologies.....148
Cultural Connotation and Contemporary Educational Value of Microcosmization in Traditional Chinese Medicine.....161
Design of a Web-Based Application "Qiraapp" Using Arabic News Texts as a Learning Media for Maharah Qirā'ah at Universitas Negeri Malang.....170
Designing a Qur'an-Based Qawāʿid Learning Web: Integrating Digital Mushaf for Arabic Grammar Instruction 181
Assessing Vertical Alignment and Cognitive Progression of BSIS Course Learning Outcomes Using Bloom's Taxonomy in Outcome-Based Education Framework.....192
Mathematization challenges in infinite processes: findings from first-year university students.....204

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

When AI Replaces Thinking: Cognitive Struggle and Shortcut Behavior in Mathematics Learning

Fathimah Azzahra Nasiruddin [1], Toto Nusantara [2 *], Erry Hidayanto[3], Indriati Nurul Hidayah [4]

[1] Department of Mathematics, Universitas Negeri Malang, Malang, Indonesia 65145, Indonesia
[2] Department of Mathematics, Universitas Negeri Malang, Malang, Indonesia 65145, Indonesia
[3] Department of Mathematics, Universitas Negeri Malang, Malang, Indonesia 65145, Indonesia
[4] Department of Mathematics, Universitas Negeri Malang, Malang, Indonesia 65145, Indonesia
* Corresponding author: Toto Nusantara (icemstoffice@gmail.com)

Article Info

Article History
Received:
12 January 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Artificial intelligence in education
Cognitive offloading
Productive struggle
Mathematics learning
Reflective practice

Abstract

The rapid integration of generative artificial intelligence (AI) into learning environments has reshaped how students engage with mathematical problem solving. While AI offers potential cognitive support, growing concerns remain regarding its tendency to replace rather than support students' independent thinking. This study investigates how students respond to mathematical difficulties in AI-rich learning environments and examines whether AI use functions as cognitive support or as a cognitive shortcut. This study employed a qualitative research design supported by descriptive quantitative data. Data were collected through a self-administered questionnaire consisting of Likert-scale items capturing students' emotional responses to mathematical difficulty, tendencies in AI use as a problem-solving strategy, and reflection behavior after incorrect answers, as well as open-ended reflective responses. Quantitative data were analyzed descriptively, while qualitative data were examined using thematic analysis to capture students' cognitive processes during moments of impasse. The findings indicate that many students experience emotional discomfort when encountering mathematical difficulty and frequently rely on AI tools to obtain answers quickly. AI use was more often associated with cognitive offloading than with scaffolded reasoning. In addition, reflective practices following incorrect answers were limited, suggesting that errors were rarely used as opportunities for learning. Overall, the study highlights the need for pedagogical regulation to ensure that AI supports productive struggle and deep mathematical thinking rather than replacing essential reasoning processes.

Citation: Nasiruddin, F. A., Nusantara, T., Hidayanto, E., & Hidayah, I. N. (2026). When AI Replaces Thinking: Cognitive Struggle and Shortcut Behavior in Mathematics Learning. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 1–15), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

The rapid integration of artificial intelligence (AI) tools into higher education has significantly transformed how students engage with academic tasks, particularly in mathematics learning. AI-powered technologies now provide instant solutions, step-by-step explanations, and automated feedback, positioning themselves as readily available forms of cognitive assistance. In this study, the term AI tools refers primarily to generative and solution-based digital tools commonly accessed by students for academic purposes, such as text-based generative AI, online problem-solving platforms, and AI-supported search engines that provide immediate answers or procedural guidance.

These tools were not formally embedded within instructional design but were independently accessed by students during learning activities. While prior research highlights the potential of AI to enhance efficiency, accessibility, and individualized support, growing concerns suggest that such technologies may also reshape students' cognitive engagement during learning, particularly when they encounter difficulty or uncertainty. Mathematical problem solving inherently involves cognitive struggle, a process widely recognized as essential for conceptual understanding and durable learning. Productive struggle encourages learners to explore alternative strategies, tolerate uncertainty, reflect on errors, and gradually refine their reasoning.

However, in learning environments where AI tools provide immediate solutions, students may begin to reinterpret struggle not as a meaningful phase of learning but as an obstacle to be eliminated. This shift raises critical questions about whether AI functions primarily as a cognitive scaffold that supports thinking or as a shortcut that replaces essential reasoning processes. Research on cognitive offloading suggests that learners tend to delegate mentally demanding processes to external tools when such tools are readily available (Risko & Gilbert, 2016).

In mathematics education, cognitive offloading may manifest as reliance on AI to bypass problem analysis, strategy formulation, and evaluation of errors. While offloading can reduce cognitive load in routine or peripheral tasks, excessive reliance risks weakening metacognitive regulation and diminishing opportunities for deep conceptual learning. Despite increasing attention to AI-assisted learning, empirical evidence examining how students respond to mathematical difficulty in AI-rich contexts remains limited. Existing studies have largely focused on students' attitudes toward AI or learning outcomes, leaving a gap in understanding the process-level changes in students' thinking behaviors during moments of impasse. In particular, little is known about how emotional responses such as confusion, frustration, or uncertainty interact with students' decisions to rely on AI and how this interaction influences reflective learning practices after errors. Understanding this process is especially important for pre-service teachers, whose learning habits and beliefs about struggle, effort, and assistance will later shape their instructional practices.

Addressing this gap, the present study investigates how undergraduate pre-service teachers respond to mathematical difficulties in AI-rich learning environments. Focusing on students' emotional responses to difficulty, their tendency to use AI as a shortcut, and their reflection on incorrect answers, this study examines whether AI use supports cognitive engagement or tends to replace independent thinking during mathematical

problem solving. By integrating descriptive quantitative patterns with qualitative thematic analysis, this research offers empirical insight into how AI reshapes cognitive struggle, learning strategies, and reflective practices in contemporary mathematics education.

Research Questions

1. How do students respond to mathematical difficulties in AI-rich learning environments?
2. Does AI use function as cognitive support or as a replacement for thinking during mathematical problem solving?

Method

Research Design

This study adopted a qualitative research design supported by descriptive quantitative data to explore how undergraduate students respond cognitively and emotionally to mathematical difficulty in environments enriched with artificial intelligence (AI). This approach aligns with educational research practices in which qualitative inquiry provides contextualized, interpretive insight, while descriptive quantitative data offer breadth and pattern recognition without claiming causal inference (Colorafi & Evans, 2016; Razali et al., 2025). Unlike experimental or intervention-based designs that rely on pre- and post-measures, this qualitative-dominant approach prioritizes understanding participants' lived perspectives, especially where rich reflection and meaning-making are central to research aims. Qualitative data illuminate how students think, struggle, and choose to rely on AI, whereas descriptive quantitative measures provide supporting evidence of broader tendencies across the study group.

Participants were 76 undergraduate pre-service elementary teachers enrolled in a mathematics education course at a public university in Indonesia. All participants reported prior experience using AI tools for academic tasks, particularly for obtaining explanations or answers during mathematical problem solving. Participation was voluntary and responses were anonymized prior to analysis to protect confidentiality and comply with ethical standards.

Instruments

Data were collected using a self-administered questionnaire consisting of two main components: Likert-scale items and open-ended reflective questions.

Likert-Scale Items

Three Likert-scale items were designed to capture descriptive indicators of students' responses when encountering mathematical difficulty in AI-rich learning environments. The items focused on: (a) emotional responses to mathematical difficulty, (b) the tendency to use AI or external sources as a problem-solving strategy, and (c) reflection behavior following incorrect answers.

The Likert items were intended as descriptive indicators rather than as a psychometric measurement scale. Consequently, the items were not designed to measure latent constructs or to support inferential statistical analysis. Instead, they served to illustrate general tendencies in students' emotional regulation, AI reliance, and reflective practices, which were subsequently interpreted in relation to qualitative thematic findings.

Open-Ended Reflection Questions

The primary qualitative data were obtained through open-ended reflection questions that invited participants to describe their thoughts and actions when they felt stuck, confused, or uncertain during mathematical problem solving. These narrative responses were designed to elicit students' lived cognitive experiences, including their strategies, emotional regulation, and decisions to seek assistance through AI or other external sources.

The open-ended responses constituted the central data source for understanding how students navigated cognitive struggle in AI-rich contexts, while the Likert-scale items provided complementary descriptive patterns to support qualitative interpretation. Descriptive quantitative data were used to support interpretation of qualitative findings, consistent with educational studies where survey-based metrics complement interpretive inquiry (Razali et al., 2025).

Data Collection

Procedure

Data collection occurred during a scheduled class session following a mathematical problem-solving activity. Participants responded to the online questionnaire including both Likert items and open-ended reflections in a single session. They were assured that their responses would be used only for research purposes, and that no identifying information would be collected.

Data Analysis

Quantitative Analysis

Quantitative data from the Likert-scale items were analyzed using descriptive statistics, including frequencies and percentages. This analysis aimed to identify prevalent patterns related to emotional discomfort, reliance on AI as a shortcut, and reflection behaviors after errors. Given the exploratory and qualitative-dominant nature of the study, no inferential statistical analyses were conducted, and no causal claims were made based on the quantitative data.

Qualitative Analysis

Qualitative data from open-ended reflections were analyzed using thematic analysis following an iterative coding process (Braun & Clarke, 2006). Initial codes were generated inductively from participants' written responses,

focusing on actions taken during moments of difficulty, emotional regulation, and the use of AI or other external resources. Through multiple rounds of coding and constant comparison, related codes were grouped into broader themes that captured recurring patterns of cognitive response to mathematical struggle. The analysis prioritized meaning and process over frequency, aiming to represent the range of students' cognitive strategies and decision-making behaviors rather than quantifying theme occurrence.

Integration of Data

Descriptive quantitative patterns were used to inform and contextualize qualitative interpretation but did not independently drive analytic conclusions. Instead, qualitative themes provided explanatory insight into the cognitive processes underlying observed quantitative tendencies. This integration aligns with qualitative-dominant research approaches in educational inquiry, where survey-based descriptors support, rather than replace, interpretive analysis.

Rigor and Trustworthiness

To strengthen credibility, multiple strategies were employed: systematic coding procedures, iterative review of narrative data, and cross-referencing between quantitative patterns and qualitative themes. While this study does not adopt an experimental design with pre- and post-measures, its methodological coherence supports valid interpretation of lived experiences and descriptive patterns.

Results

This section presents the integrated findings from descriptive quantitative data and qualitative thematic analysis. The results illustrate how students respond cognitively and behaviorally when facing mathematical difficulty in AI-rich learning environments.

Table 1. Distribution of Students' Responses to Items Q7–Q9 (n = 76)

Likert Scale	Q7: Lacking specific strategies when stuck	%	Q8: Tendency to take answers from AI/peers	%	Q9: Not examining errors after mistakes	%
Strongly Agree	13	17.11	7	9.21	7	9.21
Agree	38	50.00	34	44.74	24	31.58
Somewhat Agree	17	22.37	25	32.89	16	21.05
Somewhat Disagree	4	5.26	8	10.53	13	17.11
Disagree	4	5.26	2	2.63	15	19.74
Strongly Disagree	–	–	–	–	1	1.32
Total	76	100	76	100	76	100

Q7: “I often say ‘at least I tried,’ but I do not know specific strategies when I get stuck.”
Q8: “If AI tools or peers can provide an answer, I tend to use them to finish the task quickly.” Q9: “After making mistakes, I rarely examine why I was wrong and tend to move on immediately.”

Descriptive Quantitative Results

Descriptive quantitative analysis was conducted to portray general patterns of students’ emotional responses, AI usage tendencies, and reflective behaviors when encountering mathematical difficulties. Three Likert-scale items were used as indicators.

Emotional Responses to Mathematical Difficulty

Students’ emotional responses to mathematical difficulty were examined using Item Q7. The results indicate that approximately 89,48% of students expressed agreement (strongly agree, agree, or somewhat agree) with the statement that they lacked specific strategies when becoming stuck, despite expressing effort or perseverance. This pattern suggests that while many students demonstrate motivational persistence, such persistence is often not accompanied by clearly articulated problem-solving strategies when cognitive impasses occur.

AI Usage as a Problem-Solving Strategy

The tendency to rely on AI tools or peers as a means of completing tasks was captured through Item Q8. The findings show that approximately 53,95% of students agreed or strongly agreed that they tend to use AI or external sources to obtain answers quickly. This pattern indicates a prevalent inclination toward efficiency-oriented behavior, where external assistance is prioritized to complete tasks rather than engaging in extended independent problem-solving.

Reflection Behavior After Incorrect Answers

Students’ reflection practices following incorrect answers were examined using Item Q9. The results reveal that approximately 61.84% of students acknowledged that they rarely examined the reasons for their mistakes and tended to move on immediately. This suggests that reflective analysis of errors has not yet become a habitual component of students’ mathematical learning processes, particularly in learning contexts where external support enables rapid task continuation.

While the descriptive quantitative results reveal general patterns of students’ emotional responses, reliance on external assistance, and limited reflection, they do not fully explain how students cognitively navigate moments of mathematical difficulty in AI-rich learning environments. To gain deeper insight into the processes underlying these patterns, qualitative analysis was conducted on students’ open-ended reflections. This analysis aimed to capture students’ lived experiences, decision-making processes, and immediate responses when they felt stuck or uncertain, thereby providing a richer understanding of how emotional discomfort, AI use, and reflection interact

during mathematical problem solving.

Qualitative Thematic Findings

To gain deeper insight into students' cognitive processes, a thematic analysis was conducted on open-ended responses to the question: "Saat Anda buntu, apa langkah pertama yang Anda lakukan?" Four main themes emerged from the analysis.

Theme 1: Conceptual Re-engagement

A subset of students demonstrated adaptive cognitive responses by re-engaging with fundamental concepts when encountering mathematical difficulty. Rather than immediately seeking external assistance, these students reported returning to the meaning of the problem, identifying known and unknown information, and questioning the underlying concepts being assessed. Several students explicitly described efforts to reconstruct conceptual understanding before applying procedures: "Langkah pertama yang saya lakukan adalah kembali ke makna konteks soal, bukan langsung menghitung." "Saya membaca ulang soal dengan perlahan, lalu mengidentifikasi apa yang diketahui dan apa yang ditanyakan." "Saya menanyakan pada diri sendiri: apa arti angka-angka ini dan apa yang sebenarnya diminta?"

Others emphasized the use of concrete representations and unit reasoning to support sense-making: "Mengubah ke nilai per satuan terlebih dahulu agar lebih mudah dipahami." "Membayangkan situasi nyata seperti kue, pizza, atau jeruk." "Menggunakan gambar, tabel, atau model sederhana sebelum menghitung." These responses reflect forms of productive cognitive struggle, in which students actively attempted to construct meaning and regulate their thinking processes. However, compared to shortcut-oriented behaviors, such concept-based strategies appeared less frequently and were articulated by a smaller group of students.

Theme 2: Visualization and Problem Chunking

Some students reported using visualization and problem chunking as initial strategies when they encountered mathematical difficulty. These responses involved breaking complex problems into smaller, more manageable parts and transforming abstract information into visual or concrete representations. Several students described explicitly dividing problems to identify a point of entry: "Saya memecah soal menjadi bagian kecil agar menemukan titik mulai." "Langkah pertama saya adalah membaca ulang soal dan memecahnya menjadi bagian-bagian kecil."

Others emphasized the use of visual and representational tools to support understanding: "Saya menggambarkan situasi soal secara visual agar lebih mudah dipahami." "Menggunakan tabel atau gambar untuk melihat hubungan antar nilai." "Mengaitkan soal dengan situasi sehari-hari agar lebih mudah membayangkan." These strategies reflect attempts to manage cognitive load and support sense-making through representational thinking. However, similar to conceptual re-engagement, visualization and chunking strategies were not consistently sustained throughout the problem-solving process and often preceded subsequent reliance on external assistance, such as

AI tools or peer support.

Theme 3: Emotional Regulation before Action

Another group of students responded to mathematical difficulty by regulating their emotional state before taking further action. These responses involved pausing, remaining silent, calming themselves, or taking a brief break to regain mental readiness. Several students explicitly described this emotional pause: “Tenangkan diri dulu biar otak fresh.” “Diam sebentar memikirkan langkah selanjutnya.” “Menenangkan diri dengan berdiam diri selama beberapa menit.” These responses indicate a level of emotional awareness when encountering cognitive impasses. However, emotional regulation was often not accompanied by explicit cognitive strategies, such as conceptual analysis or representational planning. In many cases, this pause functioned as a transitional phase that preceded external help-seeking, including the use of AI tools or internet resources. Thus, while emotional regulation may serve as an important initial response to difficulty, it did not consistently lead to deeper cognitive engagement or sustained problem-solving efforts.

Theme 4: Cognitive Offloading through AI and External Sources

Cognitive offloading through AI and other external sources emerged as the most dominant response when students encountered mathematical difficulty. A substantial number of students reported that their first action upon feeling stuck was to seek immediate assistance from AI tools, internet resources, or peers, often without describing intermediate cognitive exploration or strategy development. Many students explicitly stated that they relied on AI as their primary means of obtaining answers or solution steps: “Saat pertama saya buntu, saya pergi ke AI.” “Mikir sebentar lalu buka AI.” “Menggunakan ChatGPT dan Gemini AI.” “Mencari jawaban di AI.” “Membuka AI.” Similar patterns were evident in responses referring to internet searches and peer assistance: “Cari di internet.” “Mencari solusi di internet dan media sosial.” “Bertanya ke internet dan ke teman.” “Meminta bantuan AI atau bantuan teman.” These responses suggest that external resources were frequently positioned as the first line of response to cognitive impasse. In many cases, students did not describe attempts to analyze the problem, revisit underlying concepts, or experiment with alternative representations before seeking external solutions. Instead, cognitive effort was rapidly delegated to tools or individuals perceived as capable of providing quick and reliable answers. In some responses, minimal internal regulation preceded offloading behavior, indicating a brief pause before external help-seeking: “Tenangkan diri dulu biar otak fresh, lalu cari di AI.” “Diam sebentar, kemudian mencari jawaban di internet.” Such patterns indicate that emotional regulation often functioned as a transitional phase rather than a catalyst for deeper cognitive engagement. The availability of AI and online resources appeared to shorten the duration of productive struggle, shifting students’ focus from sense-making and strategy exploration toward efficient task completion. Overall, these narratives illustrate a consistent pattern of cognitive offloading, in which AI and external sources replaced extended independent reasoning. In this context, AI functioned primarily as a shortcut to completing mathematical tasks rather than as a scaffold supporting conceptual understanding, reflection, or sustained problem-solving effort.

Integrated Interpretation of Findings

When examined together, the descriptive quantitative and qualitative findings reveal a coherent pattern in how students respond to mathematical difficulty in AI-rich learning environments. Quantitative results indicate that many students experience difficulty without clear strategic regulation, frequently rely on external assistance to complete tasks efficiently, and engage in limited reflection after making errors. The qualitative findings provide process-level insight into how these tendencies unfold during moments of cognitive impasse. Qualitative themes show that although some students attempted concept-based reasoning (Theme 1) or employed visualization and problem chunking strategies (Theme 2), such cognitively demanding approaches were articulated by a smaller subset of participants. These strategies often appeared as initial efforts but were not consistently sustained throughout the problem-solving process. Similarly, emotional regulation responses (Theme 3), such as pausing or calming oneself, reflected students' awareness of cognitive and emotional discomfort; however, these pauses frequently served as transitional moments rather than catalysts for deeper strategic engagement. In contrast, cognitive offloading through AI and other external sources (Theme 4) emerged as the most dominant response to difficulty. Students commonly described turning to AI tools, internet resources, or peers as their first or immediate action when they felt stuck. This qualitative pattern aligns closely with the quantitative finding that a large proportion of students reported using AI or external help to finish tasks quickly. Together, these findings suggest that emotional discomfort often precedes rapid external help-seeking, effectively shortening the duration of productive struggle. Limited reflection after incorrect answers, as indicated by the quantitative results, further reinforces this pattern. When answers can be quickly obtained or verified through AI, opportunities to analyze errors, reconsider strategies, or reconstruct understanding are reduced. As a result, learning interactions tend to prioritize task completion over sense-making and reflective engagement. Overall, the integrated findings suggest a shift from productive struggle toward shortcut-oriented learning behavior in AI-rich contexts. While adaptive cognitive strategies remain present among some students, the widespread availability of AI appears to redistribute cognitive effort away from sustained independent reasoning toward external tools. Without pedagogical structures that guide AI use and protect opportunities for reflection, AI is more likely to function as a substitute for thinking rather than as a scaffold that supports deep mathematical understanding.

Discussion

This study examined how students respond to mathematical difficulty in AI-rich learning environments and whether AI use functions as cognitive support or as a replacement for thinking. By integrating descriptive quantitative patterns with qualitative thematic analysis, the findings offer process-level insight into how cognitive struggle, emotional regulation, and AI use interact during mathematical problem solving.

Productive Struggle in AI-rich Learning Environments

The findings indicate that while students frequently encounter cognitive struggle, such struggle is often not sustained long enough to become productive. Quantitative results show that many students express perseverance without clear strategic regulation, while qualitative findings reveal that only a subset of students engaged in concept-based reasoning or visualization strategies. These results align with research on productive struggle, which emphasizes that learning benefits arise not merely from effort, but from engaging in strategic sense-making,

representation, and reflection during periods of difficulty (Kapur & Bielaczyc, 2012; Kapur, 2016). In AI-rich contexts, however, struggle appears increasingly reframed as an undesirable state rather than a meaningful learning phase. The availability of immediate solutions through AI tools may reduce students' tolerance for uncertainty, shortening the time spent exploring alternative strategies. Recent studies on technology-supported learning similarly caution that excessive instructional efficiency can undermine deep conceptual engagement when learners bypass exploration and explanation (van der Kleij et al., 2015; Zhai, 2022). Thus, the presence of AI does not eliminate struggle, but rather reshapes how students respond to it.

Cognitive Offloading and the Redistribution of Thinking

A central contribution of this study is the empirical demonstration of cognitive offloading as a dominant response to mathematical difficulty. Both quantitative and qualitative findings consistently show that students frequently rely on AI tools, internet resources, or peers as their first line of action when they feel stuck. This behavior aligns with cognitive offloading theory, which posits that individuals tend to delegate cognitively demanding processes to external tools when such tools are easily accessible (Risko & Gilbert, 2016). While cognitive offloading can be adaptive for routine or peripheral tasks, recent research suggests that excessive offloading during learning activities may weaken metacognitive regulation and reduce opportunities for knowledge construction (Fenech et al., 2021; Fisher et al., 2022). In mathematics learning, where understanding develops through sense-making, error analysis, and strategy refinement, the delegation of reasoning to AI risks transforming learning into a process of answer acquisition rather than conceptual development. Importantly, the findings indicate that cognitive offloading is often preceded by emotional discomfort. Students commonly described confusion, hesitation, or the need to calm themselves before turning to AI. This pattern resonates with research showing that learners are more likely to offload cognition when experiencing uncertainty or cognitive strain (Pacheco et al., 2023). In such moments, AI becomes not only a cognitive shortcut but also an emotional coping mechanism.

Reflection, Error Analysis, and Shallow Learning Cycles

Another critical finding concerns students' limited reflection after incorrect answers. Quantitative results indicate that many students rarely examine why they were wrong, and qualitative responses provide little evidence of systematic error analysis. From a learning sciences perspective, reflection and feedback processing are central mechanisms through which errors become learning opportunities (Hattie & Timperley, 2007; Winne & Azevedo, 2014). Recent research on AI-assisted learning environments suggests that when answers can be rapidly generated or verified by intelligent systems, learners may perceive reflection as unnecessary, leading to shallow learning cycles characterized by rapid correction without conceptual change (Kasneci et al., 2023,, 2023; Selwyn, 2022). The findings of this study support these concerns, showing that AI-mediated efficiency may inadvertently reduce students' engagement in reflective sense-making unless reflection is explicitly required and scaffolded.

Implications for AI Pedagogy in Mathematics Education

Taken together, the findings underscore the importance of pedagogical regulation in AI-rich learning

environments. AI is not inherently detrimental to learning; rather, its educational value depends on how it is positioned within instructional practice. When AI is used primarily as an answer generator, it is likely to replace thinking. When framed as a cognitive scaffold, however, AI can support explanation, justification, and metacognitive awareness (Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G, 2023; Stahl & Hakkarainen, 2021). Recent frameworks for AI pedagogy emphasize the need to preserve opportunities for productive struggle, delay automated assistance, and design tasks that require students to critique, explain, or compare AI-generated solutions (Pacheco et al., 2023; Zhai, 2022). In mathematics education, this includes emphasizing multiple representations, valuing reasoning processes over final answers, and embedding structured reflection after errors. Developing students' AI literacy—particularly their awareness of when AI use supports learning and when it undermines cognitive engagement—is therefore essential for sustaining deep mathematical thinking in AI-rich classrooms.

Theoretical Contribution

This study contributes to the growing body of research on AI in education by empirically demonstrating how AI-rich environments can transform cognitive struggle into shortcut behavior if not pedagogically regulated. By integrating emotional responses, AI usage patterns, and reflective behaviors, the findings provide a nuanced account of how thinking is redistributed between learners and intelligent tools. In this sense, the study extends theories of productive struggle and cognitive offloading into contemporary AI-mediated mathematics learning contexts, offering evidence that when AI tends to replace opportunities for independent thinking, learning risks becoming procedurally efficient but cognitively shallow.

Conclusion and Educational Implications

Conclusion

This study examined how students respond to mathematical difficulties in AI-rich learning environments and whether AI use functions as cognitive support or as a replacement for thinking. By integrating descriptive quantitative data with qualitative thematic analysis, the findings provide a nuanced understanding of how cognitive struggle, emotional regulation, and AI use intersect during mathematical problem solving. The results indicate that many students experience difficulty without clear strategic regulation and tend to rely on external assistance—particularly AI tools to complete tasks efficiently. While some students demonstrated adaptive strategies such as conceptual re-engagement and visualization, these approaches were articulated by a smaller subset of participants and were often not sustained.

Emotional regulation frequently preceded help-seeking behavior but did not consistently lead to deeper cognitive engagement. Instead, cognitive offloading through AI and other external sources emerged as the most dominant response to difficulty, effectively shortening the duration of productive struggle. Limited reflection after incorrect answers further reinforced surface-level engagement with mathematical tasks. Taken together, the findings suggest a shift from productive struggle toward shortcut-oriented learning behavior in AI-rich contexts. Rather than supporting students' independent reasoning processes, AI was more often positioned as a substitute for

thinking. This shift does not imply a lack of effort on the part of students, but rather reflects how readily available AI tools reshape learners' responses to uncertainty, difficulty, and error.

Educational Implications

The findings of this study carry important implications for mathematics education in the era of generative AI. First, AI should be intentionally positioned as a cognitive scaffold rather than an answer generator. Without pedagogical regulation, students are likely to default to efficiency-oriented use that replaces sense-making and strategic reasoning. Educators should therefore design learning activities that require students to explain, justify, or critique AI-generated solutions, ensuring that AI use supports rather than bypasses cognitive engagement.

Second, instructional practices must deliberately protect opportunities for productive struggle. This includes delaying access to AI during initial problem exploration, presenting non-routine or conceptually rich tasks, and emphasizing multiple representations before procedural computation. Struggle should be framed as a meaningful and expected phase of learning, rather than as an obstacle to be eliminated through technological shortcuts.

Third, reflection and error analysis need to be explicitly taught, structured, and assessed. Given the tendency for students to move on quickly after receiving answers from AI, learning designs should require students to analyze incorrect solutions, compare alternative strategies, and articulate why certain approaches fail. Embedding reflective prompts and metacognitive questioning can help counteract shallow learning cycles in AI-supported environments.

Fourth, the findings highlight the growing importance of AI literacy in mathematics education. Beyond technical proficiency, students need guidance on when AI use supports learning and when it undermines cognitive engagement. Developing critical awareness of AI's limitations, encouraging verification and sense-making, and fostering responsible decision-making about AI use are essential for sustaining deep learning.

Finally, for pre-service teachers, these implications are particularly salient. The ways in which future teachers experience struggle, assistance, and reflection during their own learning are likely to shape their instructional practices. Teacher education programs should therefore model pedagogical approaches that integrate AI thoughtfully, preserve productive struggle, and cultivate reflective learning habits.

Future Directions

Future research should investigate instructional designs that integrate AI with structured reflection and delayed assistance to support productive struggle in mathematics learning. Longitudinal and experimental studies are needed to examine how sustained AI use influences students' reasoning, metacognitive development, and learner autonomy over time. Additionally, comparative studies across educational contexts and learner populations would further illuminate how cultural and institutional factors shape students' engagement with AI in learning environments.

Limitations

Several limitations of this study should be acknowledged to contextualize the findings and guide future research.

First, this study relied on self-reported questionnaire and reflection data, which may be subject to social desirability bias or variations in students' self-awareness. Students' reported AI usage and reflective behaviors may not fully capture their actual problem-solving processes during real-time mathematical activity.

Second, the quantitative component of the study was descriptive in nature and did not involve pre-test/post-test measures or inferential statistical analysis. Consequently, the findings do not claim causal relationships between AI use and cognitive outcomes. Instead, the quantitative data served to illustrate patterns that were further interpreted through qualitative thematic analysis.

Third, the study focused on a limited set of questionnaire items related to emotional responses, AI usage, and reflection behaviors. While this focused approach strengthened alignment with the research questions, it may not encompass the full complexity of students' beliefs, motivations, and long-term learning dispositions in AI-rich environments.

Fourth, the qualitative data were derived from short written reflections, which varied in depth and clarity across participants. Although thematic saturation was achieved for major patterns, richer insights could be obtained through interviews, think-aloud protocols, or classroom observations.

Finally, the study was conducted within a specific educational and cultural context, which may influence students' attitudes toward effort, help-seeking, and technology use. As such, the findings should be interpreted as context-sensitive rather than universally generalizable.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Ethical approval was obtained from the institutional review board. Participation was voluntary; informed consent was secured prior to data collection. No identifiable data were stored or reported. The study's focus on learning behaviors minimized potential risks to participants.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. *Educational Psychologist*, *48*(1), 1–16. https://doi.org/10.1080/00461520.2013.790143

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, 473–480. https://doi.org/10.1016/j.chb.2015.02.029

Boaler, J. (2016). *Mathematical mindsets: Unleashing students' potential through creative math, inspiring messages and innovative teaching*. Jossey-Bass.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, *8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, *40*(3), 291–309. https://doi.org/10.1080/02619768.2017.1315399

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning strategies. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Fenech, E. P., Drews, F. A., & Bakdash, J. Z. (2021). The effects of automation on human cognition: A systematic review. *Human Factors*, *63*(4), 561–584. https://doi.org/10.1177/0018720820947786

Fisher, M., Goddu, M. K., & Keil, F. C. (2015). Searching for explanations: How the Internet inflates estimates of internal knowledge. *Journal of Experimental Psychology: General*, *144*(3), 674–687. https://doi.org/10.1037/xge0000070

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Information Age.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for*

teaching and learning. Center for Curriculum Redesign.

Kapur, M. (2016). Examining productive failure, productive struggle, and scaffolding in learning. *Educational Psychologist*, *51*(2), 232–252. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Kasneci, G (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

- NCTM. (2014). *Principles to actions: Ensuring mathematical success for all*. National Council of Teachers of Mathematics.

OECD (2021). *Artificial intelligence in education: Challenges and opportunities*. OECD Publishing. https://doi.org/10.1787/8d79a4e8-en

Pacheco, D., Lipszyc, B., & Sigman, M. (2023). Cognitive offloading and learning in digital environments. *Computers & Education*, *190*, 104593. https://doi.org/10.1016/j.compedu.2022.104593

Panadero, E. (2017). A review of self-regulated learning. *Educational Psychology Review*, *29*(1), 1–34. https://doi.org/10.1007/s10648-015-9320-7

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Schoenfeld, A. H. (2016). Learning to think mathematically. *Journal of Education*, *196*(2), 1–38. https://doi.org/10.1177/002205741619600202

Selwyn, N. (2022). *Education and technology: Key issues and debates (3rd ed.)*. Bloomsbury Academic.

Stahl, G., & Hakkarainen, K. (2021). Theories of learning and collaboration in AI-supported education. *International Journal of Artificial Intelligence in Education, 31*(3), 417–442. https://doi.org/10.1007/s40593-021-00259-9

van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in computer-based learning environments. *Review of Educational Research, 85*(4), 475–511. https://doi.org/10.3102/0034654314564881

Winne, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed, pp. 63–87). Cambridge University Press.

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

Zhai, X. (2022). ChatGPT user experience: Implications for education. *Smart Learning Environments*, *9*(1), 1–18. https://doi.org/10.1186/s40561-022-00220-3

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Using Gapped Notes to Enhance Students' Learning and Interaction in the Calculus Classroom

Derar Serhan [1 *], Natalie Welcome [2], Adil Ali [3]

[1] Arizona State University, U.S.A, USA, derar@asu.edu, 0000-0003-3767-1046
[2] Arizona State University, U.S.A, USA, nwelcome@asu.edu, 0000-0002-7519-9136
[3] Arizona State University, U.S.A, USA, adilali@asu.edu, 0009-0006-4957-0065
* Corresponding author: Derar Serhan (icemstoffice@gmail.com)

Article Info

Article History
Received:
18 December 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Gapped notes
Students' learning
Students' attitudes
Calculus
Mathematics education

Abstract

The focus of this study was on students' attitudes toward the use of Gapped notes in learning Calculus and their perceptions of its effect on their learning and class interaction. The participants in this study were 75 undergraduate students enrolled in a Calculus course. Data was collected using a 5-point Likert-type questionnaire that included two open-ended questions. The results of this study indicated that students had a positive attitude toward the use of Gapped notes, majority of them indicated that the use of Gapped notes had a positive effect on their learning. In addition, students felt that the use of Gapped notes increased their interaction with the class instructor and their classmates. They also believed that the use of Gapped notes had mostly a positive impact on their classroom participation and engagement. In general, students had a positive experience using Gapped notes as indicated by their responses. On the other hand, students also mentioned that there were some disadvantages to using Gapped notes as being repetitive, lengthy and confusing.

Citation: Serhan, D., Welcome, N., & Ali, A. (2026). Using Gapped Notes to Enhance Students' Learning and Interaction in the Calculus Classroom. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 16–29), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Background and Scope

It might be said that the art of teaching requires strategic intentionality on the part of educators, where specific pedagogical methods are used to deliver information, and additional methods are used to examine whether that information has been reasonably encoded by learners. There are many modalities that educators use to share information in a classroom space. Common techniques used to present learners with information include lecturing (speech presentation), Socratic probing (questioning that leads to desired awareness), and discovery exploration (curated information that actively engages learners). These methods assume that the learner will actively participate. With lecturing, the learner synthesizes and writes down what is heard. Socratic learning might also involve writing down key insights, but learners simultaneously engage in answering questions. Regarding discovery learning, learners examine materials and then document logical conclusions. Across these various methods of instruction, the act of learners writing down relevant information is a commonly threaded practice. Notetaking describes the act of learners writing relevant information, and it is the standard term used in educational settings to describe externalizing information. Notes are the written artifact gleaned directly from an educator's instruction. Students sometimes encounter problems with traditional notetaking, as it is often accompanied by cognitive burden. As learners write, they are challenged with listening, making decisions about what to write, and ultimately writing. There is also a proposed solution to this problem: gapped notes. Gapped notes describe a structured document provided to learners that is developed by an instructional agent, designed to capture the specific details of an instructional experience, and maintains a specified flow of content exploration. They can potentially remove some aspects of the cognitive burden that is notoriously related to traditional notetaking by minimizing the decisions that learners make about what to write. This study seeks to explore students' attitudes and perceptions of the effect of using gapped notes on their learning and classroom interaction. The results of this study benefit educators who are required to implement gapped notes as an element of their instructional methods. With a better understanding of students' attitudes and perceptions of using these pre-structured templates known as gapped notes, classroom educators can move from simple compliance with administrative requirements to evidence-based pedagogical mastery.

Framework

There is an overarching assumption that serves as the foundation of this study: learning calculus is difficult because organizing and storing the information is inherently difficult. In "The Learning and Teaching of Calculus: Ideas, Insights and Activities," Monagan et al. (2024) note that students must experience a metacognitive shift, transitioning from the demand of rote computation in the world of algebra to the demand of abstract reasoning and conceptualization inherent to calculus. Building upon this premise, we assume that the cognitive load of processing calculus concepts can potentially be effectively managed. George Miller, pioneer of the cognitive concept of "chunking," argues that the human brain has a short-term memory capacity that allows for approximately seven chunks of information within a given time (Miller, 1956). In his seminal paper, "The Magical Number Seven, Plus or Minus Two," Miller claimed that information is processed more efficiently when bits of information are consolidated into larger chunks. Additionally, the Atkinson-Shiffrin Model, developed by Richard Atkinson and Richard Shiffrin, specifies that the stages of human memory are sequential, and that information

storage requires a systematic encoding mechanism to transition from short-term memory to long-term memory (Wixted, 2024). We assume that if learners encode and store information using a notetaking mechanism, the overwhelming majority can potentially achieve conceptual mastery of calculus.

Literature

Evolution and Mechanics of Notetaking

Functions of Notetaking

Notetaking is a fundamental aspect of education where learners are active in the process of recording information provided by an educator. A common form and purpose of notetaking is personal documentation and retention, where learners transcribe information onto paper for future review. Beyond traditional methods, notetaking serves as a useful accommodation for learners with special needs. A student with hearing impairments or limited manual mobility may utilize a notetaking assistant to record information on their behalf, which promotes equitable access to the curriculum. Furthermore, the evolution of educational technology has introduced mobile applications and electronic pen-and-paper mechanisms that streamline the documentation of information obtained during instructional time. Regardless of the modality, notetaking facilitates knowledge consolidation at the stage where learners record information that is subsequently retrieved during independent study.

Challenges of Traditional Notetaking

Although the practice of notetaking may appear superficial—recording what has been read or heard—there are significant cognitive hurdles that can hinder the intended outcomes of learning for some learners. For instance, notetakers must decide what information to prioritize when transcribing the contents of a lecture. Verbatim? New words? Overarching subtopics? Research indicates that when students have an enhanced capacity to gather information (e.g. audio recording devices or digital word processing) they demonstrate increased efficiency in discerning what to include in their notes. Conversely, when students rely on constrained methods like handwritten notes, they often experience “decision crises” regarding content selection (Bui et al., 2013). While individual abilities to write, listen, and simultaneously comprehend might vary, the challenge of processing remains evident for all learners. Cognitive overload poses a formidable challenge, particularly for students whose working memory capacity is exceeded by the instructional demand (Piolat, 2007). In the Applied Cognitive Psychology article, “Effects of Note-Taking and Working-Memory Span on Cognitive Effort and Recall Performance” Piolat (2007) found that the act of taking notes while listening requires a significantly greater cognitive demand than listening followed by summarizing. They also observed that providing structured pedagogical support, such as Gapped notes, did not increase cognitive effort when compared to the baseline effort required for traditional notetaking, suggesting that a specific instructional design can streamline an encoding process without furthering cognitive burden.

Standards of Gapped Notes

Nomenclature in Pedagogy

A standard terminology regarding gapped notes might assist with implementation and alignment when applying such strategies to pedagogy. Within educational literature, diverse terminology is employed to describe the mechanisms of pre-organized notes containing strategic blank spaces that facilitate learner engagement with lesson content. Terms such as guided notes, skeletal notes, partial notes, cued notes, and hybrid notes are used interchangeably to convey the pedagogical strategy of utilizing pre-arranged structures to foster active student participation. While the underlying methodology remains consistent, the specific phrasing often depends on the institutional context or theoretical framework applied. In this study, we adopt the phrase "gapped notes," because it is the predominant designation presently utilized at Arizona State University.

Conceptual Foundation of Gapped Notes

In exploring evidence-based practices for serving students with exceptionalities, educational theorist William Heward—widely regarded as a pioneering figure in the development of gapped notes—developed a structured method for assisting learners with notetaking challenges. Originally intended for learners with special educational needs, this strategy intentionally sought to maintain visual focus on instructional materials while maintaining kinesthetic engagement through writing. Heward argued that by providing a conceptual skeleton of the class session, educators can foster sustained student engagement, reduce extraneous cognitive load, and help students discern and prioritize the most critical aspects of the material (Heward, 2022). While initially conceived as an intervention for specialized populations, this strategy of providing pre-organized, guided notes has since been broadly adopted for general student populations. This transition highlights the universal design benefits of gapped notes, as the mechanism effectively supports the varied cognitive processing speeds and organizational needs of all learners.

Structural Elements of Gapped Notes

While the physical layout of the document may vary, gapped notes are defined by several foundational components that facilitate effective information processing. Fundamentally, gapped notes provide learners with the following:

1) Sequential Core Concepts. A gapped notes document provides the framework of the instructional content, organized in a chronologically and logically appropriate manner. By pre-defining a sequence of ideas, the educator ensures that learners follow the intended narrative arc of the lesson, which is particularly critical for the multi-step derivations found in calculus.
2) Opportunities for Active Completion. A defining characteristic of this method is the requirement for students to actively synthesize the notes to completion. This typically manifests as strategic blank spaces where learners must provide words, short phrases, or mathematical steps to complete the partial statements provided by the educator. This mechanism transforms the learner from a passive recipient of information into an active participant in the construction of knowledge.
3) Cues and Prompts. These elements provide directive signposts that signal importance or offer procedural guidance to the learner. For instance, notational symbols such as an asterisk may serve as a visual cue to remind learners that a specific formula is subject to constraints or boundary conditions. These prompts

act as metacognitive triggers, helping learners prioritize information and avoid common pitfalls during the encoding process.

Gapped Notes in Mathematics and Calculus Context

Students often consider math to be the most difficult of the core subjects, more so than social studies, science, and language arts (Purbaningrum et al., 2023). Therefore, logically instructing mathematics content requires greater effort to garner student self-efficacy that helps promote achievement.

Scaffolding Complex Mathematical Ideas

Introductory algebra teaches concepts that generally require learners to navigate procedural steps to achieve a specific solution. However, introductory calculus introduces concepts that require multiple layers of logical deduction beyond the computational requirements of algebra—concepts that reside deeper within the realm of theoretical abstraction. Consequently, learners often require targeted instructional assistance to accurately internalize concepts. Scaffolding is a pedagogical technique that can bridge disparate ideas by providing a temporary support mechanism. Within the context of gapped notes, this involves providing structured templates for the initial introduction of a concept and subsequently increasing the density and complexity of the "gaps" as lessons proceed. This gradual release of responsibility ensures that learners build the necessary stamina to navigate complex ideas while maintaining the structural integrity of any mathematical argument.

Visual-Spatial Learning in Calculus

Learners who are engaged in the study of calculus must constantly translate between abstract visual symbols and the conceptual definitions they represent. For instance, when a differential operator such as $\frac{dy}{dx}$ precedes a function, learners are taught to conceptualize the instantaneous rate of change or the slope of the tangent line to that function. Thus, mathematics functions as a specialized language—one that becomes increasingly complex as the symbolic density of introductory calculus exceeds the foundational notation of introductory algebra. As learners encounter an array of new symbols alongside established algebraic concepts, the risk of cognitive interference increases. By providing a structured layout that explicitly connects the symbol to its visual representation, gapped notes facilitate the dual-coding necessary to encode these complex relationships into long-term memory (Wixted, 2024).

Students' Attitudes and Perceptions of Scaffolding

Students generally believe that notetaking is an effective tool in the learning process and that academic success is directly linked to one's ability to take adequate notes (Kitjaroonchai et al., 2025). It is therefore not uncommon to find students participating in a classroom lecture solely through the arts of listening and notetaking. Questions remain, though, about the effectiveness of independent notetaking. As students demonstrate marked confidence

about the effectiveness of self-directed notetaking, what are their attitudes about utilizing a preorganized document for notetaking, specifically gapped notes? Are they as confident about their ability to learn?

Learner Efficacy

A significant majority of students report a deficiency in self-efficacy regarding mathematical study—a sentiment that intensifies when navigating college-level calculus. In The Learning and Teaching of Calculus: Ideas, Insights and Activities, Monaghan et al. (2024) explore student engagement in elementary, single-variable calculus, illustrating the pedagogical struggles that arise when learners undergo a crisis in their "mathematical identity." When they no longer perceive themselves as competent in the discipline, learners tend to engage in classroom activities with diminished enthusiasm and reduced persistence compared to periods when their self-efficacy was more robust. Furthermore, many university students entering calculus courses maintain a preconceived notion that the subject will be inherently difficult; for some, the task is perceived as cognitively insurmountable (Purbaningrum et al., 2023). This lowered expectancy for success often creates a barrier to learning that precedes actual content exposure. In this context, gapped notes may serve as a psychological scaffold, providing small, achievable “wins” during the notetaking process that can incrementally rebuild a learner’s sense of mastery and academic agency.

Perceived Value Versus Actual Performance

It is widely recognized in pedagogical research that there is no empirical guarantee that students have mastered the material simply because they experienced a fluid lecture or enjoyed the classroom experience (Flaherty, 2019). This discrepancy between subjective experience and objective mastery is a critical consideration in instructional design. Inside Higher Ed reported on a seminal study which revealed that students in active-learning environments often perceived that they learned less, despite demonstrating superior learning outcomes compared to peers in passive-learning settings. This phenomenon, often described as a metacognitive mismatch, suggests that the increased cognitive effort required by active-learning strategies—such as completing gapped notes—can be misinterpreted by students as a sign of ineffective instruction. However, this productive struggle is precisely what facilitates the deep encoding of complex concepts. In the context of calculus, while students might perceive traditional, “polished” lectures as easier to follow, the active synthesis required by gapped notes provides a more accurate scaffold for long-term retention, even if learners initially feel more “challenged” by the process.

Classroom Dynamics and Interaction

Impact of Instructional Pacing

Although gapped notes inherently reduce extraneous cognitive load —as established previously— the pedagogical gains realized through this method may be undermined by a misplaced sense of urgency. If the pace of delivery compels students to rush through the completion of the notes, the opportunity for meaningful encoding is diminished. Effective pacing ensures that students have sufficient cognitive “dwell time” to not only transcribe the required information but also to synthesize the underlying logic of calculus concepts. Instructional synchrony —the alignment between the educator’s delivery and the student’s processing speed—is vital; if the pace exceeds

the learner's ability to internalize the "gaps," the scaffold ceases to be a support and instead becomes a source of performance anxiety. Therefore, the educator must balance the efficiency of pre-organized notes with a deliberate pause-and-reflect strategy to ensure that procedural completion does not supersede conceptual understanding.

Engagement and Behavioral Focus

Zoning out is a real point of consideration for educators who lead learners through challenging calculus content. Gapped notes have been useful in minimizing the zone-out risk that students experience when they are authoring their own notes (Miller, 1956). Students tend to remain focused on the instructional content during the lessons when guided by notes. However, research shows gapped notes often solve some problems while creating others. Feudel and Panse (2021) found that gapped notes removed the potential for students to discover ideas on their own because the professor had revealed the solutions before the learner had an opportunity to be individually thoughtful. For example, learners complained that steps of geometry proofs were prematurely revealed in the gapped notes such that opportunities for learners' self-discoveries were diminished.

Proper Implementation

Delivery of gapped notes must be intentional if educators are maximizing the positive impact of gapped notes. Instructions about how to complete the gapped notes must be clear, the design of gapped notes must be dynamic, and follow-up with students' completion of gapped notes must be applied (Elcoro et al., 2020).

Research Questions

To enhance class learning Gapped notes are used to reduce lecturing time and encourage active learning. The process is to provide students with lecture notes with blank content, which students fill in during class as a group activity that is supervised by the class instructor. This study investigated students' perceptions of using Gapped notes in a Calculus classroom. The aim of this study is to answer the following four questions:

1. What are students' attitudes toward the use of Gapped notes in learning Calculus concepts?
2. What are students' perceptions of the effect of using Gapped notes in learning Calculus concepts?
3. What are students' perceptions of the effect of using Gapped notes on their classroom interaction?
4. What are the students' perceived advantages and disadvantages of using Gapped notes in learning Calculus?

Method

The participants in this study were 75 students at a major university in the USA. The students were enrolled in Calculus for Business class during Fall 2025. The participants filled a 5-point Likert-type 16- item survey that the researchers developed based on literature review and the aim of the study. The scale of the survey responses ranged from Strongly Disagree (1) to Strongly Agree (5). The survey consisted of three sections: students' attitudes

toward the use of Gapped notes, students' perceptions of the effect of using Gapped notes on their learning, and students' perceptions of the effect of using Gapped notes on their interaction with their instructor and classmates. In addition, the survey included two open-ended questions: What did you like about using Gapped notes in this class? What did you dislike about using Gapped notes in this class?

Results and Discussion

In the following analysis, the designation of "agree" includes all "strongly agree" and ""agree" responses while the "disagree" designation includes all "disagree" or "strongly disagree" survey responses. The responses collected from the 5-point Likert-type survey items were grouped into three categories: students' attitudes toward the use of Gapped notes, students' perceptions of the effect of using Gapped notes on their learning, and students' perceptions of effect of using Gapped notes on their interaction with their instructor and classmates. In addition, students' responses to the questions about their likes or dislikes of using gapped notes in the class were tabulated and analyzed.

To provide an answer to the first research question about students' attitudes toward the use of Gapped notes, 78.67 % agreed that they enjoyed using it while 12% disagreed. In addition, 74.67% indicated that they would like to use Gapped notes in other classes, while 12 % disagreed. Also, majority of students (89.33%) mentioned that they felt comfortable using Gapped notes and it was easy to use. This indicates a positive attitude toward the use of Gapped notes in the class (see Table 1).

Table 1. Student ' Attitudes toward Using Gapped Notes

Item	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I felt comfortable using Gapped notes in this class.	45	22	4	4	0
It was easy for me to use Gapped notes in this class.	44	23	5	3	0
I enjoyed using Gapped notes in this class.	43	16	7	9	0
I would like to use Gapped notes in other classes.	33	23	10	6	3

Regarding the second question about students' perceptions of the effect of using Gapped notes in learning Calculus concepts, (82.66%) of students found that using Gapped notes was very helpful. Majority of students thought that using Gapped notes enhanced their learning of the calculus concepts covered in the class as follows: stayed focus (81.33%), increased interest in learning (78.67%), developed confidence in course concepts (80%) and made learning the class content easy (82.67%). In addition, 77.33% of the students reported that the use of gapped prepared them for in class exams. Results indicated that students found that the use of Gapped notes had a positive impact on their learning experience in the classroom (see Table 2).

To answer the third research question about students' classroom interaction, 65.33% agreed that the use of Gapped notes increased their class participation, while 12 % disagreed. Also 66.67% mentioned that it increased their engagement in class discussions while 16% disagreed with that statement. Regarding classroom interaction with the instructor, 58.67% agreed that their interaction with their instructor increased while using Gapped notes in the class, while 13.33% disagreed. In addition, 64% indicated that the use of Gapped increased their interaction with their classmates inside the class with 44% indicated that the use of Gapped increased their interaction with their classmates outside the class, while 26.67% disagreed. This indicates that students felt that the use of Gapped notes increased their interaction with the class instructor and their classmates. They also believed that the use of Gapped notes had mostly a positive impact on their classroom participation and engagement (see Table 3).

Table 2. The Effect of Gapped Notes on Student ' Learning

Item	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Using Gapped notes helped me to stay focused on learning the course concepts.	36	25	6	7	1
Using Gapped notes increased my interest in learning the course concepts.	30	29	16	8	2
Using Gapped notes helped me develop confidence in the course concepts.	38	22	8	6	1
Using Gapped notes helped me learn the class content easily	42	20	6	5	2
Using Gapped notes made me more prepared for exams.	40	18	8	7	2
Using Gapped notes improved my grade in the class.	36	19	10	7	3
Overall, I found the use of Gapped notes very helpful.	46	16	5	5	3

Table 3. The Effect of Gapped Notes on Student ' Interaction

Item	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The use of Gapped notes increased my class participation.	31	18	17	7	2
The use of Gapped notes increased my engagement in class discussions.	29	21	13	10	2
The use of Gapped notes increased my interaction with the class instructor.	23	21	21	9	1
The use of Gapped notes increased my interaction with my classmates inside the class.	20	28	15	10	2
The use of Gapped notes increased my interaction with my classmates outside the class.	15	18	22	18	2

Regarding the fourth research question about advantages and disadvantages of using gapped notes as perceived by the students, students' responses were tabulated based on similarities. Students' perceived advantages were categorized into five categories: Engagement, Focus, Understanding, Ease of Learning, and Ease of Use. The following table (see Table 4) provides a sample of students' responses.

Table 4. Students' Perceived Advantages of Using Gapped Notes

Engagement:
-it helps me stay engage more in class.
-It required effort and I came to class more often.
-They save me time writing everything down. I can easily follow along with the problem. I can look back & study organized notes.
-They helped me follow along in class and stay engaged.
-It helps maintain engagement.
-I found myself to be more engaged in the course.
-I really liked that they made me feel present in class.
-I felt like they kept me engaged and provided me with a clear and understandable way to study.
Focus:
-They flawed well with the lecture and allowed me to more easily focus on what's being verbally said.
-It really kept me focused throughout the class.
-I had more time to listen to the professor rather than speed writing notes.
-I liked how it made me pay more attention in class.
-kept me focused.
-Gapped notes allowed me to listen more to the professor's lecture and worry less about frantically writing everything down.
-Helped me focus more in class, helpful for exams.
Understanding:
-The notes were easy to understand even if I am bad at math.
-The concept of the math is easier to grasp.
-They made it easier to understand the material during the lecture.
-helped break up the material of the course, better understanding of each unit.
Helped me understand the concepts.
-It gave me all the info I needed without needing to write other than the key concepts.
-good information.
-The Gapped notes helped me better understand the concepts of calculus and helped me achieve good scores on tests.
-I find myself normally losing interest in class or paying attention but with the Gapped notes I was able to do a lot better because it guided me in what was important to write down. Especially it was organized rather than

my messy notes.

Ease of Learning:

-It helped me stay focused and made the lessons easier to follow.
-They made learning the course easy.
-I learned very easy when following along.
-very hands on.
-I was able to review them outside of class.
-Being able to follow along during class.
-I like that the concepts were easily laid out.
-They cover all important topics in the unit.
-helpful to look back on what we learned.
-Gave me a more visual example
- Not having to write on a lot of stuff down helps me keep pace with class and helps me focus on the content.

Ease of Use:

-simplicity of the notes.
-They were easy to follow along with and had examples.
-easier to follow along & know what is important.
-easy to follow and study tool.
-It made the class straight forward.
-It was a good way to be able to review the course material before a quiz/exam.
-They were clear, simple and easy to follow.
-Made it very easy to follow along.
-The Gapped notes were easy to use and convenient.
-They're straight to the point & concise.
-It helps me stay organized with my note taking.

While majority of students, 66.67% indicated that they had no dislikes about the use of Gapped notes, others mentioned few dislikes. Students perceived disadvantages of using Gapped notes in the class were categorized as follows: Pacing, Repetitive and Lengthy, Confusing, Learning, Organization and Nothing. The following table (see Table 5) provides a sample of students' responses.

Table 5. Students Perceived Disadvantages of Using Gapped Notes

Pacing:

-sometimes the notes felt a little too fast to fill in during class.
-Sometimes it would get a little fast.
-I disliked the pacing, I feel like some units were much more difficult and took longer to understand than

others.

Repetitive and Lengthy:

-The only thing I disliked was how some questions were repetitive.
-Its very repetitive and not engaging.
-They were repetitive sometimes.
-Had a lot of extra filler info.
-lengthy
-sometimes too wordy.
-it can be wordy.
-long

Confusing:

-It looked confusing and unorganized.
-The way some concepts are phrased can be confusing.
-unnecessary words.
-a bit confusing.
-sometimes confusing.
-sometimes felt that like the transition between concepts was a little confusing.

Learning:

-not the best to study with.
-I don’t believe that the Gapped notes increased my class engagement.
-I don’t appreciate having notes done for me.
-If you miss a part you gotta ask to go back or ask someone next to you.

Organization:

-Not enough space to do work or examples fully.
-long questions/poor layout made learning difficult.

Nothing:

-I liked it, I don’t dislike anything about it.
-Nothing was bad that I really took notice to.

Conclusion

Gapped notes—alternatively identified as guided notes, slotted notes, skeletal frameworks, etc—are increasingly implemented in higher education settings, particularly within rigorous mathematics curricula, to assist students in managing the intensive cognitive load demands inherent to academic achievement and enhance their learning and engagement in the classroom. This study investigates the intersection of instructional design and student experience through four primary inquiries: 1) What are students' attitudes toward the use of gapped notes in learning abstract calculus concepts? 2) What are students' perceptions of the pedagogical efficacy of using gapped notes in learning calculus concepts? 3) What are students' perceptions of the effect of using gapped notes on their perceived level of classroom interaction? 4) What are students' perceived advantages and disadvantages of using gapped notes in learning higher-level mathematics? As educators in the higher education community consider integrating gapped notes into their curricula, it is imperative to appreciate learner perspectives on such instructional intervention. Research has shown that implementing gapped notes significantly impacts instructional pacing, influences the behavioral engagement of learners, and requires strategic, intentional implementation to ensure optimal learning outcomes. This research aims to help establish agreement between administrative mandates and lived experiences of learners in the higher education calculus classroom. The results of this study indicated that students had a positive attitude toward the use of Gapped notes and that they perceived these as having a positive effect on their learning and it increased their interaction with the class instructor and their classmates. In general, students had a positive experience using Gapped notes as indicated by their responses.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Ethical approval was obtained from the institutional review board. Participation was voluntary; informed consent was secured prior to data collection. No identifiable data were stored or reported. The study's focus on learning behaviors minimized potential risks to participants.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Bui, D. C., Myerson, J., & Hale, S. (2013). Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology*, *105*(2), 299–309. https://doi.org/10.1037/a0030367

Elcoro, M., Ghirardi-Broughton, K., & Hebert, L. (2020). Guided notes in undergraduate instruction. *Journal of Scholarly Engagement*, *3*(2), 48–62. https://doi.org/10.9743/jse.2020.3.2.3

Feudel, F., & Panse, A. (2021). Can guided notes support students' note-taking in mathematics lectures? *International Journal of Research in Undergraduate Mathematics Education*, *8*(1), 8–35. https://doi.org/10.1007/s40753-021-00146-9

Flaherty, C. (2019). *Study: How smooth-talking professors can lull students into thinking they've learned more than they have*. Inside Higher Ed | Higher Education News, Events and Jobs. https://www.insidehighered.com/news/2019/09/09/study-how-smooth-talking-professors-can-lull-students-thinking-theyve-learned-more

Heward, W. L. A. (2022). *Exceptional children: An introduction to special education (12th ed.)*. Merrill.

Kitjaroonchai, N., Marshall, P., Phasook, J., & Thararuenroeng, N. (2025). University students' perceptions on note-taking and its impact on academic performance. *JET (Journal of English Teaching)*, *11*(1), 1–19. https://doi.org/10.33541/jet.v11i1.6373

Miller, G. A. (1956). *The magical number seven, plus or minus two: Some limits on our capacity for processing information*. Bobbs-Merrill.

Monaghan, J., Ely, R., Pinto, M. M. F., & Thomas, M. O. J. (2024). *The learning and teaching of calculus: Ideas, insights and activities*. Routledge.

Piolat, A. (2007). Chapter 8: Effects of note-taking and working-memory span on cognitive effort and recall performance. Writing and Cognition, 109–124. https://doi.org/10.1163/9781849508223_009

Purbaningrum, M., Ramadhan, S., & Thauzahra, R. (2023). Why is math difficult? : Beliefs that affecting students' mathematics skills. *Jurnal Paedagogy*, *10*(4), 1000. https://doi.org/10.33394/jp.v10i4.8652

Wixted, J. T. (2024). Atkinson and Shiffrin's (1968) influential model overshadowed their contemporary theory of human memory. *Journal of Memory and Language*, *136*, 104471. https://doi.org/10.1016/j.jml.2023.104471

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Prompt Engineering as Metacognitive Scaffolding: A Self-Regulated Learning Theory Framework for AI-Mediated Education

Eyüp Yurt

Bursa Uludağ University, Turkiye, 0000-0003-4732-6879
Corresponding author: Eyüp Yurt (icemstoffice@gmail.com)

Article Info

Article History
Received:
11 January 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Prompt engineering
Self-regulated learning
Generative AI
Metacognitive awareness

Abstract

The emergence of generative artificial intelligence (AI) in education expands opportunities for student self-regulation while introducing new challenges for meaningful learner–AI interaction. This theoretical study conceptualizes prompt engineering—the intentional formulation and iterative refinement of queries to generative AI—as a core metacognitive process within self-regulated learning (SRL) frameworks. Rather than a peripheral technical skill, prompt engineering serves as metacognitive scaffolding that supports learners' planning, monitoring, and evaluation. This iterative cycle is framed as second-order metacognition, where learners reflect on task performance as well as the adequacy of their goals, strategies, and mental representations. Drawing on six recent empirical studies, the paper demonstrates that explicit instruction in prompt engineering enhances students' metacognitive awareness, self-regulatory capacity, and learning outcomes. Consequently, effective AI-mediated education requires reconceptualizing prompt engineering as an integral component of metacognitive co-regulation. As AI increasingly provides adaptive feedback, communicating effectively with these systems emerges as a fundamental learning competence. The paper concludes by proposing an instructional framework structured around an iterative Plan–Iterate–Evaluate cycle, designed to integrate prompt engineering into SRL instruction to foster both AI literacy and advanced metacognitive skills.

Citation: Yurt, E. (2026). Prompt Engineering as Metacognitive Scaffolding: A Self-Regulated Learning Theory Framework for AI-Mediated Education. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 30–48), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

The integration of generative artificial intelligence into educational contexts has fundamentally altered how students encounter learning resources and receive instructional support. Rather than consulting textbooks, asking human tutors, or conducting traditional research, students increasingly interact with AI systems by formulating questions, requesting explanations, and seeking guidance. This shift represents not merely a change in the medium through which instructional support is delivered but a transformation in the cognitive demands placed on learners. To benefit from generative AI, students must develop sophisticated capacity to formulate effective queries—to understand what questions will elicit useful responses, how to refine questions when initial responses prove inadequate, and how to integrate AI-generated information into their learning.

This capacity —commonly termed "prompt engineering"— has typically been understood as a technical skill: a collection of techniques for optimizing queries to AI systems. Practitioners and educators have developed heuristics for effective prompting: providing context, specifying desired format, requesting step-by-step reasoning, asking the AI to adopt particular personas or roles. These techniques have been compiled into guides and tutorials presenting prompt engineering as a practical craft to be learned and applied. However, recent research suggests a more profound conceptualization of prompt engineering. Rather than merely a technical skill peripheral to learning processes, prompt engineering functions as a sophisticated form of metacognitive scaffolding—a mechanism through which learners externalize their thinking, reflect on the quality of their own mental representations, and iteratively refine their understanding. When students carefully craft prompts, evaluate AI responses, and reformulate queries based on their evaluation, they engage in metacognitive processes essential for self-regulated learning. They are not simply operating a tool; they are engaging in deliberate reflection on their own goals, understanding, and learning strategies.

Moreover, contemporary educational contexts increasingly demand sophisticated measurement and understanding of self-regulation skills in AI-assisted learning environments. Traditional self-regulation scales were developed in pre-AI contexts and inadequately capture the unique dynamics of learning with AI systems. Recent instrument development work has begun to address this gap, creating comprehensive scales that measure self-regulation specifically designed for AI-augmented learning contexts. This theoretical study reconceptualizes prompt engineering through a self-regulated learning framework, arguing that effective AI-mediated education requires explicit instruction in prompt engineering as a metacognitive practice. Drawing on six recent empirical investigations—including studies examining measurement of self-regulation in AI contexts—the paper demonstrates that students who receive instruction positioning prompt engineering as metacognitive scaffolding develop superior self-regulation capacity, metacognitive awareness, and learning outcomes compared to students using AI without such instruction.

Theoretical Foundations: Self-Regulated Learning and Metacognition

Self-Regulated Learning: Definition and Core Processes

Self-regulated learning (SRL) refers to learners' capacity to direct their own learning processes—to set goals,

select and implement learning strategies, monitor progress, and evaluate outcomes (Zimmerman, 2002). Rather than passively receiving instruction, self-regulated learners actively manage their learning by making deliberate choices about what to learn, how to approach learning tasks, and whether their current approach is effectively achieving their learning goals.

Zimmerman's cyclical model of self-regulated learning identifies three phases: (1) forethought, wherein learners analyze tasks, set goals, and activate relevant prior knowledge; (2) performance/volitional control, wherein learners implement strategies, maintain focus, and monitor their progress; and (3) self-reflection, wherein learners evaluate outcomes, reflect on what worked and what did not, and adjust strategies accordingly. These phases are not linear but cyclical; reflection from one learning attempt informs planning for subsequent attempts.

Self-regulated learning proves consistently predictive of academic achievement across diverse educational contexts and student populations (Zimmerman & Schunk, 2011). Students who actively regulate their learning—who deliberately select strategies suited to their tasks, who monitor whether strategies are working, who adjust approaches based on evidence—demonstrate superior learning outcomes. Conversely, students who passively receive instruction without actively directing their learning processes demonstrate lower achievement and develop less sophisticated learning capacities.

Metacognition as Core SRL Mechanism

Metacognition—awareness and regulation of one's own cognitive processes—operates as a central mechanism through which self-regulated learning functions. Metacognitive awareness involves explicit knowledge about one's cognitive strengths and limitations, understanding of different learning strategies and their effectiveness, awareness of one's current understanding, and recognition of when one does not understand. Metacognitive regulation involves actively monitoring one's thinking as it unfolds, evaluating whether current approaches are working, recognizing when comprehension is breaking down, and adjusting strategies accordingly.

Flavell (1979) distinguishes between metacognitive knowledge (understanding about cognition) and metacognitive monitoring (ongoing awareness of cognitive processes). Both dimensions prove essential for self-regulated learning. A learner might possess knowledge that elaboration enhances memory but fail to actively monitor whether they are adequately elaborating new information. Conversely, a learner might actively monitor their comprehension but lack knowledge of effective strategies to address comprehension failures.

Critically, metacognition is not a unitary cognitive capacity but emerges through engagement with specific tasks and domains. A student might demonstrate sophisticated metacognitive awareness in mathematics but less developed metacognitive capacity in writing. Moreover, metacognition develops progressively through educational experiences that explicitly cultivate reflection and self-monitoring. Students who receive instruction emphasizing metacognitive awareness and regulation develop stronger metacognitive capacities than students receiving standard instruction.

Scaffolding: Supporting Self-Regulation and Metacognition

Scaffolding—temporary support structures that enable learners to accomplish tasks beyond their current independent capacity—represents a crucial mechanism for developing self-regulated learning and metacognitive capacity. Effective scaffolding involves providing external support that gradually decreases as learners develop capability, maintaining learners at the edge of their current competence.

Pea (1987) identifies several scaffolding functions: constraining task difficulty by breaking complex tasks into manageable components, maintaining orientation toward task goals, highlighting critical task features, providing models of expert performance, and offering reflective prompts that cue metacognitive awareness. Importantly, different learners require different types and amounts of scaffolding; effective scaffolding is responsive to individual learner needs.

In the context of AI-mediated learning, scaffolding might involve explicit instruction in effective prompting strategies, guided practice formulating and refining prompts, feedback on prompt quality, or prompts within the AI system itself cuing effective query formulation. Such scaffolding supports learners in developing self-regulation with AI by helping them understand which prompting strategies are effective, how to monitor whether prompts are generating useful responses, and how to adjust prompting approaches when initial attempts prove unsuccessful.

Second-Order Metacognition: Reflecting on Reflecting

An important distinction emerges between first-order metacognition (reflecting on one's own cognitive processes) and second-order metacognition (reflecting on the quality and effectiveness of one's metacognitive processes themselves). When students reflect on whether they understand a concept, they engage in first-order metacognition. When students reflect on the quality of their own reflection—wondering whether their evaluation of their understanding is accurate, questioning whether they are using the most effective monitoring strategies—they engage in second-order metacognition. Alatoai and Alshahri (2025) introduce a crucial insight: prompt engineering with generative AI functions as second-order metacognition. When students craft prompts and evaluate AI responses, they are essentially asking: "Did my query effectively communicate what I wanted to know? What does the AI's response tell me about the clarity and specificity of my query? How can I refine my query to better communicate my actual question?" This recursive reflection on the quality of one's own questioning represents sophisticated second-order metacognition: learners use external feedback (AI responses) to infer the quality of their own internal mental representations and question formulation.

Measuring Self-Regulation in Ai-Assisted Learning Contexts

The Need for AI-Specific Self-Regulation Measurement

Traditional self-regulation scales, developed in pre-AI educational contexts, inadequately capture the unique dimensions of self-regulation required in AI-assisted learning environments. These conventional instruments

assess general learning strategies, time management, and metacognitive processes, but they fail to address the distinctive self-regulation challenges and opportunities created by AI integration.

Yurt (2025) directly addresses this gap through development of the Self-Regulation for AI-Based Learning Scale (SRAILS), a comprehensive instrument specifically designed to measure self-regulation in university students' AI-assisted learning. This instrument development work responds to a critical need: as AI technologies become integral to university learning environments, educators and researchers require valid, reliable measurement tools that capture the distinctive self-regulation competencies students must develop.

The Self-Regulation for AI-Based Learning Scale (SRAILS): Comprehensive Framework

Yurt (2025) developed SRAILS through extensive psychometric validation with 750 university students from various Turkish faculties. The scale comprises four main dimensions with nine sub-dimensions, reflecting a sophisticated understanding of self-regulation in AI contexts:

Motivational Components (2 Sub-Dimensions):

- Intrinsic and extrinsic motivation
- Self-efficacy beliefs regarding AI use

Cognitive and Metacognitive Strategies:

- Strategic planning and goal-setting with AI
- Metacognitive monitoring of learning processes
- Reflective evaluation of AI-assisted learning

Time and Task Management:

- Effective temporal organization of AI-assisted learning
- Strategic task decomposition and sequencing
- Persistence and effort regulation with AI tools

Resource Management and Technological Adaptation:

- Effective selection and utilization of AI tools
- Strategic use of supplementary learning resources
- Adaptation to technological affordances and constraints

The scale's psychometric properties are robust: Exploratory Factor Analysis supported construct validity, while Confirmatory Factor Analysis confirmed the nine-factor structure with good model fit. Cronbach's alpha and

McDonald's omega reliability coefficients ranged from .853 to .913, demonstrating excellent internal consistency. Convergent and discriminant validity were established, and criterion-related validity was confirmed through significant positive correlations with external criteria: academic GPA ($r = .20–.60$), technology interest, and digital literacy levels ($p < .01$).

Yurt (2025) concludes that SRAILS "provides a comprehensive, reliable assessment of students' self-regulation skills in AI-assisted learning environments" and "contributes originally to literature by enabling personalized learning experience design and supporting effective instructional strategy development."

Implications of AI-Specific Self-Regulation Measurement

The development of SRAILS and similar AI-context-specific instruments has important implications for understanding self-regulated learning with AI. First, it validates the theoretical claim that self-regulation in AI-assisted contexts differs meaningfully from traditional self-regulation. The distinctive sub-dimensions identified in SRAILS—particularly technological adaptation and strategic AI tool selection—represent forms of self-regulation that were absent or marginal in pre-AI contexts.

Second, SRAILS enables educators to diagnose students' specific self-regulation strengths and weaknesses in AI contexts. Rather than assuming all students require identical support, educators can use SRAILS assessment results to identify which students need stronger metacognitive scaffolding, which require better motivation support, and which need development in technological adaptation skills. This diagnostic capability enables the personalized learning experience design Yurt (2025) emphasizes.

Third, SRAILS provides a research instrument enabling investigation of how self-regulation in AI-assisted learning develops, what pedagogical approaches most effectively support such development, and how different self-regulation dimensions interact with learning outcomes. Future research using SRAILS can build evidence regarding which self-regulation competencies are most critical for learning success with AI, and how these competencies can be most effectively cultivated.

Empirical evidence: Prompt Engineering as Metacognitive Scaffolding in SRL

AI-Enhanced SELF-Regulated Learning Frameworks

Alatoai and Alshahri (2025) developed a comprehensive framework extending self-regulated learning theory to explicitly incorporate AI-specific learning behaviors. Their "AI-Supported Self-Regulated Science and Mathematics Learning Scale" (AI-SSRSML) for secondary students identifies prompt engineering as a distinctive AI-era learning behavior: "prompt engineering strategies for soliciting tailored help from AI tutors."

Critically, Alatoai and Alshahri (2025) position prompt engineering as fundamentally metacognitive. They explain that prompt engineering constitutes "a form of second-order metacognition in which learners use the quality of model output to infer the quality of their own queries, goals, and mental representations." In other words, when

students carefully craft prompts and evaluate AI responses, they engage in sophisticated self-reflection about their own thinking. The quality of AI output provides evidence about the quality of their prompting; gaps between desired responses and actual responses signal that their query formulation needs refinement or their underlying question needs clarification.

This reconceptualization has profound implications. Rather than treating prompt engineering as a technical skill peripheral to learning, it becomes central to metacognitive self-regulation. Within the cyclical SRL framework, prompt engineering represents a key mechanism through which learners: (1) clarify their goals and questions (forethought phase); (2) formulate and implement strategies for soliciting help (performance phase); and (3) evaluate the effectiveness of their prompting strategies based on AI responses (reflection phase).

Alatoai and Alshahri (2025) emphasize that this metacognitive dimension of prompt engineering represents a significant expansion of traditional SRL theory. When learners reflected on their understanding using only human resources, they engaged in one type of metacognitive monitoring. When they reflect on their understanding by formulating queries to AI, evaluating responses, and reformulating queries, they engage in a qualitatively different metacognitive process where external systems provide feedback about their question quality.

Generative AI Chatbots Enhancing Self-Regulated Learning Outcomes

Ng et al. (2024) conducted a pilot experimental study comparing a ChatGPT-based learning chatbot ("SRLbot") against traditional rule-based tutors for secondary science instruction. The SRLbot was specifically designed to support students' self-regulated learning through adaptive prompts and personalized recommendations aligned with SRL principles.

The study examined 120 secondary science students across a three-week unit. The ChatGPT-enhanced group demonstrated significantly higher gains in science knowledge compared to the rule-based tutor control group. More importantly for understanding prompt engineering's role in SRL, the number of AI interactions—the frequency with which students engaged in dialogue with the AI—strongly predicted variations in students' self-regulated learning outcomes.

Students reported that the SRLbot's adaptive prompts and personalized recommendations helped them in crucial SRL processes: setting learning goals, solving problems, and establishing regular study habits. The ChatGPT-based design reduced student anxiety and boosted sustained effort—key indicators of self-regulated learning commitment. Notably, students described the interaction pattern with the SRLbot as an iterative dialogue: they formulated questions, received responses, evaluated responses, and refined subsequent questions based on their evaluation.

Ng et al. (2024) conclude that "generative AI prompts can empower learners' self-regulation." The critical mechanism appears to be that adaptive AI—responding to students' specific queries with personalized guidance—creates conditions supporting iterative prompt refinement. As students engage in successive interactions with

the AI, they learn which prompting strategies elicit helpful responses, progressively developing more sophisticated prompting capability and associated metacognitive awareness.

Explicit Prompt Engineering Instruction and Autonomous Learning

Mzwri and Turcsányi-Szabo (2025) evaluated a university-level elective course explicitly teaching students how to formulate effective prompts for large language models. The course, titled "I Learn with Prompt Engineering," integrated prompt-engineering lessons with self-regulated learning principles and used an AI-driven assessment tool to provide immediate feedback.

The study tracked 45 university students across a full semester. Results demonstrated that students significantly improved their prompting skills, measured through pre-post assessments of their ability to formulate various prompt types (persona-based prompts, instructional prompts, creative prompts). Importantly, beyond prompting skill gains, students showed substantial improvements in academic English proficiency—the actual disciplinary learning outcome. Gains in English proficiency were largest among students with lower starting levels, suggesting that explicit prompt engineering instruction particularly benefits struggling learners.

Students praised the AI assessment tool (EnSmart) for providing immediate, human-like feedback on their written prompts and resulting outputs. Crucially, Mzwri and Turcsányi-Szabo (2025) attribute the course's success to "careful course design incorporating SRL phases (planning, monitoring, evaluating)." The course explicitly scaffolded students through each phase of the SRL cycle as they engaged with prompt engineering:

Planning phase: Students analyzed learning objectives, identified what information or explanation they needed, and formulated specific prompts seeking that information.
Monitoring phase: Students reviewed AI responses, evaluated whether responses addressed their actual question, and reflected on whether their prompt was specific enough.
Evaluation phase: Students assessed the quality of both their prompts and the resulting outputs, identifying what made prompts effective or ineffective.

This explicit structure kept students engaged in metacognitive reflection throughout their engagement with AI, transforming prompt engineering from mechanical operation into a self-regulated learning process.

Human-in-the-Loop Prompt Engineering within Iterative Learning Cycles

Anders and Speltz (2025) implemented a mixed-methods writing course explicitly weaving generative AI literacy into a self-regulated learning cycle structured as Plan–Iterate–Evaluate. Students engaged in writing projects incorporating generative AI as a collaborative tool while remaining the primary decision-makers and evaluators of the work.

By semester's end, students demonstrated large gains in AI-literacy self-efficacy—their confidence in using AI

effectively for writing. More importantly, qualitative analysis revealed concrete "human-in-the-loop" practices reflecting sophisticated self-regulated prompt engineering:

In the planning phase, students activated prior knowledge about their writing projects, researched their topics, and developed clear understanding of their writing goals before formulating initial prompts to AI. Rather than asking AI to immediately generate writing, they used AI as a thinking tool to explore their own understanding.

In the iteration phase, students developed multi-step workflows engaging with AI across multiple prompting cycles. They formulated initial prompts, evaluated responses, identified gaps or issues in responses, reformulated prompts addressing those gaps, and progressively refined their work through successive AI-assisted cycles. Critically, students-maintained agency throughout: they made deliberate choices about which AI suggestions to incorporate, which to modify, and which to discard based on their own evaluation.

In the evaluation phase, students reflected on their collaboration with AI, examining what prompting strategies worked effectively, where they needed to refine prompting approaches, and how their own writing developed through the iterative AI-assisted process. They developed explicit awareness of how different prompt formulations generated different quality outputs.

Anders and Speltz (2025) emphasize that this structure represents "scaffolded prompt-based projects helping students direct their own learning with AI." The explicit Plan–Iterate–Evaluate cycle provided scaffolding maintaining students within their zone of proximal development while supporting the development of more sophisticated prompting strategies and associated metacognitive awareness. Students concluded that "scaffolded prompt-based projects help students direct their own learning with AI" (Anders & Speltz, 2025).

Prompt Engineering as Core SRL Strategy in Language Education

Zhang et al. (2025) conducted a systematic scoping review surveying empirical studies of self-regulated learning with AI in language education. Their analysis of 64 studies published between 2019 and 2024 explicitly conceptualizes prompt engineering as a central SRL strategy.

Zhang et al. (2025) emphasize that "SRL with AI often involves co-regulation through prompting cycles." Rather than individual learners alone regulating their learning, learners and AI systems engage in collaborative regulation wherein the learner's prompting strategies shape AI responses, which in turn inform learner metacognition and subsequent prompt refinement. The authors note that learners "actively participate in a cycle of prompting and re-prompting" with AI tools, using their goals to continuously adjust queries based on AI output.
Explicitly defining prompt engineering as "the process of designing, refining, and optimizing input prompts to elicit desired responses," Zhang and colleagues (2025) describe how language learners employed prompt engineering as an SRL strategy. Learners iteratively reformulated questions and critically evaluated AI-generated feedback. Importantly, this iteration was not random trial-and-error but systematic: learners modified specific aspects of prompts based on their analysis of previous responses, progressively developing more targeted,

effective queries.

Zhang et al. (2025) conclude by portraying "prompt formulation and refinement as intrinsic SRL strategies: students monitor the effectiveness of their prompts as part of metacognitive regulation." This positioning represents a significant theoretical shift: prompt engineering is not a technical skill parallel to learning strategies but rather a fundamental mechanism through which learners engage in metacognitive self-regulation.

AI as Co-Regulator of Learning: Distributed Metacognitive Scaffolding

Tsakeni and colleagues (2025) conducted a bibliometric-systematic review analyzing 187 studies examining AI in STEM education from 2005-2025. Their analysis reveals important insights about how contemporary AI systems function within self-regulated learning frameworks.

Tsakeni et al. (2025) observe that modern AI tools, including generative platforms, are shifting "the focus from internal reflection to distributed cognition." Rather than learners alone engaging in metacognitive reflection, AI systems increasingly function as co-regulators of learning, externalizing self-monitoring processes through dashboards, analytics, adaptive hints, and personalized feedback.

For example, when an adaptive AI tutoring system provides hints, these hints externalize the planning process: rather than students internally deliberating about problem-solving approaches, the system presents possible approaches. When AI provides performance analytics, it externalizes the evaluation process: rather than students solely internally assessing their progress, the system displays progress visually. This externalization of metacognitive processes represents what Tsakeni and colleagues term "distributed cognition"—metacognitive regulation dispersed across human and machine systems.

Critically, Tsakeni et al. (2025) argue that this distribution does not eliminate human metacognition but rather transforms it. Learners still engage in metacognitive reflection, but they do so in dialogue with external AI systems providing information, suggestions, and scaffolding. Within this framework, understanding how to effectively query AI systems—prompt engineering—becomes part of the metacognitive scaffolding. As students collaborate with AI, "formulating the right questions is itself a cognitive skill supported by the system."

Integrating Prompt Engineering into Self-Regulated Learning Instruction

Extending SRL Theory: AI-Augmented Self-Regulated Learning Cycles

The empirical evidence reviewed above, informed by comprehensive measurement frameworks like SRAILS, suggests that self-regulated learning theory requires expansion to explicitly address how learners self-regulate with AI systems. Traditional SRL theory conceptualized learners as directing their own learning through individual goal-setting, strategy selection, monitoring, and reflection. However, when AI systems are integrated as learning partners, the self-regulated learning process becomes more complex: learners must not only regulate their learning but also regulate their interaction with AI systems.

The expanded SRL cycle for AI-mediated learning, incorporating dimensions measured by SRAILS, might be conceptualized as follows:

Planning Phase (Enhanced):

- Analyze learning task/goal (as in traditional SRL)
- Identify what information or explanation is needed to accomplish the goal
- Activate motivation for AI-assisted learning (intrinsic/extrinsic motivation)
- Develop self-efficacy beliefs about effective AI use
- Formulate initial query/prompt to AI, specifying desired information type and format
- Activate prior knowledge about effective prompting strategies
- Activate technological adaptation awareness regarding particular AI tools' affordances

Performance/Iteration Phase (Expanded):

- Implement initial prompt to AI
- Monitor AI response quality relative to actual information need (time and task management)
- If response is adequate, integrate AI output into learning process
- If response is inadequate, engage in metacognitive reflection: What about my prompt was unclear? What assumptions did the AI make? How should I refine my query?
- Select appropriate technological resources and strategies for refinement (resource management)
- Reformulate and resubmit prompt with refinements
- Repeat iterative prompting cycles until satisfactory response obtained
- Integrate AI output with prior knowledge and engage in elaboration
- Manage temporal demands and task sequencing effectively

Evaluation/Reflection Phase (Enhanced):

- Evaluate learning outcome and AI-assisted process
- Reflect on what prompting strategies were effective/ineffective
- Assess self-efficacy development and motivational changes through AI use
- Identify implications for subsequent AI use (What prompting approaches work best for this type of information? How should I adjust prompting for different query types?)
- Develop emerging prompting expertise and technological adaptation awareness
- Plan modifications to approach for future AI-assisted learning

This expanded cycle explicitly positions prompt engineering as integral to self-regulated learning rather than peripheral to it, and incorporates the distinctive dimensions of self-regulation in AI contexts measured by instruments like SRAILS.

Pedagogical Principles for Scaffolding Prompt Engineering as Metacognitive Practice

Drawing on the empirical studies reviewed above, several pedagogical principles emerge for explicitly scaffolding prompt engineering within self-regulated learning instruction:

Principle 1: Make Metacognitive Processes Explicit

Rather than assuming students will spontaneously recognize the metacognitive dimensions of prompt engineering, instruction should explicitly address how prompt quality reflects question clarity and comprehensiveness. Students should engage in activities where they explicitly examine their own prompts, identify vague or incomplete elements, and consciously reformulate prompts. This metacognitive transparency helps students develop second-order metacognitive awareness—awareness of their own query formulation processes.

Principle 2: Scaffold Iterative Prompt Refinement

Rather than expecting students to formulate perfect prompts independently, instruction should scaffold the iterative refinement process. This might involve: (1) providing templates or frameworks for different prompt types; (2) guided practice where students formulate prompts, receive feedback on prompt quality, and engage in guided refinement; (3) explicit discussion of what made some prompts more effective than others; (4) opportunities for students to compare their prompts and refined versions, reflecting on what changed and why.

Principle 3: Connect Prompting to Learning Goals and Motivation

Prompt engineering instruction should emphasize the relationship between clear prompting and goal achievement, addressing the motivational and self-efficacy dimensions measured in SRAILS. Students should engage in authentic tasks where they must formulate prompts to accomplish specific learning objectives, then evaluate whether prompts successfully generated the information needed for those objectives. This connection between prompting, goal-achievement, and metacognitive reflection helps students develop both intrinsic motivation and self-efficacy regarding AI-assisted learning.

Principle 4: Develop Prompt Strategy Knowledge and Technological Adaptation

Students should develop explicit knowledge of different prompting strategies suited to different query types and different AI tools: strategies for generating explanations versus generating examples, strategies for obtaining step-by-step reasoning versus conceptual overview, strategies for role-based or persona-based responses. This metacognitive knowledge about prompting strategies, combined with understanding how different AI tools afford different capabilities, allows students to deliberately select strategies and tools matched to their actual information needs rather than using trial-and-error. This dimension directly addresses technological adaptation, a key component of SRAILS.

Principle 5: Integrate Prompting Into Authentic Disciplinary Tasks

Rather than teaching prompt engineering as an isolated skill, it should be integrated into authentic disciplinary learning tasks. Students in writing courses might engage in AI-assisted writing where they develop prompting expertise within the context of actual writing projects. Science students might use AI for learning explanations within the context of actual scientific problem-solving. Language learners might use AI for practice dialogues integrated into communicative language tasks. This authentic integration helps students develop prompting expertise as a natural dimension of disciplinary learning rather than as a separate skill.

Principle 6: Foster Critical Evaluation of AI Outputs and Resource Management

Prompt engineering instruction should emphasize that obtaining AI responses is not the endpoint but the beginning of critical engagement. Students should develop capacity to evaluate AI outputs for accuracy, bias, completeness, and alignment with their actual information needs. This critical evaluation represents crucial metacognitive regulation: students must monitor not only their own prompting but also the quality and appropriateness of AI-generated information. Additionally, instruction should help students manage AI tool selection and supplementary resource use (resource management), recognizing when AI is appropriate versus when other learning resources are preferable.

Assessment Approaches for Prompt Engineering in Self-Regulated Learning

If prompt engineering is understood as a metacognitive SRL process encompassing dimensions like those in SRAILS, assessment should evaluate not merely prompt quality as a technical product but the metacognitive reasoning and self-regulatory capacity underlying prompt formulation. Assessment might include:

Prompt Rubrics:

Assessment rubrics should evaluate prompts based on clarity, specificity, alignment with learning goals, evidence of technological adaptation, and evidence of metacognitive refinement. Rather than scoring prompts as simply "good" or "poor," rubrics should recognize increasingly sophisticated prompting reflecting deeper metacognitive awareness and self-regulation.

Metacognitive Reflection Prompts:

Students might be asked to explain their reasoning for specific prompt formulations: "Why did you formulate your prompt this way? What information or explanation were you seeking? How did you decide what details to include? What self-regulation strategies did you employ?" These reflections reveal students' metacognitive reasoning and self-regulatory processes.

Self-Regulation Scale Assessment:

Using instruments like SRAILS, educators can assess students' broader self-regulation development in AI-assisted contexts, identifying specific dimensions (motivation, metacognitive strategies, time management, resource management, technological adaptation) requiring additional support.

Iterative Prompt Revision:

Rather than assessing a single prompt, instructors might assess students' ability to revise prompts based on evaluation of AI responses. Do students identify specific issues in AI responses? Do their revised prompts address those issues? This assesses metacognitive regulation in context.

Transfer of Prompting Strategies:

Perhaps most importantly, the assessment should evaluate whether students transfer prompting strategies to new contexts and query types. Have students learned specific prompting approaches or have they developed generalizable metacognitive capacity to formulate effective prompts across diverse situations and tools?

Implications for Educational Practice and Instructional Design

Diagnostic Assessment and Personalized Support

The development of comprehensive self-regulation measurement instruments like SRAILS enables educators to conduct diagnostic assessment identifying specific self-regulation strengths and weaknesses in AI-assisted learning contexts. Rather than assuming uniform support needs, educators can use SRAILS or similar instruments to identify which students demonstrate strong intrinsic motivation but weak metacognitive strategies; which have developed technological adaptation skills but lack self-efficacy; which excel at prompt engineering but struggle with time management.

This diagnostic capability enables personalized support aligned with individual student needs. A student with weak self-efficacy regarding AI use might receive confidence-building exercises and scaffolded success experiences. A student struggling with metacognitive awareness might receive explicit instruction in monitoring strategies. A student weak in time management might receive temporal scaffolding support. This differentiated approach represents more effective support than uniform instruction assuming all students require identical assistance.

Teacher Professional Development

Recognizing prompt engineering as metacognitive scaffolding within SRL requires that teachers develop sophisticated understanding of both prompt engineering and self-regulated learning principles. Teacher professional development should address:

- Understanding how prompt engineering functions as second-order metacognition

- Familiarity with comprehensive self-regulation measurement (SRAILS and similar instruments) enabling diagnostic assessment
- Strategies for scaffolding iterative prompt refinement within authentic learning tasks
- Designing learning activities that integrate prompting scaffolding with SRL phases and address all SRAILS dimensions
- Assessing metacognitive and self-regulatory dimensions of prompt engineering
- Facilitating critical evaluation of AI outputs
- Addressing potential risks of over-reliance on AI scaffolding, ensuring students develop autonomous regulation
- Personalizing support based on diagnostic assessment results

Curriculum Integration

Rather than treating prompt engineering as an isolated topic, it should be integrated across curricula with attention to the full range of self-regulation dimensions. Writing courses naturally integrate prompt engineering for AI-assisted writing, with explicit attention to motivational support and metacognitive scaffolding. Science courses might integrate prompting for seeking explanations of difficult concepts, with emphasis on metacognitive monitoring. Language courses employ prompting for generating practice dialogues, with attention to time and task management. Mathematics courses use prompting for problem-solving strategies, with development of self-efficacy.

This curricular integration positions prompt engineering as a fundamental learning skill integrally connected to broader self-regulation development rather than a peripheral technical competency.

Technology Design Considerations

Educational AI systems should be explicitly designed to support metacognitive prompt engineering and broader self-regulation development. Design features might include:

- Prompt quality feedback: Systems might provide feedback on prompt clarity, specificity, completeness, and strategic appropriateness, supporting metacognitive awareness of prompt quality.
- Refinement suggestions: Rather than merely generating responses, systems might suggest how prompts could be refined to obtain better responses, supporting iterative improvement.
- Visible prompt-response relationships: Making explicit the connections between specific prompt features and resulting response characteristics helps learners develop prompting expertise and technological adaptation awareness.
- Structured prompting templates: Providing templates or frameworks supports initial prompting while scaffolding development toward more autonomous prompting.
- Reflection prompts: Systems might include metacognitive prompts cuing students to reflect on their prompting effectiveness and self-regulation processes.

- Progress tracking and motivation support: Systems might display learning progress and provide motivational feedback supporting intrinsic motivation development.

Institutional Support

Institutions should develop structures supporting teachers in implementing prompt engineering scaffolding and comprehensive self-regulation development:

- Professional learning communities examining how to integrate prompt engineering scaffolding and self-regulation support within disciplinary courses
- Resource repositories sharing exemplary prompting activities, self-regulation assessment approaches, and differentiated support strategies
- Time and support for faculty to redesign courses emphasizing prompt engineering as metacognitive SRL and addressing all self-regulation dimensions
- Research initiatives examining effectiveness of various scaffolding approaches and how different self-regulation dimensions contribute to learning outcomes with AI
- Training in administering and interpreting self-regulation scales like SRAILS to enable diagnostic assessment

Conclusion

This theoretical study has synthesized six empirical investigations—including recent work on measurement of self-regulation in AI-assisted learning contexts—examining prompt engineering within self-regulated learning frameworks. The evidence demonstrates convincingly that prompt engineering is not merely a technical skill for operating AI systems but rather a sophisticated form of metacognitive scaffolding integral to self-regulated learning with AI. Furthermore, comprehensive measurement frameworks like SRAILS enable educators to understand the full range of self-regulation dimensions students must develop in AI-assisted contexts.

The central findings are:

1. Second-Order Metacognition: Prompt engineering functions as second-order metacognition wherein learners use AI output quality to reflect on and refine their own query formulation, goals, and mental representations (Alatoai & Alshahri, 2025).
2. AI Interaction Predicts SRL Outcomes: The frequency and quality of students' interactions with AI—their engagement in iterative prompt refinement—significantly predicts variations in self-regulated learning outcomes (Ng et al., 2024).
3. Explicit Instruction Enhances Development: When prompt engineering is explicitly taught within SRL frameworks, students develop significantly stronger prompting skills, metacognitive awareness, and disciplinary learning outcomes (Mzwri & Turcsányi-Szabo, 2025).
4. Iterative Refinement as Core SRL Process: Careful scaffolding of iterative prompt refinement within

Plan–Iterate–Evaluate cycles supports both AI literacy development and metacognitive sophistication (Anders & Speltz, 2025).

5. Co-Regulation of Learning: Prompt engineering represents a mechanism of co-regulation wherein learner metacognition and AI system guidance are intricately interconnected; effective AI-mediated learning requires learners who can thoughtfully query systems (Tsakeni et al., 2025; Zhang et al., 2025).
6. Multidimensional Self-Regulation in AI Contexts: Self-regulation in AI-assisted learning encompasses multiple distinct dimensions—motivational components, cognitive/metacognitive strategies, time and task management, resource management, and technological adaptation—requiring comprehensive assessment and differentiated support (Yurt, 2025).

Theoretical Contributions

This study contributes to self-regulated learning theory by expanding traditional SRL frameworks to explicitly address how learners self-regulate when engaging with AI systems. It demonstrates that incorporating AI into learning contexts requires not merely adding AI tools but reconceptualizing self-regulated learning processes themselves. The study further contributes to understanding metacognition by highlighting second-order metacognitive processes wherein learners reflect on the quality of their own cognitive outputs using external feedback from AI systems. Additionally, it recognizes the importance of comprehensive measurement frameworks capturing the distinctive self-regulation dimensions emerging in AI-augmented educational contexts.

Practical Implications

For educators and instructional designers, this framework suggests urgent priorities:

1. Explicitly integrate prompt engineering scaffolding into SRL instruction rather than treating it as peripheral
2. Employ comprehensive self-regulation assessment (using instruments like SRAILS) to diagnose student strengths and weaknesses
3. Provide differentiated support based on diagnostic assessment across all self-regulation dimensions
4. Design learning activities where students iteratively formulate, evaluate, and refine prompts within authentic disciplinary tasks
5. Develop assessment approaches evaluating metacognitive reasoning about prompting and broader self-regulation development
6. Provide professional development supporting teachers in recognizing and scaffolding prompt engineering as metacognitive SRL practice
7. Design educational AI systems explicitly supporting metacognitive prompt engineering through feedback and refinement scaffolding

Future Research Directions

Future studies should examine: (1) whether explicit prompt engineering instruction transfers to autonomous prompting in novel contexts; (2) how prompting expertise develops over extended periods and how different self-regulation dimensions contribute to sustained growth; (3) what types of scaffolding most effectively support metacognitive prompt engineering across diverse learner populations, disciplines, and AI tool types; (4) how students' prompting strategies evolve as they develop disciplinary expertise; (5) whether and how developing sophisticated prompt engineering skills transfers to other metacognitive competencies; (6) how comprehensive self-regulation assessment using instruments like SRAILS can guide personalized instructional design; (7) what institutional structures and professional development approaches most effectively support implementation of these approaches.

Final Reflection

The integration of generative AI into education presents both opportunities and risks for learner autonomy and metacognitive development. If AI is deployed without explicit scaffolding of metacognitive prompting and without attention to the full range of self-regulation dimensions students must develop, students may become passive consumers of AI outputs, their metacognitive capacities and self-regulation atrophying through outsourcing of thinking to external systems. Conversely, if AI is deliberately integrated within pedagogies emphasizing metacognitive prompt engineering, scaffolded within comprehensive self-regulation frameworks, and supported by diagnostic assessment enabling personalized instruction, AI becomes a catalyst for developing more sophisticated self-regulated learning and deeper metacognitive awareness. The difference lies not in AI's inherent properties but in how educators deliberately design learning environments supporting learners in using AI as a tool for metacognitive development and self-regulation cultivation rather than as a shortcut bypassing these essential human capacities.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Alatoai, A. A., & Alshahri, A. S. (2025). The development and validation of the AI-supported self-regulated science and mathematics learning scale (AI-SSRSML) among secondary school students in Saudi Arabia. *BMC Psychology*, *13*(1), 1330. https://doi.org/10.1186/s40359-025-03764-z

Anders, A. D., & Speltz, E. D. (2025). Developing generative AI literacies through self-regulated learning: A human-centered approach. *Computers & Education: Artificial Intelligence*, *9*, 100482. https://doi.org/10.1016/j.caeai.2025.100482

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Mzwri, K., & Turcsányi-Szabo, M. (2025). The impact of prompt engineering and a generative AI-driven tool on autonomous learning: A case study. *Education Sciences*, *15*(2), 199. https://doi.org/10.3390/educsci15020199

Ng, D. T. K., Tan, C. W., & Leung, J. K. L. (2024). Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. *British Journal of Educational Technology*, *55*(6), 1328–1353. https://doi.org/10.1111/bjet.13454

Pea, R. D. (1987). Socializing intelligence through computers and homeschooling. In D. N. Perkins, J. Lochhead, & J. C. Bishop (Eds.), *Thinking: The second international conference* (pp. 341–360). Lawrence Erlbaum

Tsakeni, M., Nwafor, S. C., Mosia, M., & Egara, F. O. (2025). Mapping the scaffolding of metacognition and learning by AI tools in STEM classrooms: A bibliometric–systematic review approach (2005–2025). *Journal of Intelligence*, *13*(11), 148. https://doi.org/10.3390/jintelligence13110148

Yurt, E. (2025). The Self-Regulation for AI-Based Learning Scale: Psychometric properties and validation. *International Journal of Current Educational Studies*, *4*(1), 95–118. https://doi.org/10.46328/ijces.176

Zhang, Z., Zhou, S., & Bai, B. (2025). A systematic scoping review of self-regulated learning with AI in language education. *International Journal of Applied Linguistics*, *0*(0), 1–15. https://doi.org/10.1111/ijal.70030

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Handbook of self-regulated learning and performance*. Routledge.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

An Economic Mathematical Curricular Complexity Analysis of Data Science Bachelor's Degree Programs at U.S. Public Universities

Carl Lawrence Aronson [1 *], Susannah Jean Aronson-Johnston [2]

[1] Department of Foundational Sciences, College of Marine Sciences and Maritime Studies, Texas A&M University, Galveston, Texas, USA, 0009-0004-1855-2636

[2] United States Department of the Air Force, Studies and Analysis Agency, USA, 0009-0006-7276-4255
Corresponding author: Carl Lawrence Aronson (icemstoffice@gmail.com)

Article Info

Article History
Received:
23 December 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Mathematical curricular complexity
Economic mathematical curricular complexity
Curricular analytics
Data science
Ranking inversion

Abstract

Following the rapid U.S. job market expansion in computational science, undergraduate data science degree programs have proliferated nationwide and involve complex entanglements of mathematics and computational techniques. Curricular analytics, a graph theory framework for quantifying prerequisite and laboratory corequisite tortuosity, was previously used to define a mathematical curricular complexity metric for data science programs. Because no singular accredited U.S. curricular standard exists for the B.S. Data Science degree, calculated complexity values vary widely among public universities. In this study, we compared the effective economic and academic value-added extent of U.S. public B.S. Data Science programs by normalizing mathematical curricular complexity with university cost of attendance as a share of state median household income, yielding a new variable, economic mathematical curricular complexity. We analyzed 56 B.S. Data Science programs at U.S. public universities in 36 states. This normalization, combining mathematical rigor, cost of attendance, and household income by state, further broadened the observed range of programs in terms of comparative economic and academic value-added extent. Significant inversions were observed between economic mathematical curricular complexity and U.S. News and World Report national public university rankings, contrary to perceived prestige and academic reputation.

Citation: Aronson, C. L., & Aronson-Johnston, S. J. (2026). An Economic Mathematical Curricular Complexity Analysis of Data Science Bachelor's Degree Programs at U.S. Public Universities. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 49–62), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA

OPEN ACCESS

Introduction

Mathematics as a Central Feature Throughout U.S. Undergraduate Curricula

From student challenges observed directly within both the lower division undergraduate mathematics and chemistry lecture classrooms, we hypothesized that the perseverance of a student toward mastering increasingly higher-level mathematical concepts sat not only at the heart of exactly how an undergraduate student solves quantitative problems but also how the student ultimately chooses a major (Aronson *et al.*, 2025) in direct opposition to previous econometric analyses by other research groups (Arcidiacono, 2004). In general, U.S. undergraduate science, technology, engineering, and mathematics (STEM) undergraduate curricula tend to follow quite a winding (tortuous) pathway due to prerequisite courses and laboratory corequisites. Nevertheless, mathematics holds a unique place as a learning requirement throughout virtually all undergraduate majors at U.S. Universities ranging from merely a one or two course core component for humanities, arts, and social sciences (HASS) majors, business (commerce) majors, and health majors to a *multi*-course ancillary mathematics requirement for the most technical STEM majors. By utilizing curricular analytics (Heilemann, 2018), a mathematical framework for quantifying curricular tortuosity due to subject matter entanglements arising from prerequisite courses and laboratory corequisites, we were able to quantify mathematical rigor in the form of a novel metric entitled the mathematical curricular complexity (Aronson *et al.*, 2025). From this novel factor, we were able to analyze undergraduate major curricula across the university landscape, which we observed to be not only statistically correlated but more exactly inversely proportional to (*i.e.* the multiplicative inverse of) their graduation rate. Herein, we assert that an increasingly mathematical computational degree program provides both permanent and substantive academic value-added career erudition (*i.e.* fundamental knowledge base).

Mathematical Curricular Complexity of B.S. Data Science Degree Programs at U.S. Colleges

Following the relatively recent rapid U.S. job market expansion in computational science, undergraduate data science degree programs have utterly burgeoned nationwide. Undergraduate data science degree programs are available to state residents at their own public universities in virtually every U.S. state at the present time. In addition, many private universities throughout the U.S. also offer B.S. Data Science degree programs. B.S. Data Science degrees throughout the U.S. routinely encompass a daunting set of subject matter entanglements involving mathematics and statistics as well as intricate, state-of-the-art computational techniques. However, unlike chemistry (ACS Committee on Professional Training, 2023) no single standard accreditation agency exists currently in terms of U.S. academic credentialing of B.S. Data Science degree program curricula. In addition, U.S. universities tend to use a variety of National Center for Education Statistics (NCES) classification of instructional program (CIP) codes in order to categorize the data science major within a College of Arts and Sciences, College of Engineering, or School of Computing and Information Sciences. Hence, at present the B.S. Data Science degree curriculum is merely a function of the particular U.S. university's faculty decisions as an autonomous academic entity. Analysis of B.S. Data Science programs offers a glimpse into an intriguing, modern, applicable and yet nationally unstandardized STEM major at U.S. universities. A large span of calculated mathematical curricular complexity values ranging from 65 to 194 ($\overline{\chi}$ = 121 ± 33) were observed for the thirty-two (32) B.S. Data Science degree programs studied in both public and private universities across the U.S. (Aronson & Aronson, 2025). No

statistical correlation was observed between the mathematical curricular complexity of B.S. Data Science programs at the thirty-two (32) U.S. universities studied and either world university ranking, U.S. national university ranking, or public versus private U.S. institutions of higher learning.

Hypothesis with Respect to the Economic Mathematical Curricular Complexity

In terms of our previous purely academic curricular analysis (Aronson & Aronson, 2025), we hypothesized herein that the observed gamut in mathematical curricular complexity calculated for B.S. Data Science among degree programs at public universities across the U.S. may, in fact, significantly broaden upon incorporating economic factors. Previous differences in mathematics integration and subject matter entanglements noted among academic programs at U.S. universities may be further accentuated by considering financial facets. More significantly, the economic mathematical curricular complexity may be, in some cases, significantly inverted with respect to the 2026 U.S. national university ranking by U.S. News and World Report (U.S. News and World Report, 2026 A) contrary to the perceived prestige and academic reputation of the college. A resident in virtually any U.S. state can now pursue a B.S. Data Science undergraduate degree at a public university within their own state. However, the B.S. degree program at a particular public university can vary significantly in mathematical rigor and integration in terms of subject matter entanglements as quantified by the mathematical curricular complexity. Furthermore, we query herein whether students in particular U.S. states can acquire a rigorous data science undergraduate degree, which is highly mathematically integrated, at a relative low cost making the degree a truly value-added higher education product. We firmly assert that an increasingly mathematically oriented undergraduate computational degree program provides persisting and substantive value-added erudition to the student in terms of a fundamental knowledge base. Forming opinions based solely upon outward appearances by merely recruiting students to B.S. Data Science programs at prestigious and elite universities may, in fact, be unadvisable (Angelo, 2026). True academic worth and quality may not be immediately visible wherein first impressions based solely on external rankings can be deceptive. Nevertheless, we hypothesize that the economic mathematical curricular complexity calculated herein may, in fact, reveal true value-added extent for both the student and family.

Method

Curricular Analytics Methods to Calculate the Mathematical Curricular Complexity

Curricular analytics is a graph theory framework for quantifying curricular tortuosity, which provides a quantitative measurement of subject matter entanglements for each course ($\square_i$) within a total set of (i) number of curricular courses. The blocking factor (A_i) measures the extent to which the course of interest ($\square_i$) blocks the ability to take other courses within the curriculum wherein the course of interest ($\square_i$) serves as a prerequisite. The delay factor (B_i) represents the length of the longest path within the corresponding curriculum flowchart graph, which contains the course of interest ($\square_i$). STEM curricula routinely contain courses which must be completed in sequential order (*i.e.* prerequisites and laboratory corequisites) wherein the ability to navigate these pathways remains absolutely critical for student success (Wigdahl *et al.*, 2014). Course complexity (C_i) is calculated as the

arithmetic sum ($\sum_i (A_i + B_i)$) of the blocking factor (A_i) and delay factor (B_i) for the course of interest ($\square_i$). The mathematical curricular complexity (a novel metric) is calculated as the arithmetic sum complexity ($\sum_i C_i$) for the curricular subset of all required and elective mathematics courses, wherein purely statistics courses are not counted. In each B.S. Data Science curriculum considered herein, the science electives were set to engineering (calculus based) Physics I (mechanics) and Physics II (electricity and magnetism) courses each with their laboratory corequisite. In addition, electives within the B.S. Data Science major degree curricula were set to upper division mathematics courses wherever applicable. Flow charts were created with curricular MS Excel® code *via* the curricularanalytics.org site.

Calculating the Economic Mathematical Curricular Complexity

Herein, we extend the value-added academically through integrating mathematics into B.S. Data Science degree programs (*i.e.* the mathematical curricular complexity) at U.S. public universities by incorporating economic data features organized by U.S. state. Normalizing the mathematical curricular complexity with the university cost of attendance as a share of median annual household income within that particular state produces a second novel metric entitled the *economic mathematical curricular complexity*. The normalization process for the data science major herein involved mathematical rigor from curricular analytics, the average annual cost of public college attendance including tuition, room, board, and fees by U.S. state as well as the median annual household income by U.S. state. Median household income is the midpoint for the census ranking of state household gross earnings commonly categorized as typical as the median remains somewhat insensitive to statistical outliers. The mathematical curricular complexity was divided by a fractional variable consisting of the cost of attendance as a share of median household income in order to calculate the economic mathematical curricular complexity. The fractional value for the cost of attendance (tuition, room, board, and fees) as a share of median household income by state is shown to three significant figures in Table 1 below for each of the thirty-six states studied herein (McCann, 2025). Both the cost of undergraduate attendance at public universities and the median household income vary significantly by state and regionality across the U.S. (McCann, 2025) as shown in Table 1. The ratio of undergraduate yearly cost of attendance to median household income varied significantly from a high in Pennsylvania (0.725) to a low in Utah (0.277), where $\bar{x} = 0.497 \pm 0.105$ for the thirty-six states studied herein. Herein, we analyzed B.S. Data Science degree program curricula at fifty-six (56) public universities from across the U.S. (fifty-four public national research universities and two public regional universities) in thirty-six different U.S. states as shown by location in Table 1and Figure 1.

Table 1. Cost of Attendance (tuition, room, board, and fees) as a Share of Median Household Income Expressed as a Fraction to Three Significant Figures for Each of the Thirty-Six States Studied herein from McCann, 2025

Overall Rank	U.S. State	Cost of Education as a Share of Median Annual Household Income
1	Pennsylvania	0.725
2	Rhode Island	0.712
3	New York	0.683

Overall Rank	U.S. State	Cost of Education as a Share of Median Annual Household Income
4	Massachusetts	0.622
5	Illinois	0.619
6	Vermont	0.604
7	Connecticut	0.597
8	Louisiana	0.578
9	Oregon	0.578
10	Missouri	0.566
11	Tennessee	0.563
12	Wisconsin	0.547
13	Mississippi	0.543
14	South Carolina	0.519
15	Indiana	0.516
16	California	0.515
17	Nebraska	0.477
18	Michigan	0.476
19	West Virginia	0.471
20	Minnesota	0.460
21	Arizona	0.459
22	Washington	0.457
23	New Jersey	0.457
24	Iowa	0.456
25	Florida	0.452
26	North Carolina	0.445
27	Texas	0.438
28	Virginia	0.416
29	New Mexico	0.411
30	Delaware	0.399
31	Colorado	0.397
32	Maryland	0.376
33	South Dakota	0.371
34	Nevada	0.366
35	North Dakota	0.331
36	Utah	0.277



Figure 1. Cost of Attendance as Share of Income

Results

Curricular Analytics Analysis for Mathematical Curricular Complexity

An example curricular analytics flowchart graph for the B.S. Data Science degree curriculum at Colorado State University is shown in Figure 2 below. The liberal arts core courses have been suppressed. The complexity (C_i) value of each course in Figure 2 is shown within the dark circle. The mathematical curricular complexity, equal to 175 in Figure 2, was calculated from the arithmetic sum complexity ($\sum_i C_i$) for the mathematics courses only.

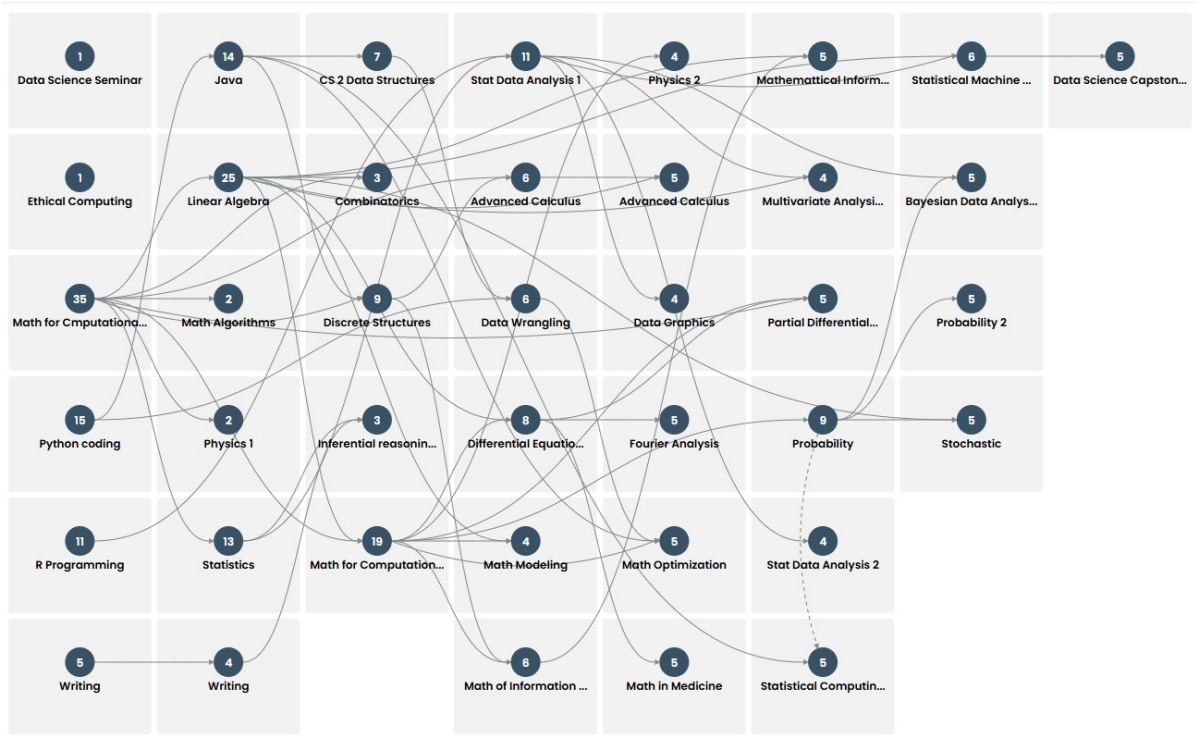


Figure 2. Curricular Analytics Flowchart for Colorado State University

The curricular analytics flowchart for the B.S. Data Science degree program at Florida International University is shown in Figure 3 below. The individual course complexity (C_i) value within the Florida International University

B.S. Data Science degree program in Figure 3 was shown as the number within the dark circle. Mathematical curricular complexity in Figure 3 for the B.S. Data Science degree program at Florida International University (equal to 180) was calculated from the arithmetic sum complexity ($\sum_i C_i$) of the mathematics courses only, which were each denoted with a dashed circle.

In contrast to the degree programs at both Colorado State University and Florida International University, which had relatively high mathematical curricular complexity values of 175 and 180 respectively, the curricular analytics flowchart graph for the B.S. Data Science program at the University of Vermont is shown in Figure 4 below.

The individual course complexity within the University of Vermont B.S. Data Science degree program in Figure 4 was shown as the number within the dark circle. Mathematical curricular complexity in Figure 4 for the B.S. Data Science degree program at University of Vermon, equal to merely 44, was calculated from the arithmetic sum complexity ($\sum_i C_i$) of the mathematics courses only, which were each denoted with a dashed circle. Hence, the University of Vermont B.S. Data Science degree program (Figure 4) had far less mathematical integration and subject matter entanglement compared to the B.S. Data Science degree programs at both Colorado State University (Figure 2) and Florida International University (Figure 3).

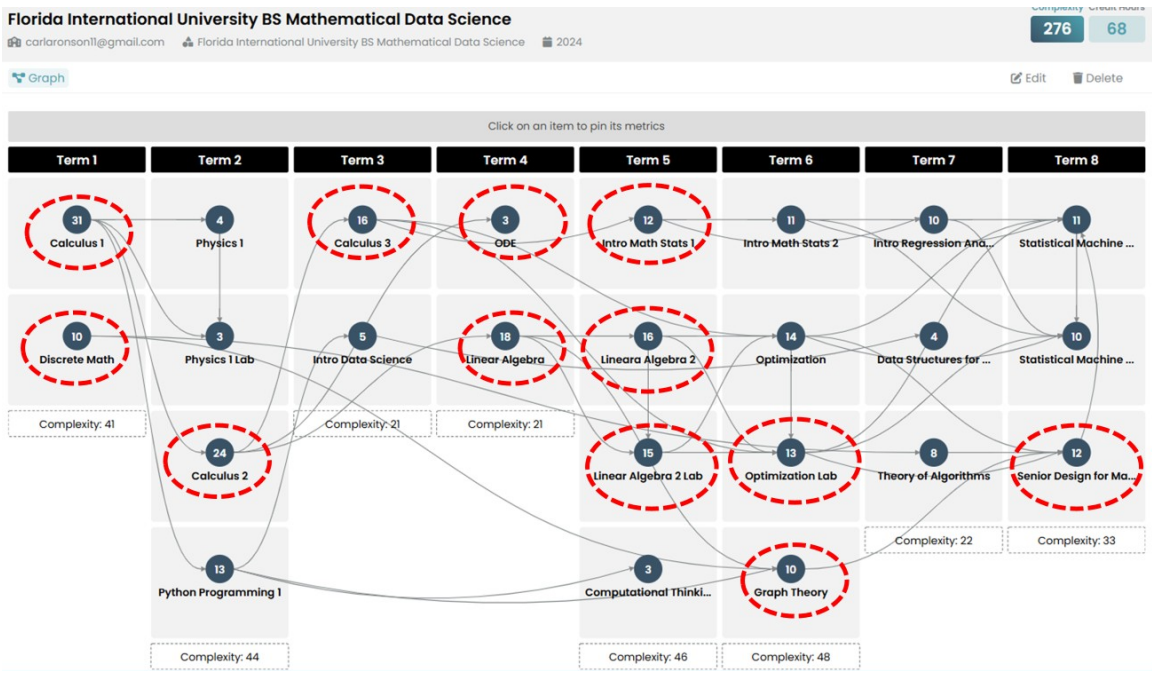


Figure 3. Curricular Analytics Flowchart for Florida International University

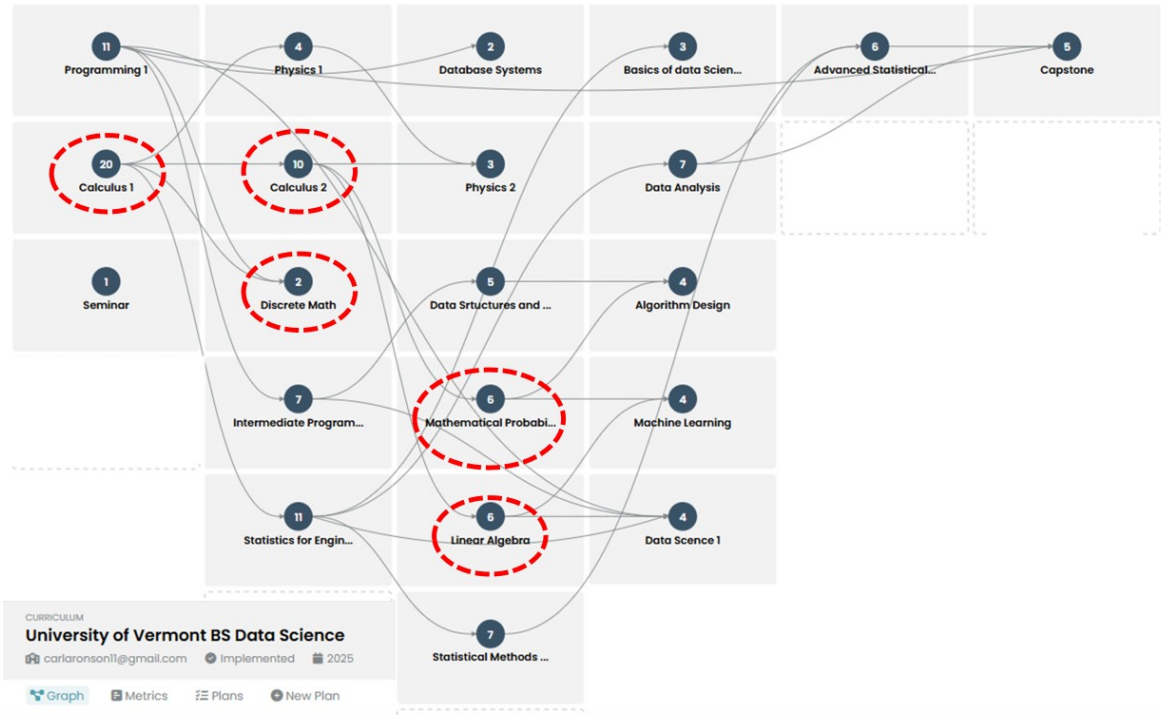


Figure 4. Curricular Analytics Flowchart for University of Vermont

Economic Mathematical Curricular Complexity B.S. Data Science Degree Program Ranking by University

The economic mathematical curricular complexity university ranking is shown in Table 2 below. Beginning from the left-hand side of Table 2 below, column number 1 shows the ranking (#1 - #56) for the mathematical curricular complexity calculated for each university degree program. Column 2 depicts the ranking (#1 - #56) for the economic mathematical curricular complexity calculated for each university degree program.

Table 2. The Economic Mathematical Curricular Complexity Ranking of Fifty-Six U.S. University B.S. Data Science Degree Programs in Thirty-Six Different States

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12	Column 13
RANK	RANK	RANK	RANK	RANK	RANK	RANK	University Name	State	Math Curr.	College $	Econ. Math. Curr.	Relative
MATH	ECONOMICS	DIFFERENCE	USNWR	ENGR.	RELATIVE	CHANGE			Complexity	Income $	Complexity	Change, %
3	1	2	79	77	38	37	Colorado State University	Colorado	175	0.397	441	505
11	2	9	136	161	48	46	University of North Dakota	North Dakota	133	0.331	402	452
1	3	-2	46	125	26	23	Florida International University	Florida	180	0.452	398	447
23	4	19	79	64	37	33	University of Utah	Utah	108	0.277	390	435
2	5	-3	16	21	10	5	University of Washington	Washington	178	0.457	389	435
25	6	19	136	112	47	41	Utah State University	Utah	106	0.277	383	425
6	7	-1	21	14	13	6	Virginia Tech University	Virginia	155	0.416	373	411
5	8	-3	54	77	29	21	University of Texas at Dallas	Texas	156	0.438	356	389
8	9	-1	68	77	35	26	University of Houston	Texas	147	0.438	336	361
4	10	-6	14	42	9	-1	University of California, Santa Barbara	California	160	0.515	311	326
9	11	-2	3	5	2	-9	University of Michigan - Ann Arbor	Michigan	145	0.476	305	318
19	12	7	100	61	42	30	University of Idaho	Idaho	118	0.399	296	306
10	13	-3	57	35	31	18	Arizona State University	Arizona	134	0.459	292	301
7	14	-7	9	32	6	-8	University of California, Davis	California	150	0.515	291	300
14	15	-1	21	17	14	-1	Texas A&M University (B.S. Data Engineering)	Texas	126	0.438	288	295
27	16	11	16	21	11	-5	University of Maryland, College Park	Maryland	101	0.376	269	269
20	17	3	4	50	3	-14	University of North Carolina Chapel Hill	North Carolina	117	0.445	263	261
16	18	-2	84	86	40	22	University of Nebraska – Lincoln	Nebraska	123	0.477	258	254
12	19	-7	9	35	7	-12	University of California, Irvine	California	132	0.515	256	252
13	20	-7			55	35	San Jose State University (Regional U.)	California	132	0.515	256	252
15	21	-6	19	8	12	-9	Purdue University	Indiana	124	0.516	240	230
21	22	-1	40	77	23	1	New Jersey Institute of Technology	New Jersey	109	0.457	239	227
26	23	3	7	11	4	-19	University of Texas at Austin	Texas	102	0.438	233	220
36	24	12	145	176	50	26	South Dakota State University	South Dakota	84	0.371	226	211
17	25	-8	12	14	8	-17	University of Wisconsin – Madison	Wisconsin	121	0.547	221	204
34	26	8	43	50	25	-1	University of Deleware	Deleware	88	0.399	221	203
28	27	1	26	28	16	-11	University of Minnesota Twin Cities	Minnesota	99	0.460	215	195
22	28	-6	36	50	22	-6	Clemson University	South Carolina	109	0.519	210	188
29	29	0	122	135	45	16	West Virginia University	West Virginia	95	0.471	202	177
18	30	-12	42	72	24	-6	University of Illinois Chicago	Illinois	119	0.619	192	164
41	31	10	122	135	46	15	New Mexico State University	New Mexico	78	0.411	190	161
31	32	-1	57	96	33	1	San Diego State University	California	94	0.515	183	151
38	33	5	57	50	32	-1	Iowa State University	Iowa	82	0.456	180	147
30	34	-4	178	176	53	19	University of Tennessee Chattanooga	Tennessee	95	0.563	169	132
44	35	9	7	32	5	-30	University of Florida	Florida	76	0.452	168	131
24	36	-12	26	64	17	-19	Stony Brook University SUNY	New York	108	0.683	158	117
35	37	-2	115	96	44	7	Mississippi State University	Mississippi	85	0.543	157	115
33	38	-5	54		30	-8	University of Oregon	Oregon	89	0.578	154	111
37	39	-2	92	150	41	2	University of Mississippi	Mississippi	83	0.543	153	110
51	40	11	187		54	14	University of Northern Colorado	Colorado	60	0.397	151	107
32	41	-9	29	57	19	-22	University of Massachusetts - Amherst	Massechusetts	93	0.622	150	105
49	42	7	49	72	27	-15	University of Iowa	Iowa	65	0.456	143	96
47	43	4	170	194	52	9	Oakland University	Michigan	67	0.476	141	93
48	44	4	29	48	18	-26	Michigan State University	Michigan	67	0.476	141	93
39	45	-6	170	150	51	6	Louisiana Tech University	Louisiana	81	0.578	140	92
45	46	-1	49	96	28	-18	University of Missouri	Missouri	75	0.566	133	82
42	47	-5	32	69	20	-27	University of Connecticult (B.S. Data Engineering	Connecticut	77	0.597	129	77
50	48	2	1	3	1	-47	University of California Berkeley	California	65	0.515	126	73
52	49	3	57	96	34	-15	George Mason University	Virginia	50	0.416	120	65
40	50	-10	79	135	39	-11	University of Rhode Island	Rhode Island	80	0.712	112	54
43	51	-8	32	57	21	-30	University of Pittsburgh	Pennsylvania	77	0.725	106	46
46	52	-6	26	21	15	-37	Pennsylvania State University	Pennsylvania	71	0.725	98	34
53	53	0	104	77	43	-10	Washington State University	Washington	44	0.457	96	32
55	54	1	136	161	49	-5	Northern Arizona University	Arizona	38	0.459	83	14
56	55	1			56	1	Nevada State University (Regional U.)	Nevada	30	0.366	82	13
54	56	-2	68	135	36	-20	University of Vermont	Vermont	44	0.604	73	0

Column 3 of Table 2 shows the change in ranking from the mathematical curricular complexity to the economic mathematical curricular complexity (column one minus column two). A negative number within column 3 of Table 2 indicated a decrease in ranking upon incorporating economic factors whereas a positive number in column 3 represented a rise in ranking upon considering monetary elements. Column 4 represented the 2026 U.S. News and World Report national public university ranking (U.S. News and World Report, 2026 A). Any ties found within the ranking in column 4 of Table 2 were broken or subdivided by taking into account the data in column 5 representing the 2026 U.S. News and World Report engineering programs ranking for universities offering doctoral level engineering degree programs (U.S. News and World Report, 2026 B). Column 6 represents the overall ranking (#1 through #56) for just the fifty-six U.S. national public universities studied herein from the U.S. News and World Report rankings with respect to both column 4 and column 5. Column 7 of Table 2 represents the change in ranking for the economic mathematical curricular complexity relative to the overall university ranking (column 6 minus column 2) from U.S. News and World Report. A negative value within column 7 represented a fall in ranking relative to the 2026 U.S. News and World Report national public university ranking whereas a positive value in column 7 represented a rise in ranking relative to the U.S. News and World Report ranking for the university. Column 8 of Table 2 lists the name of the university and column 9 represents

the university's state geographic location as depicted in Figure 1. Column 10 of Table 2 lists the mathematical curricular complexity value calculated for each university's B.S. Data Science degree program from curricular analytics modeling following the flowcharts shown in Figure 2, Figure 3, and Figure 4. Column 11 shows the annual cost of college attendance (tuition, fees, room, and board) as a share of median household income in that particular U.S. state listed as a fraction with three significant figures from Table 1. Column 12 of Table 2 shows the economic mathematical curricular complexity calculated for each university's B.S. Data Science degree program from curricular analytics (i.e. column 10 divided by column 11). Column 13 of Table 2 represents the percentage (%) change for the economic mathematical curricular complexity value ($EMCC_j$) from the university of interest from column 12 relative to the lowest calculated economic mathematical curricular complexity value, which was for the University of Vermont ($EMCC_o$) herein, among the fifty-six universities studied as [($EMCC_j - EMCC_o$) / $EMCC_o$) x 100%].

Discussion

Ranking Changes Observed Upon Calculating the Economic Mathematical Curricular Complexity from the Mathematical Curricular Complexity

The mathematical curricular complexity among the fifty-six (56) U.S. public university programs varied from 30 to 180, (i.e. 500% change), wherein the average ($\overline{\chi}$) ± the standard deviation (σ) was $\overline{\chi}$ = 102 ± 36 within column 10 of Table 2. The economic mathematical curricular complexity was observed to vary from 73 to 441 (*i.e.* approximately a 500% change as shown in column 13), wherein the average ($\overline{\chi}$) ± the standard deviation (σ) was $\overline{\chi}$ = 219 ± 96 within column 12 of Table 2. Hence, the percent (%) change from the lowest value to the highest value was virtually unchanged when incorporating the economic aspects in addition to the academic factors. However, a significant inversion of ranking was observed in many cases when comparing the economic mathematical curricular complexity (columns 2 and 12 of Table 2) to merely the mathematical curricular complexity (columns 1 and 10 in Table 2). A movement up the rankings by as many as nineteen places out of fifty-six and a movement down the rankings by as many as twelve places out of fifty-six were observed for column 3 of Table 2. Relative changes were consistently observed in column 3 of Table 2 upon moving from the mathematical curricular complexity to the economic mathematical curricular complexity ranging from (+19) to (-12), wherein on average ($\overline{\chi}$) the ranking movement (column 3 of Table 2) was equal to ±5 places ($|\overline{\chi}|$ = 5) upon moving from purely academic considerations in column 1 values towards taking economic factors into consideration within column 2 of Table 2.

Ranking Changes Observed in the Economic Mathematical Curricular Complexity Relative to the U.S. News and World Report National Public University Rankings

Moreover, considerably larger changes were consistently observed for the economic mathematical curricular complexity relative to the 2026 U.S. News and World Report (USNWR) ranking of public national universities in column 7 of Table 2 (U.S. News and World Report, 2026 A). A movement up the rankings was observed by as many as forty-six places out of fifty-six total universities. A movement down the rankings was observed by as

many as forty-seven places out of fifty-six universities studied. The economic mathematical curricular complexity average ($\overline{x}$) ranking change was an astonishing ±16 places ($|\overline{x}| = 16$) out of fifty-six universities studied herein. Inversion of economic mathematical curricular complexity ranking relative to USNWR national public university ranking, both large and small, was observed consistently throughout all fifty-six (56) B.S. Data Science university degree programs studied herein (column 7 of Table 2). The combination of requisite mathematical subject matter entanglements for the university undergraduate student and economic value-added for the student's family ofttimes produced an economic mathematical curricular complexity value far exceeding the prestige and public perception of the particular college ($EMCC_j$) relative to their peer U.S. institutions of higher learning. Thus, regarding B.S. Data Science programs within the U.S. at present, you most definitely "cannot judge a book by its cover" wherein curricula from household university names such as the University of California, Berkeley and The University of Michigan-Ann Arbor were ranked far below universities with less popular national public identities such as the University of North Dakota and Colorado State University within our study as shown in columns 2 and 7 of Table 2.

Conclusion

A novel curricular analytics metric was calculated herein from a combination of both academic and economic factors for fifty-six U.S. public university B.S. Data Science degree programs in thirty-six separate states. A resident in virtually any U.S. state can now get a B.S. Data Science degree at a public university within their own state with varying degrees of mathematical rigor and subject matter entanglement as shown in columns 1 and 10 of Table 2. Furthermore, students in selected U.S. states can get a B.S. Data Science degree that is not only academically sound as bolstered by significant mathematical integration but also economical as a value-added educational graduation accomplishment (column 12 of Table 2) due to the relatively low cost of education within their state compared to the state median annual household income as shown in both Table 1 and column 11 of Table 2. Relatively large changes to the original mathematical curricular complexity were observed in merely incorporating the economic ratio of educational cost of attendance to the median family income due to significant differences by U.S. state and region for this ratio (Table 1). Among the fifty-six (56) U.S. public university programs in thirty-six (36) separate states studied herein, very large inversions were observed for the economic mathematical curricular complexity relative to the 2026 U.S. News and World Report (USNWR) ranking of national public universities and their engineering programs (U.S. News and World Report, 2026 A; U.S. News and World Report, 2026 B). Hence, the takeaway herein is the significant inversion of both prestige and reputation with respect to the economic mathematical curricular complexity ofttimes observed for B.S. Data Science programs across the U.S. Thus, when it comes to U.S. national public university B.S. Data Science programs at the present time surface appearances can be highly misleading wherein true value is revealed only upon a significantly closer quantitative examination *via* the economic mathematical curricular complexity (column 12 of Table 2). Herein, we warn against forming opinions based solely on outward appearances in the form of national ranking of prestige. The true academic (mathematical) quality and economic worth of a B.S. Data Science degree program may not be immediately visible, wherein first impressions can, in fact, be truly deceptive. Students moving blindly toward B.S. Data Science programs based solely on the prestige and elite nature of the university (Angelo, 2026) may not be advisable as mathematical subject matter entanglement provides substantive and

persisting value.

Recommendations

We have previously reported that the graduation rate within a particular undergraduate major correlates statistically with the mathematical curricular complexity across the university landscape (Aronson *et al.*, 2025). More precisely, the mathematical curricular complexity is inversely proportional to (*i.e.* the multiplicative inverse of) the number of graduates recorded annually for colleges across the U.S. for that particular major (Aronson *et al.*, 2025). Hence, the more mathematically integrated the major is, the less students will advance through the curricula and ultimately graduate in that particular undergraduate major. A relatively wide range of mathematical curricular complexity values were observed specifically for B.S. Data Science programs across the U.S. (Aronson & Aronson, 2025). Herein, we report that a particularly low ratio for the cost of university attendance to the median state household income can, when used to normalize, greatly enhance the mathematical curricular complexity calculated solely from academic rigor and extent of mathematical subject matter entanglement (column 12 of Table 2). Hence, no matter how impressive a B.S. Data Science major curriculum appears on paper, in order to satisfy the U.S. "business of education" (*i.e.* the complex marketplace balancing profitability with instruction and innovation) students at the college must actually elect, progress through, and ultimately finish the major upon graduation. Hence, designing an intricate, highly mathematical data science major for student cohorts who are not particularly mathematically inclined will serve as an effective mismatch and result in a low student yield for the data science major at the particular university of interest. Future research will focus on optimization between mathematical subject matter entanglements and the quantitative aptitude of the student body for B.S. Data Science programs across the U.S.

Acknowledgements

The authors wish to thank the Department of Foundational Sciences within the College of Marine Sciences and Maritime Studies at the Texas A&M University Galveston campus for their generous financial support of this research and its remote presentation at the 2026 International Conference on Education in Mathematics, Science, and Technology (ICEMST) of the International Society for Technology, Education, and Science (ISTES) organization in Cappadocia (Nevsehir), Turkiye.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

ACS Committee on Professional Training (2023). *Coursework guidelines*. In ACS Guidelines for Bachelor's Degree Programs, Washington, DC: American Chemical Society. Retrieved from https://www.acs.org/education/policies/acs-approval-program/guidelines.html https://www.acs.org/education/policies/acs-approval-program/guidelines/coursework.html

Angelo, J. (May 9, 2026) 'Employers are increasingly turning to degree and GPA' in hiring: Recruiters retreat from 'talent is everywhere,' double down on top colleges, *Fortune*, New York, NY: Fortune Media Group Holdings (Chatchaval Jiaravanon).

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *J. Econom*, *121*(1–2), 343–375.

Aronson, C. L., Ravi, R., Black, K. R., & Aronson, S. J. (2025). The Effect of Mathematical Curricular

Complexity on College Major Stratification. In M. Shelley & O. T. Ozturk (Eds.), *Proceedings of ICEMST 2025-- International Conference on Education in Mathematics, Science and Technology, Antalya, Turkiye.* International Society for Technology, Education, and Science (ISTES) Organization, pp. 34 - 53.

Aronson, C. L., & Aronson, S. J. (2025). A Mathematical Curricular Complexity Analysis of U.S. Data Science and Data Engineering Bachelor's Degree Programs. In M. Shelley & O. T. Ozturk (Eds.), *Proceedings of ICONEST 2025-- International Conference on Engineering, Science and Technology, Los Angeles, California, USA* (pp. 1–20). International Society for Technology, Education, and Science (ISTES) Organization.

Heileman, G.L., Abdallah, C.T., Slim, A., & Hickman, M. (2018). *Curricular Analytics: A Framework for Quantifying the Impact of Curricular Reforms and Pedagogical Innovations*. arXiv preprint, arXiv:1811.09676. Retrieved from https://arxiv.org/pdf/1811.09676

McCann, A. (November 13, 2025). Where College Education Costs the Most & Least. *WalletHub*. Retrieved from https://wallethub.com/edu/states-where-college-education-costs-the-most-least/145825

U.S. News and World Report (2026 A). *2026 Best National University Rankings*. Washington, DC: U.S. News and World Report. Retrieved from https://www.usnews.com/best-colleges/rankings/national-universities?myCollege=national-universities&_sort=myCollege&_sortDirection=asc.

U.S. News and World Report (2026 B). Best Undergraduate Engineering Programs Rankings (Doctorate). Washington, DC: U.S. News and World Report. Retrieved from https://www.usnews.com/best-colleges/rankings/engineering-doctorate?myCollege=engineering-doctorate&_sort=myCollege&_sortDirection=asc

Wigdahl, J., Heileman, G. L., Slim, A, & Abdallah, C. T., (June 15-18, 2014). Curricular Efficiency: What Role Does It Play in Student Success? *Proceedings of the 2014 American Society for Engineering Education*, Indianapolis, Indiana, Paper ID# 9609.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Critical Thinking in the Age of Generative AI: A Neo-Piagetian Cognitive Development Framework

Eyüp Yurt

Bursa Uludağ University, Turkiye, 0000-0003-4732-6879
Corresponding author: Eyüp Yurt (icemstoffice@gmail.com)

Article Info

Article History
Received:
15 January 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Critical thinking
Generative AI
Neo-Piagetian theory
Pedagogical design
Higher-order reasoning

Abstract

The integration of generative artificial intelligence into education raises fundamental questions about how students develop critical thinking when AI systems rapidly generate plausible responses to complex questions. This theoretical study argues that understanding AI's role in critical thinking development requires a neo-Piagetian framework acknowledging both cognitive demands and opportunities AI creates. Rather than treating critical thinking as a static skill, the paper positions it as progressive cognitive transformation wherein learners' capacity to analyze, evaluate, and reflect evolves through appropriately challenging cognitive tasks. Drawing on empirical investigations and foundational work in neo-Piagetian cognitive development theory, critical thinking pedagogy, and cognitive mediators, the paper demonstrates that generative AI, when deliberately integrated within pedagogically-grounded frameworks, functions as a "cognitive mediator" enhancing critical thinking by presenting novel problem-solving scenarios, generating personalized cognitive challenges, and scaffolding progression from concrete to abstract reasoning. However, unguided AI use precipitates cognitive offloading wherein students outsource critical thinking, becoming passive consumers without engaging in analytic, evaluative, and reflective processes essential to cognitive development. The paper proposes a neo-Piagetian framework for AI-mediated critical thinking grounded in Piaget's cognitive conflict, assimilation, and accommodation theories, integrated with contemporary scaffolding and cognitive mediator research. The study concludes that critical thinking in the AI age requires cultivating cognitive sophistication through guided engagement with appropriately calibrated intellectual challenges presented through, with, and about AI systems.

Citation: Yurt, E. (2026). Critical Thinking in the Age of Generative AI: A Neo-Piagetian Cognitive Development Framework. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 63–84), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

OPEN ACCESS

Introduction

The emergence of generative artificial intelligence systems capable of generating coherent, plausible responses to virtually any question presents an unprecedented pedagogical challenge grounded in fundamental questions about cognitive development. How can educators cultivate students' capacity for rigorous critical thinking when AI systems can rapidly produce convincing answers to complex questions without requiring users to engage in deep analytic work?

This challenge is not merely practical but conceptual, requiring theoretical frameworks capable of explaining how critical thinking develops and why cognitive engagement matters for development. The foundational theories of Jean Piaget (1954, 1972), developed through decades of careful observation of children's cognitive development, offer surprisingly prescient insights into this contemporary challenge. Piaget's theory posits that cognitive development occurs not through passive reception of information but through active engagement with intellectual challenges that force learners to reconcile new information with existing mental structures—a process he termed "accommodation" (Piaget, 1954, p. 385).

Contemporary neo-Piagetian theorists, including Case (1985), Pascual-Leone (1970), and Demetriou (2002), have extended and refined Piaget's foundational insights, recognizing that cognitive development is more gradual and context-specific than Piaget's stage theory suggested, but maintaining his central principle: meaningful cognitive development occurs through engagement with cognitive challenges that create what Piaget termed "cognitive conflict"—situations where existing mental models prove inadequate for understanding or solving problems (Chapman, 1988; Piaget, 1954, pp. 386–388).

Simultaneously, recent empirical research on generative AI in educational contexts reveals a troubling paradox. Essel et al. (2024) report that when ChatGPT was integrated into a university research methods course with deliberate pedagogical scaffolding, students demonstrated significantly higher gains in critical thinking, creative thinking, and reflective thinking compared to control groups using traditional resources. Yet Ortega-Ochoa and colleagues (2024) conducted a systematic mini-review of 64 empirical studies (2019–2024) on generative AI in educational contexts and found that "in practice, current AI tools often fall short of fostering critical thinking: students tend to rely on model outputs without questioning them" (Ortega-Ochoa et al.,, 2024, p. 5).

This apparent paradox is resolved when understood through neo-Piagetian frameworks: generative AI itself is pedagogically neutral. The technology's effects on critical thinking development depend entirely on pedagogical design. When AI is deliberately integrated within frameworks emphasizing guided analysis and critical evaluation, it can enhance critical thinking development by creating opportunities for meaningful cognitive engagement. When AI is deployed without pedagogical scaffolding requiring critical analysis, it facilitates cognitive offloading —outsourcing of thinking to external systems—such that critical thinking capacity actually diminishes (Ortega-Ochoa et al.,, 2024).

This theoretical study synthesizes contemporary empirical research with foundational developmental theory to

articulate a neo-Piagetian framework for understanding AI's role in critical thinking development. The framework draws on Piaget's foundational insights regarding cognitive conflict and accommodation, contemporary neo-Piagetian extensions regarding scaffolding and cognitive mediators, and recent empirical investigations of AI-mediated learning to propose pedagogical principles for leveraging AI to support critical thinking development rather than undermining it.

Theoretical Foundations: Cognitive Development Theory and Critical Thinking

Piaget's Theory of Cognitive Development: Schema, Assimilation, and Accommodation

Jean Piaget's theory of cognitive development, articulated across numerous works spanning decades (Piaget, 1954, 1972, 1975), proposes that cognitive development involves progressive construction of increasingly sophisticated mental structures through active engagement with the environment. The central construct in Piaget's theory is the schema—organized patterns of thought and action representing understanding of how the world operates. Piaget (1954) defines schemas as "the mental or behavioral structures by which individuals intellectually adapt to their environment" (p. 385).

Schemas develop progressively through two complementary mechanisms: assimilation and accommodation. Piaget (1954) explains: "Assimilation is the integration of external elements into evolved structures of an organism or its behavior. Accommodation consists of the modification of assimilatory schemes by the demands of external situations" (p. 386). In other words, when learners encounter new information compatible with existing schemas, they assimilate it—incorporating new information into current understanding without requiring schema modification. When they encounter information incompatible with existing schemas, they must accommodate—modifying or expanding schemas to incorporate new understanding.

Critical to Piaget's theory is his recognition that accommodation (cognitive development) is not automatic. Rather, it occurs in response to what Piaget terms "cognitive conflict"—situations where existing schemas prove inadequate for understanding or solving problems. Piaget (1954, p. 388) explains that cognitive conflict creates "disequilibrium," a state of cognitive tension that motivates cognitive work to resolve the inconsistency. When cognitive conflict is resolved through schema accommodation, genuine cognitive development occurs.

Chapman (1988), in his comprehensive analysis of Piaget's theory, emphasizes that "accommodation is the process by which schemas are modified to handle situations to which they cannot presently apply" and that "disequilibrium between assimilatory and accommodatory tendencies drives development forward" (Chapman, 1988, pp. 48–49). This principle—that cognitive development is driven by cognitive conflict requiring schema modification—remains central to understanding learning with AI.

Formal Operational Thinking and Critical Reasoning

Piaget (1972) identified four major developmental stages, culminating in the formal operational stage (approximately age 12 onward, though Piaget recognized substantial individual variation). In this stage,

adolescents and adults develop sophisticated capacity for abstract reasoning, hypothetical thinking, systematic evaluation of multiple possibilities, and metalevel reflection on their own thinking.

Piaget (1972, pp. 75–98) describes formal operational thought as involving capacity for propositional logic, hypothetical-deductive reasoning, combinatorial analysis, and the ability to reason about abstract propositions independent of concrete referents. Rather than reasoning only about concrete objects and events, formal operational thinkers can reason about abstract concepts, possible worlds, logical propositions, and the structure of arguments themselves.

Critical thinking, in its sophisticated forms, represents formal operational reasoning applied to intellectual and ethical domains. When individuals think critically, they engage in the cognitive operations characteristic of formal operational thought: They analyze abstract arguments for logical validity; they consider multiple perspectives and evaluate competing viewpoints; they construct hypothetical scenarios and reason through their implications; they evaluate evidence systematically and identify logical fallacies; and they reflect on their own thinking, examining underlying assumptions.

Brookfield and Preskill (2005), in their foundational work on critical thinking in educational contexts, argue that critical thinking involves "identifying and challenging assumptions that underlie our beliefs and behaviors" and "imagining and exploring alternatives" (p. 5). These processes directly mirror formal operational reasoning: examining abstract logical structures (identifying and challenging assumptions about how arguments work) and hypothetical reasoning (imagining and exploring alternatives).

Neo-Piagetian Extensions: Gradual Development, Context-Specificity, and Scaffolding

Contemporary neo-Piagetian theorists have extended and refined Piaget's foundational work in important ways while maintaining his central insight: cognitive development involves progressive construction of more sophisticated cognitive structures through engagement with appropriately challenging cognitive tasks.

Pascual-Leone (1970), in seminal work extending Piaget's theory, introduces the concept of "mental capacity" or "M-capacity," arguing that cognitive development involves progressive increases in the amount of information individuals can simultaneously process and coordinate. Critically, Pascual-Leone (1970, p. 297) notes that "the development of mental capacity determines which cognitive operations become executable at each developmental stage." This neo-Piagetian insight suggests that cognitive development is gradual rather than stage-based, with individuals progressively developing capacity to handle increasing complexity.

Case (1985), another influential neo-Piagetian theorist, extends Piaget's framework through detailed analysis of how children's cognitive strategies become increasingly efficient and sophisticated.

Case (1985) argues that "cognitive development occurs as children develop more powerful and efficient strategies for processing information within particular domains" (p. 164) and that "the development of domain-specific

expertise involves not just acquisition of knowledge but progressive reorganization of that knowledge into more efficient, higher-order structures" (p. 165). Case's work is particularly important for understanding that cognitive development is not uniform across domains but rather develops progressively within specific fields of knowledge and practice.

Demetriou (2002), synthesizing neo-Piagetian work, argues that "the core of cognitive development involves the progressive coordination of multiple cognitive systems" and that "cognitive development proceeds through cycles of increasing integration and differentiation within and across systems" (p. 74). This theoretical perspective emphasizes that cognitive sophistication emerges through progressive coordination of different thinking processes rather than through movement through discrete stages.

Critically, contemporary neo-Piagetian theorists emphasize the essential role of scaffolding—temporary support structures enabling learners to operate at higher cognitive levels than they could achieve independently. Wood et al. (1976) introduce the metaphor of scaffolding, arguing that effective instruction involves gradually withdrawing support as learners develop competence. Wood et al. (1976, p. 90) explain: "The tutor should control those elements of the task that are beyond the learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence."

This scaffolding concept connects to Vygotsky (1978) foundational work on the Zone of Proximal Development (ZPD)—the region between what learners can accomplish independently and what they can accomplish with expert guidance. Vygotsky (1978, p. 86) defines the ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance." Contemporary research emphasizes that development occurs optimally within this zone: tasks too simple generate no development; tasks too difficult generate failure; tasks within the ZPD, appropriately scaffolded, generate cognitive development (Rogoff, 1990; Wertsch, 1991).

Cognitive Mediators and the Externalization of Cognition

A crucial extension of Piagetian theory comes from Vygotsky (1978) concept of psychological tools—cultural artifacts and symbols that mediate human thought. Vygotsky (1978, pp. 54–57) argues that human cognition is fundamentally mediated through psychological tools including language, notation systems, and technologies. Critically, these tools are not neutral instruments; rather, they reshape cognition itself. Language does not merely express thought; it structures and transforms thinking. Writing does not merely record ideas; it enables new forms of thought impossible without written externalization.

Norman (1993), in influential work on cognitive artifacts, extends this insight, arguing that "cognitive artifacts are tools that enhance, amplify, and extend human cognitive abilities" (p. 18). Norman emphasizes that artifacts shape not only what we can accomplish but fundamentally how we think. A calculator extends mathematical reasoning capability but also changes how we think about mathematics. A scientific instrument extends perceptual capacity but transforms scientific observation.

Crucially, Norman (1993, p. 23) identifies a critical risk: "The same artifacts that extend human cognitive capacity can facilitate cognitive offloading—the outsourcing of cognitive work to external systems—such that internal cognitive capacity atrophies." This observation is particularly relevant for understanding generative AI. Salomon and Globerson (1987), in research on technology and cognition, distinguish between cognitive residue (capacities developed through interaction with tools that internalize into enhanced cognitive capacity) and cognitive offloading (outsourcing of cognitive work such that internal capacity diminishes). They argue (Salomon & Globerson, 1987, pp. 623–625) that whether a cognitive artifact enhances or attenuates capacity depends on how it is used: tools used to extend thinking promote cognitive residue and development; tools used to bypass thinking facilitate cognitive offloading.

This theoretical distinction proves crucial for understanding generative AI's effects on critical thinking. The same AI system can enhance critical thinking development when used to extend and challenge students' thinking, or it can facilitate critical thinking atrophy when used to bypass thinking.

Critical Thinking as a Developmental Achievement

Critical thinking scholarship, grounded in educational research and philosophy, recognizes critical thinking as involving sophisticated cognitive capacities that develop progressively rather than appearing fully formed. Ennis (1962), in foundational work defining critical thinking, proposes that critical thinking is "the correct assessing of statements" and involves "ability to determine whether a statement follows necessarily from premises" and "ability to distinguish between premises and conclusions" (p. 84).

Brookfield and Preskill (2005) argue that critical thinking involves six key capacities: (1) identifying and challenging assumptions, (2) recognizing the importance of context, (3) imagining and exploring alternatives, (4) recognizing implications and consequences, (5) engaging in perspective-taking, and (6) integrating all these into reflective skepticism (Brookfield & Preskill, 2005, pp. 7–12). These capacities directly reflect formal operational reasoning: they involve abstract analysis, consideration of multiple possibilities, and metacognitive reflection.

Importantly, Paul and Elder (2008), in their work on critical thinking standards, argue that critical thinking involves both intellectual standards (clarity, accuracy, precision, consistency, relevance, depth, breadth) and intellectual virtues (intellectual humility, autonomy, empathy, confidence in reason). This framework suggests that critical thinking is not merely a technical competency but involves epistemological commitments and intellectual dispositions that develop progressively through educational experience.

Empirical Evidence: Critical Thinking Development with Generative AI

Guided AI Use and Enhanced Critical Thinking Development

Essel et al. (2024) conducted a quasi-randomized experimental study examining the effects of ChatGPT integration into undergraduate university research methods instruction. Their study, involving 240 Ghanaian

undergraduates, compared learning outcomes between students using ChatGPT with pedagogical scaffolding for in-class research tasks versus control groups using traditional instructional resources.

The results demonstrated statistically significant differences between conditions. Students in the ChatGPT-integrated condition showed significantly higher gains in critical thinking (measured through pre-post assessment of critical thinking competencies) compared to control groups (Essel et al., 2024, p. 7). Beyond critical thinking, students also demonstrated higher gains in creative thinking and reflective thinking—capacities closely aligned with formal operational reasoning.

Importantly, Essel and colleagues (2024) emphasize that these gains occurred not through passive consumption of AI-generated responses but through deliberate pedagogical design requiring active cognitive engagement. The instructional design required students to generate research questions, receive AI-generated responses, analyze those responses for quality and appropriateness, evaluate reasoning, and engage in iterative refinement of inquiry. Students were not treated as passive consumers of AI output but as active agents evaluating and improving upon AI contributions.

From a Piagetian perspective, Essel et al. (2024, p. 8) explain that "with proper instructional design, GenAI tools may help learners practice Piagetian-style abstract reasoning by engaging them in novel problem-solving scenarios." Their findings suggest that AI, when used to present novel problems requiring abstract analysis and evaluation, can function as a source of cognitive conflict—situations where existing mental models prove inadequate—thereby driving the accommodation and cognitive development that characterizes progression toward more sophisticated critical thinking.

Unguided AI Use and Critical Thinking Atrophy

In sharp contrast to Essel et al.'s findings, Ortega-Ochoa and colleagues (2024) conducted a comprehensive systematic mini-review examining 64 empirical studies (published 2019–2024) investigating generative AI in educational contexts. Their analysis reveals a concerning pattern: without explicit pedagogical guidance, generative AI frequently fails to foster critical thinking and may actually undermine it.

Ortega-Ochoa et al., (2024, p. 5) report that "students tend to rely on model outputs without questioning them" and that the review "reveals substantial deficiencies in GAI's capacity to promote critical thinking and maintain response accuracy." More fundamentally, they conclude that "without explicit pedagogical strategies, AI may not spur the intended cognitive development" (Ortega-Ochoa et al., 2024, p. 8).

This finding aligns precisely with Piagetian developmental theory. When students encounter AI-generated responses without requirement to analyze, evaluate, or refine them, they experience no cognitive conflict. Their existing schemas—"AI systems provide correct answers; my task is to accept them"—prove adequate for the instructional task. No accommodation is required; no cognitive development occurs. Rather, the opposite may happen: as students outsource thinking to AI, they engage in less cognitive work, leading to atrophy of critical

thinking capacities (consistent with Salomon & Globerson's, 1987, concept of cognitive offloading).

Critically, Ortega-Ochoa et al. (2024, p. 8) do not conclude that AI should be abandoned. Rather, they call for "curricula that intentionally teach students to analyze and verify AI-generated information." This prescription aligns with neo-Piagetian principles: effective pedagogy requires designing tasks creating cognitive conflict—situations where students' existing mental models prove inadequate—thereby forcing accommodation and cognitive development.

AI as Cognitive Mediator: Theoretical Integration

Monib and colleagues (2024) offer a comprehensive theoretical overview synthesizing literature on generative AI in education. They explicitly acknowledge both risks (students offloading thinking onto AI) and opportunities (AI stimulating higher-order thinking when used thoughtfully). Critically, they articulate a theoretical framework positioning AI as a cognitive mediator—a tool that can either extend or attenuate cognitive capacity depending on how it is deployed.

Monib et al. (2024, p. 8) argue that "when AI is used thoughtfully it can stimulate higher-order thought. For example, AI can present tailored, level-appropriate questions or tasks that require analysis and evaluation." This observation maps directly onto neo-Piagetian scaffolding theory: effective cognitive development requires tasks calibrated to learners' current cognitive level—challenging enough to force accommodation but within reach with support.

Monib and colleagues (2024) identify several mechanisms through which AI can function as a cognitive mediator supporting critical thinking development:

Personalized Cognitive Challenge: Generative AI systems, through adaptive capabilities, can generate personalized assignments and problems calibrated to individual students' cognitive levels. Rather than providing uniform instruction, AI can create appropriately challenging tasks for each learner—more concretely scaffolded for some, more abstractly demanding for others.

Probing Student Reasoning: Instructors can leverage AI to pose carefully crafted questions designed to probe students' reasoning processes, forcing explicit articulation and evaluation of their thinking. Questions designed to create productive cognitive conflict—situations where students must reconcile AI responses with their existing understanding—drive accommodation and cognitive development.

Critical Evaluation of AI Outputs: Monib et al. (2024, p. 9) emphasize that "asking students to critique AI outputs can itself cultivate critical thinking." When students must evaluate AI-generated responses—assessing accuracy, identifying biases, recognizing limitations, articulating what is missing—they engage in the evaluative, analytic reasoning characteristic of critical thinking and formal operational thought.

Monib et al. (2024, p. 10) conclude that "generative AI by itself does not automatically teach critical thinking—rather, guided use of AI (with appropriate scaffolding and feedback) can act as a 'cognitive mediator' that enhances reasoning and problem-solving." This theoretical formulation elegantly integrates Piagetian theory (cognitive conflict and accommodation), neo-Piagetian scaffolding theory (providing appropriately calibrated challenges with temporary support), and cognitive mediator theory (tools that can extend cognition): AI functions as a cognitive tool that enhances human cognitive capacity when deliberately designed and deployed to present appropriately challenging cognitive tasks.

Developmental Progression: from Concrete to Abstract Reasoning

Monib et al. (2024) propose that "as learners' cognitive capacity grows, educators should scale AI-mediated tasks from concrete drills toward more abstract, reflective activities." This prescription aligns with neo-Piagetian developmental theory and reflects the observation that cognitive development is gradual and progressively more demanding.

This pedagogical progression can be understood through neo-Piagetian lenses as corresponding to developmental stages, though understood as flexible and domain-specific rather than rigidly stage-determined:

Concrete Operational Stage: Learners benefit from AI-generated concrete examples, step-by-step explanations, and immediate feedback grounded in specific instances. Cognitive demand involves applying concrete operational reasoning to new domains. For example, AI might generate variations of problems where students must recognize patterns and apply learned procedures in new contexts. According to Piaget (1954), concrete operational thinkers reason effectively about concrete objects and events but struggle with abstract propositions.

Transitional Stage: As learners develop foundational understanding, AI-mediated tasks should introduce increasing cognitive demand, bridging concrete and abstract reasoning. Students might compare multiple AI responses to the same question, identifying similarities, differences, inconsistencies, and alternative approaches. This comparative analysis supports transition toward formal operational reasoning by forcing recognition that multiple valid approaches may exist and that evaluation requires criteria extending beyond "correctness" of a single answer.

Formal Operational Stage: Mature learners benefit from abstract, reflective AI engagement. They critique AI reasoning for logical consistency, evaluate unstated assumptions, consider alternative theoretical perspectives, and reason about implications and consequences. Rather than AI merely providing information, AI functions as a thinking partner in sophisticated intellectual dialogue.

Importantly, Ortega-Ochoa et al. (2024, p. 9) note that "from a Neo-Piagetian perspective, younger learners (in concrete stages) need concrete guidance, while older learners (in formal stages) benefit from challenges that force them to reconcile AI responses with their existing knowledge." This observation emphasizes that effective pedagogical design requires responsiveness to developmental level—providing appropriately calibrated cognitive

challenges for learners at different stages of cognitive development.

A Neo-Piagetian Framework for AI-Mediated Critical Thinking Development

Cognitive Conflict as the Engine of Critical Thinking Development

The central principle underlying neo-Piagetian cognitive development theory—that cognitive growth occurs through engagement with cognitive conflict—provides the foundation for understanding how to effectively integrate AI to support critical thinking development. This principle, rooted in Piaget (1954) foundational insight, remains powerful and current.

Cognitive conflict occurs when learners' existing mental models, schemas, or assumptions prove inadequate for understanding phenomena or solving problems. When learners encounter information contradicting their understanding, unexpected results, or logical inconsistencies, they experience cognitive disequilibrium—a state of tension between what they expect and what they observe. Resolving this disequilibrium requires schema accommodation: modification or expansion of mental models to incorporate new understanding. This process of conflict, tension, and resolution characterizes cognitive development (Chapman, 1988; Piaget, 1954, pp. 386–388).

Generative AI, with its capacity to generate responses that may be insightful, unexpected, incorrect, incomplete, or biased, is inherently a potential source of cognitive conflict. An AI-generated response might contradict a student's assumptions, reveal gaps in their reasoning, present alternative perspectives they hadn't considered, contain subtle errors requiring critical detection, or exhibit biases they must recognize and evaluate. When pedagogy is deliberately designed to leverage this potential for cognitive conflict—when students are required to analyze, evaluate, critically engage with, and potentially challenge AI responses—cognitive development toward more sophisticated critical thinking occurs.

Conversely, when pedagogy treats AI as an authoritative source to be passively accepted, cognitive conflict is eliminated. Students' schema "AI systems provide correct answers; I should accept them" proves entirely adequate for the task. No disequilibrium is created; no accommodation is required; no cognitive development occurs. In this scenario, as Salomon and Globerson (1987) theory suggests, students engage in cognitive offloading: they outsource thinking to AI, and through disuse, their critical thinking capacities may actually atrophy.

Designing Productive Cognitive Conflict while Using AI

A critical challenge in AI-mediated critical thinking development is deliberately designing learning experiences that create productive cognitive conflict while leveraging AI as a resource. This requires pedagogical design that avoids reducing cognitive demand while using AI to enhance intellectual challenge.

Designing for Analysis Rather than Acceptance: Traditional instruction often positions AI responses as authoritative answers to be learned. Critical thinking development requires a different pedagogical orientation.

Rather than asking "What does AI say about this problem?" pedagogy should ask "What does AI say about this problem, and what are the strengths and weaknesses of its response?" This shift moves cognitive work from information retrieval (low cognitive demand) to critical analysis (high cognitive demand).

Brookfield and Preskill (2005, pp. 45–62) describe pedagogical strategies for fostering critical thinking including structured controversy, where students defend and critique different positions, and assumption-hunting, where students identify and challenge underlying assumptions. Similar pedagogical strategies can organize AI engagement: rather than accepting AI responses, students engage in structured critique—identifying assumptions, evaluating evidence, recognizing limitations.

Creating Cognitive Conflict Through Deliberately Imperfect AI Responses: One particularly effective strategy involves providing AI responses that are partially correct but contain subtle errors, biases, or incompleteness—without explicitly marking them as flawed. Students must engage in careful analysis to identify and articulate what is problematic. This creates genuine cognitive conflict: their initial inclination to accept a plausible-sounding response must accommodate to recognition of genuine problems with the response.

Case (1985, p. 168) argues that "cognitive development is stimulated when learners encounter problems whose solution requires them to coordinate their existing strategies in new ways." Deliberately flawed or incomplete AI responses create exactly this situation: solving the problem (evaluating the response) requires students to coordinate multiple evaluation strategies, thereby promoting cognitive development.

Requiring Integration with Broader Disciplinary Understanding: Rather than asking students merely "Is this AI response correct?" pedagogy should ask "How does this AI response relate to the conceptual framework we learned? Does it provide a complete answer? What would you add or modify? How does this response compare to alternative approaches?" These questions require students to integrate AI responses with broader disciplinary understanding, creating productive cognitive conflict when AI responses prove incomplete or misaligned with disciplinary knowledge.

Scaling Cognitive Demand Progressively: Effective pedagogical design progressively increases cognitive demand. Initially, students might engage in straightforward evaluation of AI responses—identifying clear errors or omissions. Progressively, evaluation tasks become more sophisticated: identifying subtle biases, recognizing unstated assumptions, evaluating theoretical consistency, considering alternative frameworks. This progressive scaling maintains optimal cognitive challenge (Case, 1985), keeping learners in the zone of proximal development.

Scaffolding Critical Evaluation of AI

While cognitive conflict drives development, scaffolding provides essential support enabling learners to successfully engage with challenging cognitive tasks. Several scaffolding strategies support critical evaluation of AI:

Explicit Evaluation Frameworks: Students should receive explicit frameworks and heuristics for evaluating AI-generated responses. What criteria should be used to assess quality? What evidence should support claims? What assumptions underlie the response? What perspectives might be missing? Such frameworks scaffold the formal operational reasoning required for sophisticated evaluation.

Ennis (1962) identifies standards for critical thinking including consistency, clarity, and relevance. These standards can be translated into concrete evaluation frameworks: "Does the AI response address the actual question asked?" (relevance), "Does it provide clear explanations and define terms?" (clarity), "Are the claims internally consistent with each other?" (consistency). Such frameworks provide cognitive scaffolding supporting students in more sophisticated evaluation.

Guided Questioning: Rather than leaving evaluation entirely to students, instructors can scaffold through carefully designed questions: "What assumptions does this AI response make about the problem? Do you agree with these assumptions? How might someone with different assumptions evaluate this response differently?" Such questions prompt metacognitive reflection and critical analysis.

Collaborative Analysis: Students should engage in peer dialogue analyzing and critiquing AI responses. Discussing with peers why certain AI responses are inadequate, what information they lack, what biases they might reflect, what alternative approaches exist—this collaborative evaluation scaffolds critical thinking development. Rogoff (1990) concept of "guided participation" and Wertsch (1991) emphasis on "dialogic processes" highlight how collaborative interaction supports cognitive development.

Iterative Refinement: Rather than a single evaluation of an AI response, students should engage in iterative cycles where they evaluate an initial response, identify problems, suggest refinements, see refined responses, and evaluate again. This iterative cycle mirrors Piaget (1954) assimilation-accommodation cycle and provides repeated opportunities for cognitive engagement.

Metacognitive Reflection: Students should explicitly reflect on their own evaluation processes: How did you arrive at your assessment? What criteria did you use? What reasoning led you to identify these issues? What alternative interpretations did you consider? This metacognitive reflection supports the formal operational capacity for thinking about one's own thinking (Piaget, 1972).

Developmental Responsiveness: Differentiated Scaffolding

A sophisticated application of neo-Piagetian theory recognizes that not all learners benefit from identical instructional approaches. Demetriou (2002) emphasizes that cognitive development is domain-specific and uneven across domains. A learner might demonstrate formal operational reasoning in mathematics while remaining in concrete operational reasoning in social domains.

This recognition suggests that effective AI-mediated critical thinking instruction must be responsive to learners'

current developmental levels in specific domains. Several implications follow:

Assessing Current Cognitive Levels: Instruction should begin with assessment of students' current cognitive capacities in relevant domains. Can they engage in abstract analysis of disciplinary problems? Can they consider multiple perspectives? Can they evaluate evidence systematically? Such assessment enables diagnostic understanding of where learners currently function.

Providing Appropriately Scaffolded Challenges: For learners operating primarily in concrete operational reasoning, initial AI engagement might focus on concrete analysis: identifying clear errors in AI responses, comparing AI responses to concrete examples, recognizing obvious gaps. For learners demonstrating formal operational reasoning, engagement might focus on abstract analysis: evaluating theoretical consistency, identifying implicit assumptions, considering paradigm-level alternative frameworks.

Progressive Scaffolding Reduction: As learners develop competence, explicit scaffolding should be gradually reduced. Evaluation frameworks that were initially explicit can become internalized. Guided questions can be gradually withdrawn as students independently prompt their own critical analysis. This progressive reduction, consistent with Wood et al. (1976) scaffolding theory, supports development toward independent critical thinking.

Domain and Discipline Specificity: Critical thinking development should occur within disciplinary contexts, not as generic training. Each discipline has distinctive ways of thinking critically, distinctive standards for evidence, distinctive ways of constructing arguments. AI engagement should be grounded in discipline-specific critical thinking practices. A biologist's critical evaluation of an AI response about ecology involves different criteria than a philosopher's critical evaluation of an AI response about ethics (Case, 1985).

Implications for Educational Practice and Institutional Design

Reconceptualizing AI Literacy as Critical Engagement

Traditional conceptualizations of AI literacy—often framed as technical understanding of how AI systems work—prove insufficient from a critical thinking perspective. Paul and Elder (2008) framework suggests that critical thinking involves both intellectual standards (clarity, accuracy, relevance, depth) and intellectual virtues (intellectual humility, intellectual autonomy, intellectual confidence).

Critical AI literacy, from this perspective, involves more than technical understanding. It involves capacity for rigorous critical evaluation of AI: recognizing AI's genuine capabilities and legitimate limitations, identifying when AI responses are reliable versus when they require verification, understanding biases and epistemic limitations that might be embedded in AI systems, and appreciating when human judgment, disciplinary expertise, and ethical reasoning should override or supplement AI recommendations.

This critical literacy develops not through abstract instruction about AI but through engagement with appropriately challenging cognitive tasks requiring analysis and evaluation of AI. Teaching critical AI literacy requires

designing pedagogical experiences where students must evaluate AI critically, often discovering for themselves AI's strengths and limitations through experience.

Professional Development for Educators

Educators designing and facilitating AI-mediated critical thinking development require sophisticated understanding of both cognitive development theory and pedagogical strategies for maintaining and amplifying cognitive demand. This requirement suggests priorities for professional development:

Understanding Neo-Piagetian Cognitive Development Theory: Professional development should provide educators with conceptual understanding of how cognitive development occurs through cognitive conflict and accommodation, how scaffolding supports development, and how cognitive development is progressive and domain-specific rather than stage-based. Educators need to understand cognitive mediation theory and the distinction between cognitive enhancement (tools extending capacity) and cognitive offloading (tools replacing internal capacity).

Pedagogical Design for Cognitive Demand: Educators should develop competence in designing learning activities that create productive cognitive conflict—situations where students' existing understanding proves inadequate and they must engage in deeper analysis. This includes understanding how to identify appropriate cognitive challenges (within students' zone of proximal development), how to scaffold support, and how to progressively reduce scaffolding as competence develops.

Recognizing and Addressing Cognitive Offloading: Educators need to recognize when students are passively accepting AI outputs versus actively analyzing them. This requires developing assessment approaches focused on understanding rather than mere task completion. It also requires developing awareness of pedagogical structures that facilitate offloading and replacing them with structures requiring critical engagement.

Discipline-Specific Critical Thinking: Professional development should be organized within disciplinary communities, recognizing that critical thinking is inherently discipline-specific. Chemistry faculty, history faculty, and business faculty must develop understanding of how critical thinking operates within their disciplines and how AI engagement can either support or undermine domain-specific critical thinking.

Curriculum Design for Critical Thinking Development

Rather than treating critical thinking as a generic skill taught in isolated courses, critical thinking development should be integrated throughout curricula with deliberate progression and discipline-specific emphasis.

Lower-Level Courses: Courses introducing disciplinary perspectives and foundational concepts can use AI to provide concrete examples, variations on problems, and worked solutions that students analyze for correctness, completeness, and implicit assumptions. The emphasis is on developing foundational understanding while

beginning to recognize that problems can be approached in multiple ways and that evaluation requires judgment.

Mid-Level Courses: As students develop foundational understanding, courses should engage students in more sophisticated critical thinking. AI can generate multiple approaches to problems, different perspectives on issues, or synthesized information that students must critically analyze, compare, and evaluate. Students learn to recognize that complex problems often have multiple defensible approaches and that evaluating quality requires disciplinary judgment, not merely calculation or rule application.

Advanced Courses: Upper-level and specialized courses engage students in sophisticated intellectual dialogue about complex, ambiguous problems central to the discipline. AI serves as a thinking partner: generating ideas that students refine, challenging students' assumptions, presenting alternative frameworks. Students practice the kind of rigorous critical thinking practiced in professional and scholarly communities.

Capstone Experiences: Capstone courses, senior theses, and research-based learning provide opportunities for students to apply developed critical thinking capacities to problems of genuine disciplinary importance. AI engagement in these contexts should support rather than replace student thinking.

Assessment of Critical Thinking Development

Assessment should evaluate genuine critical thinking—the capacity to analyze, evaluate, and reflect on complex problems—rather than mere skill performance or knowledge recall. Several assessment approaches align with neo-Piagetian theory:

Analysis of Student Evaluation of AI Outputs: Rather than asking students for answers, ask them to evaluate and critique AI-generated responses. Do they identify legitimate issues? Do they provide evidence for their critiques? Can they articulate what would improve inadequate responses? Do their critiques demonstrate understanding of disciplinary standards? This directly assesses critical thinking capacity.

Brookfield (2012), in comprehensive work on assessing critical thinking in higher education, advocates for "critical incident questionnaires" where students describe instances where they recognized assumptions or perspectives they hadn't previously considered. Similar approaches could be adapted to AI contexts: asking students to describe times when AI responses challenged their thinking, forced them to refine their understanding, or revealed gaps in their reasoning.

Reasoning Justification and Metacognitive Reflection: Rather than asking students for answers, ask them to explain and justify their reasoning. How did they arrive at their conclusions? What evidence do they cite? What alternative perspectives did they consider and reject? What uncertainties remain? Such assessment assesses the depth and sophistication of reasoning.

Problem Decomposition and Strategy Selection: Present complex, ambiguous problems (potentially involving

AI-generated information). Assess students' capacity to break problems into components, identify what information is needed, recognize and articulate assumptions, identify multiple possible approaches, and explain the reasoning for selected approaches. This assesses formal operational reasoning: ability to consider multiple possibilities, coordinate multiple factors, and make strategic choices.

Conceptual Integration Across Contexts: Assess students' capacity to transfer critical thinking strategies across different contexts and problem types. Can they apply evaluation criteria learned in one context to novel problems? Do they recognize when different approaches are appropriate for different problem types?

Challenges, Risks, and Considerations

The Cognitive Offloading Risk and Pedagogical Failure

The most significant challenge in AI-mediated critical thinking development is cognitive offloading: students outsourcing thinking to AI, becoming passive consumers of AI outputs without engaging in analysis or evaluation. Ortega-Ochoa et al., (2024) document this risk empirically: in the absence of explicit pedagogical strategies, students tend to rely on AI outputs without questioning them.

From a neo-Piagetian perspective, cognitive offloading represents a fundamental pedagogical failure: instruction has failed to create cognitive conflict—situations where students' existing approaches prove inadequate—thereby failing to drive cognitive development. When students can accomplish instructional tasks (obtaining answers, completing assignments) by accepting AI outputs without critical engagement, no cognitive conflict exists. No accommodation is required; no cognitive development occurs.

The solution is not to abandon AI but to redesign instruction such that cognitive conflict becomes unavoidable. If students cannot complete learning tasks through passive acceptance of AI outputs, if genuine understanding requires critical evaluation, if assessment measures understanding rather than mere completion, then cognitive offloading becomes pedagogically impossible.

However, this requires fundamental pedagogical redesign. Many current instructional practices inadvertently facilitate offloading by valuing task completion over understanding, by assessing knowledge recall rather than reasoning, by allowing students to hide behind AI-generated answers rather than requiring articulation of their own thinking.

Equity, Access, and Developmental Readiness

Essel et al. (2024) conducted their research with Ghanaian undergraduates, demonstrating that AI-mediated critical thinking development is not limited to privileged populations. However, effective AI-mediated critical thinking requires not merely access to technology but also pedagogical expertise, institutional support, and students' foundational understanding in disciplinary domains.

Several equity Considerations Merit Attention:

Pedagogical Expertise and Institutional Support: Effective AI-mediated critical thinking development requires substantial pedagogical expertise, professional development, and institutional support. Institutions with fewer resources—less wealthy institutions, institutions serving lower-income students—may lack capacity to develop and implement such approaches. This creates risk that AI amplifies educational inequities rather than reducing them.

Foundational Understanding: Critical thinking development requires foundational understanding in disciplinary domains. Students who lack foundational understanding cannot productively evaluate AI responses because they lack disciplinary knowledge enabling discrimination between good and poor responses. This suggests that critical AI literacy development requires prerequisite learning.

Digital Literacy and Technological Comfort: While many learners are technologically sophisticated, others may struggle with technical aspects of AI interaction, limiting their capacity to engage productively with AI. Effective implementation requires ensuring all students develop necessary technological competence.

Developmental Readiness: Not all students are developmentally ready for the same cognitive challenges. Some students may still be primarily concrete operational in their reasoning about particular domains and benefit from different scaffolding than formal operational thinkers. Equitable implementation requires responsive differentiation.

Quality of AI Outputs and Reliability

Critical thinking development requires that AI outputs be substantive enough to merit critical analysis. Trivial or nonsensical AI responses do not provide productive cognitive conflict; they may merely frustrate learners or waste instructional time.

Moreover, effective critical thinking development requires that AI responses be reliable enough that students' evaluation efforts produce accurate judgments. If AI outputs are substantially unreliable—containing frequent errors, biases, or misrepresentations—students may form incorrect conclusions about AI capabilities and limitations through their evaluation efforts.

This creates a tension: AI must be sophisticated enough to warrant serious analysis, but not so unreliable that evaluation becomes futile. Effective practice requires careful selection of AI systems and thoughtful framing of their limitations.

The Epistemological Challenge: What Counts as Knowledge?

An important consideration, touched on but not fully developed, concerns epistemology—what counts as

legitimate knowledge and how it is produced. Traditional critical thinking in academic contexts has involved evaluation of human-authored texts, empirical research, scholarly arguments. These sources bear epistemic authority based on disciplinary credentials, methodological rigor, peer review, and explicit reasoning.

Generative AI systems operate according to fundamentally different epistemological principles. They generate plausible responses based on statistical patterns in training data, not through disciplinary reasoning, empirical research, or explicit argumentation. Their outputs may be coherent and persuasive without being epistemically reliable or defensible.

This creates what we might call an "epistemological friction" between how students should think about human-produced disciplinary knowledge and how they should think about AI-generated responses. Students must learn to evaluate AI critically while simultaneously developing disciplinary understanding based on human-produced knowledge. This requires explicit epistemological instruction: helping students understand different modes of knowledge production and the distinctive standards appropriate for evaluating each.

Conclusion

This theoretical study has synthesized contemporary empirical investigations with foundational cognitive development theory to articulate a neo-Piagetian framework for understanding AI's role in critical thinking development. The framework, grounded in Piaget's foundational theory of cognitive conflict and accommodation, extended through contemporary neo-Piagetian work on scaffolding and cognitive mediators, demonstrates that critical thinking development with AI is not predetermined by technology but rather depends fundamentally on pedagogical design.

The central findings are:

1. **Cognitive Conflict as Development's Engine:** Critical thinking develops through engagement with cognitive conflict—situations where existing mental models prove inadequate—requiring accommodation (modification of schemas). Generative AI, when deliberately designed to create such conflict, can drive development toward more sophisticated critical thinking. Conversely, when AI is used to bypass thinking, cognitive development stagnates or atrophies (Essel et al., 2024; Ortega-Ochoa et al., , 2024; Piaget, 1954).
2. **AI as Cognitive Mediator:** Generative AI functions as a cognitive mediator that can either enhance or attenuate critical thinking capacity depending on pedagogical deployment. When used to extend thinking through challenging analysis and evaluation, AI enhances cognitive capacity and supports development. When used to replace thinking through passive acceptance of outputs, AI facilitates cognitive offloading and capacity atrophy (Monib et al., 2024; Norman, 1993; Salomon & Globerson, 1987).
3. **Pedagogical Design is Determinative:** The difference between AI use that enhances versus undermines critical thinking is not inherent to technology but rather depends on pedagogical design. Pedagogy that creates cognitive conflict through requiring critical analysis, that maintains genuine intellectual

challenge, and that assesses understanding rather than mere task completion supports critical thinking development. Pedagogy that allows passive acceptance of AI outputs facilitates cognitive offloading (Essel et al., 2024; Ortega-Ochoa et al., , 2024).

4. **Scaffolding Supports But Does Not Replace Challenge:** Effective critical thinking development requires scaffolding—temporary support structures enabling learners to engage with challenges beyond their independent capacity. However, scaffolding supports engagement with challenge; it does not reduce cognitive demand. Scaffolding that merely simplifies tasks without maintaining intellectual challenge fails to promote development (Case, 1985; Wood et al., 1976).
5. **Developmental Responsiveness is Essential:** Critical thinking development is progressive and domain-specific. Effective pedagogy is responsive to students' current cognitive development levels in specific domains, provides appropriately calibrated cognitive challenges (within the zone of proximal development), and progressively increases demand as competence develops. Different students require different forms of scaffolding; differentiated instruction, responsive to developmental level, supports equitable critical thinking development (Demetriou, 2002; Pascual-Leone, 1970).

Theoretical Contributions

This study contributes to educational theory by demonstrating the enduring relevance of Piaget's foundational cognitive development theory for understanding contemporary pedagogical challenges. Piaget's insights regarding cognitive conflict and accommodation, developed through careful observation of children's learning, remain theoretically powerful and empirically validated for understanding how critical thinking develops in the age of AI. The study further contributes by articulating the conceptual distinction between cognitive enhancement (cognitive mediators that extend capacity) and cognitive offloading (outsourcing that atrophies capacity), applying this distinction to understand AI's variable effects.

The study also contributes by integrating multiple theoretical traditions—Piagetian cognitive development theory, neo-Piagetian extensions regarding scaffolding and domain-specificity, critical thinking philosophy emphasizing intellectual virtues and standards, and cognitive mediator theory—into a coherent framework applicable to contemporary questions about AI and learning.

Practical Implications

For educators, administrators, and instructional designers, this framework suggests several urgent imperatives:

1. Recognize that critical thinking development requires maintaining or increasing cognitive demand, not reducing it through readily available AI answers. Every pedagogical choice should be evaluated through the question: Does this design maintain or increase genuine intellectual challenge?
2. Design learning experiences requiring students to analyze, evaluate, and refine AI-generated responses rather than passively accepting them. Create tasks where cognitive engagement with AI is unavoidable if students are to succeed.

3. Scaffold critical evaluation of AI through explicit frameworks, guided analysis, and collaborative dialogue, while maintaining sufficient challenge to require genuine intellectual work.
4. Progress instruction from concrete AI engagement toward abstract, reflective AI interaction, responsive to students' developing cognitive capacity.
5. Assess critical thinking through evaluation of students' reasoning, analysis, and explanation rather than through conventional testing or task completion.
6. Provide professional development supporting educators in understanding cognitive development theory and designing appropriately challenging AI-mediated learning experiences.
7. Provide differentiated pedagogical support responsive to students' current cognitive development levels in specific domains, recognizing that not all students benefit from identical instruction.
8. Organize critical thinking development within disciplinary contexts, recognizing that critical thinking is inherently discipline-specific, with distinctive standards, methods, and ways of reasoning.

Future Research Directions

Future studies should examine:

1. What specific pedagogical strategies most effectively create productive cognitive conflict within AI-mediated contexts across diverse disciplines?
2. How do students' critical thinking capacities develop over extended periods of carefully designed AI-mediated learning? What is the longitudinal trajectory of development?
3. How do students' developmental levels in specific domains influence their capacity to benefit from various forms of AI engagement? What assessment approaches effectively identify developmental readiness?
4. Does critical thinking developed with AI transfer to contexts without AI support? How robust is transfer?
5. What forms of cognitive offloading occur in practice, and what pedagogical interventions most effectively address them?
6. What institutional structures, professional development approaches, and policy frameworks most effectively support implementation of critical thinking-focused AI pedagogy?
7. How do different AI systems—with different capabilities, different epistemic limitations, different biases—differentially support critical thinking development?

Final Reflection

Generative AI presents an unprecedented choice for education with profound consequences for students' cognitive development. Pedagogy that treats AI as providing authoritative answers will facilitate cognitive offloading, atrophying students' critical thinking capacity and moving cognition further from the rigorous disciplinary thinking that constitutes genuine expertise. The ease of obtaining plausible-sounding answers from AI makes it tempting to reduce cognitive demand, accepting AI outputs rather than engaging in difficult thinking.

Conversely, pedagogy that treats AI as a thinking partner requiring rigorous critical evaluation—that leverages AI's capacity to generate novel problems, alternative perspectives, and intellectual challenges—can cultivate more sophisticated critical thinking than traditional approaches. The same technology that enables cognitive offloading can, with appropriate pedagogical design, extend intellectual capacity and support development toward more sophisticated reasoning.

The difference lies not in technology but in human pedagogical wisdom: the deliberate design of learning experiences that maintain cognitive demand, create productive cognitive conflict, scaffold engagement with appropriately challenging intellectual work, and assess understanding rather than mere task completion. In Piaget (1954, p. 412) formulation: "Education should aim to develop individuals capable of doing new things, not simply of repeating what previous generations have done." AI-mediated education can support this aim if, and only if, it is designed to maintain and amplify intellectual challenge rather than reducing it.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Brookfield, S. D. (2012). *Teaching for critical thinking: Tools and techniques to help students question their assumptions*. Jossey-Bass.

Brookfield, S. D., & Preskill, S. (2005). *Discussion as a way of teaching: Tools and techniques for democratizing the classroom (2nd ed.)*. Jossey-Bass.

Case, R. (1985). *Intellectual development: Birth to adulthood*. Academic Press.

Chapman, M. (1988). *Constructive evolution: Origins and development of Piaget's thought*. Cambridge University Press.

Demetriou, A. (2002). Tracing psychological structures and processes across the life-span. In A. Demetriou, A. Raftopoulos, & A. Efklides (Eds.), *Metacognition: Psychological and educational perspectives* (pp. 69-101). Routledge.

Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review*, *32*(1), 81–111.

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers & Education: Artificial Intelligence*, *6*, 100198. https://doi.org/10.1016/j.caeai.2023.100198

Monib, W. K., Qazi, A., Apong, R. A., Azizan, M. T., De Silva, L., & Yassin, H. (2024). Generative AI and future education: A review, theoretical validation, and authors' perspective on challenges and solutions. *PeerJ Computer Science*, *10*, e2105. https://doi.org/10.7717/peerj-cs.2105

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Addison-Wesley.

Ortega-Ochoa, E., Sabaté, J.-M., Arguedas, M., Conesa, J., Daradoumis, T., & Caballé, S. (2024). Exploring the utilization and deficiencies of generative artificial intelligence in students' cognitive and emotional needs: A systematic mini-review. *Frontiers in Artificial Intelligence, 7,* Article 1493566. https://doi.org/10.3389/frai.2024.1493566

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica*, *32*, 301–345.

Paul, R. W., & Elder, L. (2008). *The miniature guide to critical thinking: Concepts and tools (5th ed.)*. Foundation for Critical Thinking.

Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). Basic Books.

Piaget, J. (1972). *The principles of genetic epistemology* (W. Mays, Trans.). Basic Books.

Piaget, J. (1975). *The child's conception of the world* (J. Tomlinson & A. Tomlinson, Trans.). Rowman and Littlefield.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford University Press.

Salomon, G., & Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research*, *11*(6), 623–637. https://doi.org/10.1016/0883-0355(87)90006-1

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Harvard University Press.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

A Mathematical Curricular Complexity Analysis of U.S. Primary and Lower Secondary School Mathematics Curricula

Carl Lawrence Aronson [1*], Susannah Jean Aronson-Johnston [2]

[1] Department of Foundational Sciences, College of Marine Sciences and Maritime Studies, Texas A&M University, Galveston, Texas, USA, 0009-0004-1855-2636

[2] United States Department of the Air Force, Studies and Analysis Agency, USA, 0009-0006-7276-4255
Corresponding author: Carl Lawrence Aronson (icemstoffice@gmail.com)

Article Info

Article History
Received:
2 November 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Curricular analytics
Primary school
Lower secondary school
Mathematics curricula
Zearn mathematics curriculum

Abstract

Mathematical learning requires an intricate building process in which prerequisite concepts must be mastered before advanced material. Curricular analytics, using graph theory to quantify prerequisite and corequisite entanglements, provides a way to measure subject matter tortuosity. In this study, curricular analytics techniques were applied to a model U.S. kindergarten through eighth grade (K-8) mathematics curriculum to quantify the breadth and depth of formal mathematical integration through a novel metric, mathematical curricular complexity. To simulate academic stress-strain in student tension versus perseverance, mathematical curricular complexity was calculated as a function of sequentially removing particularly critical and challenging subject matter through modular extraction. The model showed that decreases in academic credits caused disproportionate decreases in mathematical curricular complexity because of prerequisites and the propagation of mathematics subject learning gaps, serving as a proof of concept. Mathematical learning cavities were proposed as proportionate to overall decreases in student mastery and cumulative learning success, pending validation in a longitudinal mathematics classroom study. These findings suggest that subject matter entanglements in K-8 mathematics curricula may help explain functional innumeracy and highlight the importance of mastering foundational modules for later achievement.

Citation: Aronson, C. L., & Aronson-Johnston, S. J. (2026). A Mathematical Curricular Complexity Analysis of U.S. Primary and Lower Secondary School Mathematics Curricula. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 85–99), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA

OPEN ACCESS

Introduction

Mathematics Learning as a Building Process Leading to Helical Tortuosity and Persistence

Mathematical learning requires an intricate and gradual building process involving mastering prerequisite concepts before moving on to subsequent advanced material. Hence, with regard to mathematics we must learn to walk before we can run. Furthermore, periodically reviewing previous concepts conventionally known as helical learning has also been shown to be highly advantageous in both mathematics and science (O'Connor, 2006). A helix is a three-dimensional geometric shape like that of a wire wound uniformly around a cylinder. In helical learning, at the base level merely concrete experiences and sensory engagement are initially emphasized whereas successive turns lead to increasing abstraction, analysis, synthesis, and evaluation. Revisiting topics at higher levels of mathematical sophistication greatly reinforces connections and leads to a significantly deeper mathematical understanding. This helical type of learning progression mirrors a spiral curricular concept, wherein key concepts are revisited over time with greater and greater depth, leading to significantly stronger long□term student retention. Helical learning tortuosity refers to the winding, spiral-like nature of the mathematical curricular flow path. Helical persistence refers to the student's tendency, within the helical curricular structure's curvature, to maintain learning intensity and mastery over time.

Deficiency in Kindergarten Through Eighth Grade Mathematics and U.S. Adult Innumeracy

A keen understanding of arithmetic operations, algebraic functions, and geometric transformations from both primary school (kindergarten through fifth grade) and lower secondary school (sixth grade through eighth grade) mathematics modules is absolutely a key predictor of achieving functional numeracy toward increasingly higher-level mathematical learning achievement. Proficiency throughout the U.S. kindergarten through eighth grade school mathematics sequence has proven to optimally prepare students for both high school (upper secondary school, ninth through twelfth grade) and science, technology, engineering, and mathematics (STEM) college programs, although in practice only a minority fraction of students is actually willing to persevere toward complete mastery. Nevertheless, at present a substantial fraction of U.S. adults is functionally innumerate stemming directly from an apparent inability to master fundamental primary and lower secondary school mathematics modules. The scale of this mathematical education problem within the U.S. is somewhat daunting.

According to the U.S. Department of Education's National Center for Education Statistics (NCES), using the Organization for Economic Cooperation and Development's (OECD) Program for the International Assessment of Adult Competencies (PIAAC) data from 2012–2014, about 70% of U.S. adults aged 16–65 scored at or below the lowest PIAAC numeracy level (level 1), or could not participate due to language or cognitive barriers (National Center for Education Statistics, 2020). Hence, roughly two in three U.S. adults are considered "innumerate" in the OECD sense as merely unable to handle basic mathematical tasks from kindergarten through eighth grade mathematics within their daily life. This data translates most prominently to the U.S. Bureau of Labor and Statistics (BLS), the principal federal agency responsible for collecting, analyzing, and publishing labor market and economic data within the United States, as innumerate U.S. adults simply not possessing the mathematical competencies needed for many modern jobs within our increasingly technical world, such as in the health care,

energy, materials, and information technology economic sectors.

Curricular Analytics Modeling Applied to the Kindergarten Through Eighth Grade Mathematics Learning Sequence

Curricular analytics techniques, a graph theory framework for quantifying curricular tortuosity in the form of prerequisite and corequisite modules, provides a quantitative measurement of subject matter entanglements (Heileman et al., 2018). Using curricular analytics techniques herein on a model U.S. kindergarten through eighth grade (K-8) mathematics curricula, the breadth and depth of integration of formal mathematical concepts was evaluated quantitatively through a novel metric entitled the *mathematical curricular complexity* (Aronson *et al.*, 2025). Herein we apply this modeling to a purely mathematical curriculum of specifically sequenced learning modules instead of an entire collegiate curriculum containing merely a subset of mathematics courses. In order to simulate academic stress-strain in elementary and middle school student tension versus perseverance, the mathematical curricular complexity was calculated as a function of sequentially removing particularly critical and challenging subject matter throughout the K-8 mathematics curriculum. Modular extraction within the prescribed progression was meant herein to mimic a student's partial understanding, *non*-mastery, or limited proficiency within the mathematical area of interest.

Modular Deficiency Propagation Leading to Learning Cavity Areas

Modular learning deficiencies were simulated by merely extracting (removing) the module within the modeling herein. Modular deficiency routinely causes a concomitant decrease in academic credits as well as a disproportionate decrease in the mathematical curricular complexity (*i.e.* winding curricular tortuosity) due directly to mathematics subject entanglements. The learning gaps or cavities initiated by modular extraction were hypothesized herein to be proportional to losses within both overall student mastery and cumulative learning success. This is the fourth ISTES conference presentation using the novel metric entitled the mathematical curricular complexity from curricular analytics techniques (Aronson & Aronson, 2025; Aronson & Aronson, 2026; Aronson et al., 2025). However, herein the mathematical curricular complexity metric is utilized for the first time on both a *pre*-college level curriculum as well as on an entire *multi*-year purely mathematics curriculum.

Hypothesis Toward Proof of Concept

In this particular work, we concentrate on early school mathematics education and the implications of learning gaps leading directly to missing cavity areas of *non*-proficiency. Mathematical modular deficiencies are hypothesized herein to propagate towards not only a mere loss in academic credits (*i.e.* learning time) but also significant learning gaps. Due to prerequisite subject matter entanglements, learning loss (modular extraction) can ultimately result in a disproportionately greater loss in overall student mastery and cumulative learning success. The curricular analytics modeling herein is offered as an effective proof of concept as a small-scale test or pilot project designed to demonstrate the technical feasibility of using curricular analytics modeling techniques to quantitatively analyze *pre*-college mathematics curricula. The proof of concept herein is pending a subsequent

longitudinal kindergarten through eighth grade mathematics classroom in-depth study.

Method

Curricular Analytics to Calculate the Mathematical Curricular Complexity

Curricular analytics is a graph theory framework for quantifying curricular tortuosity, which provides a quantitative measurement of subject matter entanglements for each module of interest ($\square_i$) within a set of (i) total mathematical learning modules (Heileman et al., 2018). The blocking factor (A_i) measures the extent to which the learning module of interest ($\square_i$) blocks the ability to master other modules within the curriculum wherein the module of interest ($\square_i$) merely serves as a prerequisite. The delay factor (B_i) represents the length of the longest path within the corresponding curriculum flowchart graph, which contains the module of interest ($\square_i$).

STEM curricula routinely contain learning modules, which must be completed in sequential order (*i.e.* prerequisites and laboratory corequisites), wherein the ability to navigate these pathways is absolutely critical for student success (Wigdahl *et al.*, 2014). Modular Complexity (C_i) is calculated as the arithmetic sum ($\sum_i (A_i + B_i)$) of the blocking factor (A_i) and delay factor (B_i) for the mathematical learning module of interest ($\square_i$). The mathematical curricular complexity, a novel metric, is calculated as the arithmetic sum complexity ($\sum_i C_i$) for all required and elective mathematics learning modules within the curriculum. Herein, prerequisites for each module were determined by our particular research group. Flow charts were created with curricular MS Excel® code *via* the curricularanalytics.org site.

Zearn® Kindergarten Through Eighth Grade Mathematics Curricula

The Zearn® curricular pacing map (Zearn Curricular Pacing Map, 2022®) shown in Figure 1 below was used herein as a model mathematics curriculum by week, semester, and year (Grade, G) wherein one curricular week was set equal to a single academic credit within the curricular analytics coding, and an academic semester was set equal to eighteen weeks or eighteen academic credits.

Zearn® is a U.S. *non*-profit educational software organization who launched their kindergarten through fifth grade mathematics curriculum in 2017, and their kindergarten through eighth grade mathematics curriculum in 2022 through the Bill and Melinda French Gates® Foundation (Mollenkamp, 2022). Zearn® content addresses topics like adaptive mathematics skills, conceptual understanding, and procedural fluency. Zearn® mathematics platform is modular and moves in helical rotation.

Zearn® math is now used by 25% of U.S. elementary (kindergarten through fifth grade) school students and more than 1.5 million middle school (sixth grade through eighth grade) students and 50,000 *pre*-college teachers in total. The true efficacy of Zearn® K-8 mathematics has been clearly demonstrated within a large, urban school

district (Morrison *et al.*, 2019). Furthermore, the Zearn® mathematics platform is routinely used by accelerated mathematics education programs in numerous U.S. states.



Figure 1. The Zearn K-8 Mathematics Curriculum Analyzed Herein using Curricular Analytics Techniques (Semester 1, Semester 2)

Curricular Analytics Modeling of the Zearn® K-5th and 6th through 8th Grade Curriculum

The Zearn® kindergarten through fifth grade mathematics curriculum encompassed the first fifty (50) coded lines for curricular analytics analysis (curricularanalytics.org) as shown in Table 1. The Zearn® kindergarten through

fifth (K-5) grade mathematics curricular analytics flowchart graph encompassed 252 academic credits with a total mathematical curricular complexity of 955 as shown in Figure 2. The curricular analytics flowchart connects each module with its prerequisite within key learning areas wherein the prerequisites for each module were determined by our particular research group.

Table 1. The Zearn Kindergarten through Fifth (K-5) Grade Mathematics Curriculum Encompassed the First Fifty (50) Coded Lines within the Curricular Analytics Analysis

Course ID	Course Name	Prefix	Number	Prerequisites	Corequisites	Strict-Corequisites	Credit Hours	Institution	Canonical Name	Term
1	Numbers to 10	KM	1				9			1
2	2D and 3D Shapes	KM	2				2			1
3	Comparison of Length	KM	3				7			1
4	Numbers to 5 Digital	KM	11		1		11			1
5	Numbers to 10 Digital	KM	31		1;3		7			1
6	Comparison of Capacity	KM	310				1			2
7	Number Pairs	KM	4	1;3			9			2
8	Numbers 10 to 20	KM	5	1			6			2
9	Numbers to 10 Digital	KM	311	1;3;5	6		4			2
10	Numbers to 15 Lab	KM	41	1;3	7;8		7			2
11	Numbers to 20 Lab	KM	51	1;3	7;8		7			2
12	Analyzing Shapes	KM	6	2			2			2
13	Add and Substract Small Numbers	G1M	1	1;7			9			3
14	Meet Place Value	G1M	2	8			7			3
15	Measure Length	G1M	3	5;6			2			3
16	Measure Length	G1M	310	5;6;15			1			4
17	Add Subtract Big Numbers	G1M	4	8;13			7			4
18	Work with Shapes	G1M	5	12			4			4
19	Add Subtract to 100	G1M	6	13	17		6			4
20	Add Subtract Friendly Numbers	G2M	1	13;17;19			3			5
21	Explore Length	G2M	2	15			3			5
22	Counting and Place Value	G2M	3	8;14			5			5
23	Add Subtract Solve	G2M	4	17;19	20		7			5
24	Add Subtract and Solve	G2M	410	17;19;20;23			1			6
25	Add Substract Big Numbers	G2M	5	13;17;19;20			5			6
26	Equal Groups	G2M	6	20	25		4			6
27	Length Money and Data	G2M	7	15;21			5			6
28	Shapes Times and Fractions	G2M	8	18			3			6
29	Multiply Divide Friendly Numbers	G3M	1	20			6			7
30	Measure It	G3M	2	15;21;27			5			7
31	Multiply Divide Tricky Numbers	G3M	3	25	29		7			7
32	Find the Area	G3M	4	18;28;29;31			4			8
33	Fractions as Numbers	G3M	5	29;31			8			8
34	Display Data	G3M	6	27;30			2			8
35	Shapes and Measurement	G3M	7	18;28	32		4			8
36	Add Subtract and Round	G4M	1	20;23;25			5			9
37	Measure and Solve	G4M	2	15;21;30			1			9
38	Multiply and Divide Big Numbers	G4M	3	29;31			9			9
39	Construct Lines Angles and Shapes	G4M	4	18;28;32;35			3			9
40	Construct Lines Angles and Shapes	G4M	410	18;28;32;35;39			1			10
41	Equivalent Fractions	G4M	5	26;33			10			10
42	Decimal Fractions	G4M	6	33;38	41		4			10
43	Multiply and Measure	G4M	7	29;31;37;38			3			10
44	Place Value with Decimal Fraction	G5M	1	33;41;42			5			11
45	Base Ten Operations	G5M	2	42	44		8			11
46	Add and Stracct Fractions	G5M	3	33;38;41;42			4			11
47	Multiply and Divide Fractions and Decimals	G5M	4	38;41;42;44;46			1			11
48	Multiply and Divide Fractions and Decimals	G5M	410	38;41;42;44;46;47			7			12
49	Volume Area and Shapes	G5M	5	32;35;39;40			5			12
50	The Coordinate Plane	G5M	6	30;35;37;39	49		6			12

Course ID	Course Na	Prefix	Number	Prerequisi	Corequisit	Strict-Core	Credit Hou	Institution	Canonical	Term
51	Area and S	G6M	1				5			13
52	Introducir	G6M	2				4			13
53	Rates and	G6M	3		52		4			13
54	Dividing F	G6M	4				4			13
55	Arithmetic	G6M	5				1			13
56	Arithmetic	G6M	510	55			3			14
57	Expressior	G6M	6				4			14
58	Rational N	G6M	7	52;54			5			14
59	Data Sets	G6M	8				5			14
60	Putting It A	G6M	9	51;52;53;5	56;58;59		1			14
61	Scale Dra	G7M	1	51			3			15
62	Introducir	G7M	2	52			4			15
63	Measuring	G7M	3	51	61		3			15
64	Proprtiona	G7M	4	52;53	62		4			15
65	Rational N	G7M	5	54;55;56;58			4			15
66	Expressior	G7M	6				6			16
67	Angles Tri	G7M	7	61;63			4			16
68	Probabilit	G7M	8	59			5			16
69	Putting It A	G7M	9	61;62;63;6	67;68		3			16
70	Rigid Tran	G8M	1	51;61;63;67			5			17
71	Dilations	G8M	2	51;61;63;6	70		4			17
72	Linear Rel	G8M	3	62;64;66			4			17
73	Linear Equ	G8M	4	72			4			17
74	Functions	G8M	5	73			1			17
75	Functions	G8M	510	73			4			18
76	Associatic	G8M	6	59;68			4			18
77	Exponents	G8M	7	58;65			5			18
78	Pythagore	G8M	8	58;63;67;70;71			4			18
79	Putting it A	G8M	9	70;71	76;77;78		1			18

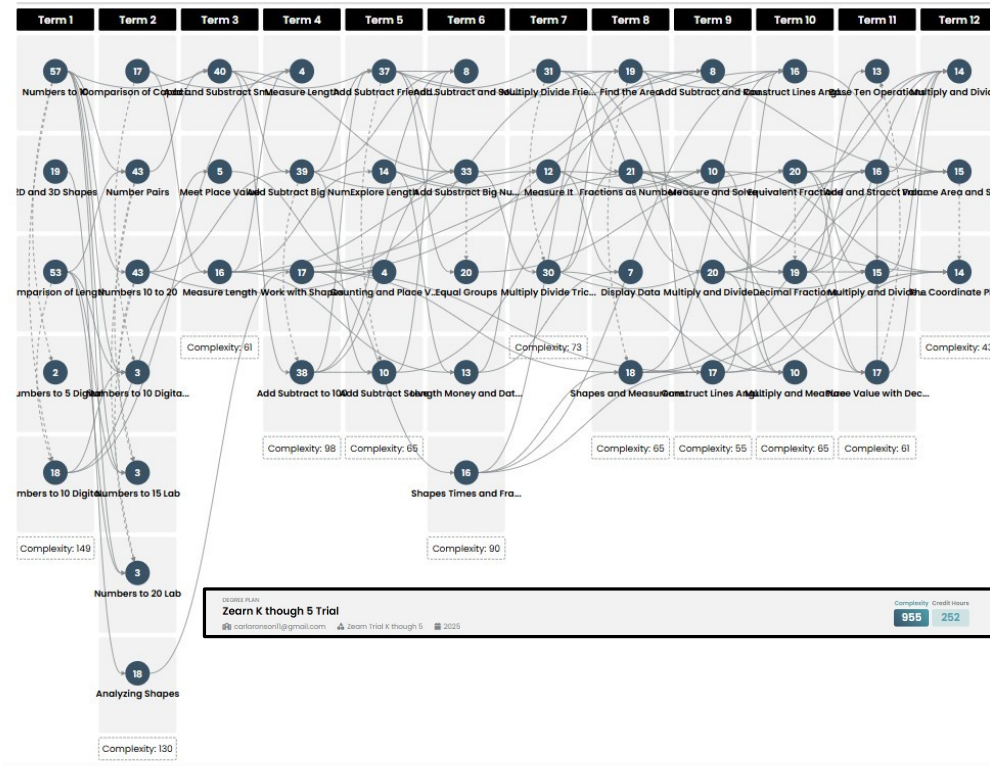


Figure 2. K-5 Curricular Analytics Flowchart

The Zearn® sixth grade through eighth grade mathematics curriculum encompassed coded lines number 51 through 79 for curricular analytics analysis (curricularanalytics.org) as shown in Figure 3.

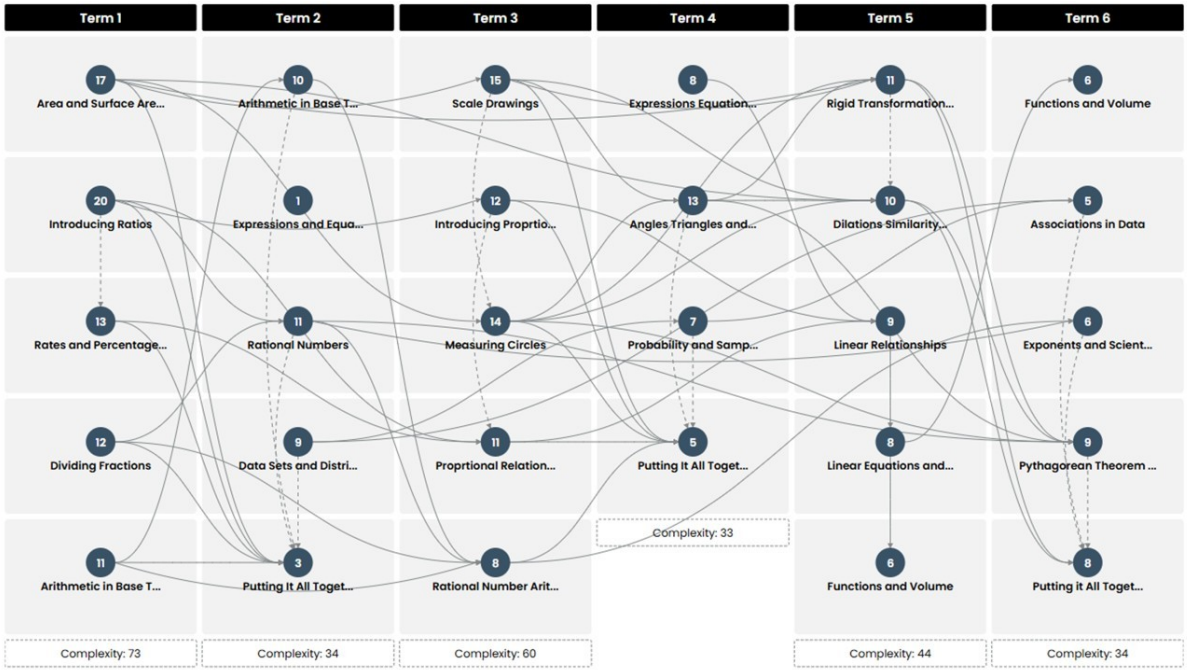


Figure 3. G6–G8 Coded Lines

The Zearn® sixth grade through eighth grade mathematics curriculum analytics flowchart graph is shown in Figure 4.

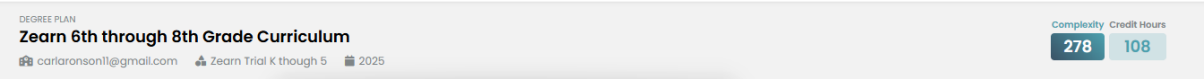


Figure 4. G6–G8 Curricular Analytics Flowchart

The Zearn® sixth grade through eighth grade mathematics curriculum encompassed 108 total credits and had a total mathematical curricular complexity of 278. The flowchart shown in Figure 4 connects mathematics learning modules with their prerequisite modules in common learning areas wherein specific prerequisites were determined by our particular research group.

Results

Curricular Analytics Modeling of Modular Deficiencies within the Zearn® K-5 Mathematics Curriculum

The Zearn® curricular pacing map shown in Figure 1 (Zearn Curricular Pacing Map, 2022) was used herein as a model kindergarten through eighth grade mathematics curriculum by both week, semester, and year (K – G8 from Figure 1). For example, within the kindergarten through fifth grade modular deficiency was modeled to propagate in a spillover effect. The spillover effect occurs wherein an inaction in one area produces a negative consequence in another related area. Spillover effects often occur in interconnected systems such as an academic curriculum.

Hence, not mastering a particular module herein during the designated mathematics learning time period was modeled to impact a student later due merely to prerequisites thereby producing negative consequences as we proceed forward within the curriculum as shown in Figure 5 below for kindergarten (K) through the fifth grade (G5) measurement modules.

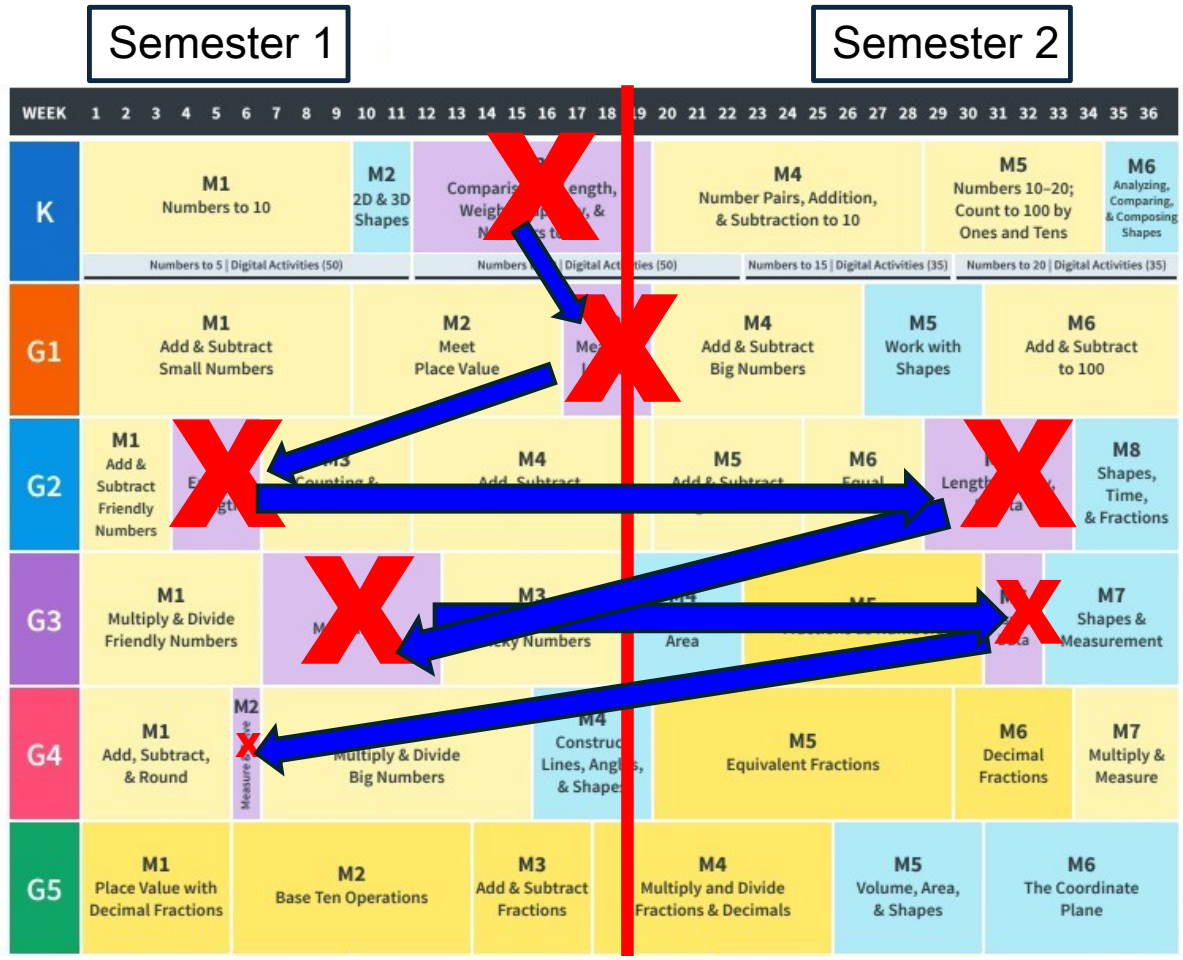


Figure 5. Modular Deficiency in Measurement, K–5

Examples of disproportionate losses in mathematical curricular complexity as a result of modular deficiency propagation within the kindergarten (K) through fifth grade (G5) are shown in Table 2. The original, intact kindergarten through fifth grade curricula encompassed 252 total credits and had a total mathematical curricular complexity of 955 (Figure 2).

The content summarized as Changes in Curricular Complexity, K–5 is presented in Table 2.

Table 2. Changes in Curricular Complexity, K–5

Math Learning Area Deficiency Cavity	Number of Credits Remaining	Loss of Credits (weeks), %	Resulting Complexity	Loss of Complexity (Learning Loss), %
Geometry (K – 2nd Grade)	245	2.78	882	7.64
Geometry (K – 5th Grade)	220	12.7	699	26.8

Math Learning Area Deficiency Cavity	Number of Credits Remaining	Loss of Credits (weeks), %	Resulting Complexity	Loss of Complexity (Learning Loss), %
Measurement (K – 5th	234	7.1	844	11.6
Operations (3rd – 5th)	205	18.7	711	25.5
Whole Numbers and Operations (K – 5th)	141	44.0	320	66.5
Measurement and Geometry Visualization (K - 5th)	205	18.7	568	40.5

From the Zearn® curricular pacing map (see Figure 1), an example of geometric modular deficiency within the sixth through eighth grade (G6 – G8 of Figure 1) is shown in Figure 6. Initial modular deficiency was modeled to catalyze an initiation-propagation series of events, wherein propagation effectively sustains the deficiency outgrowth.

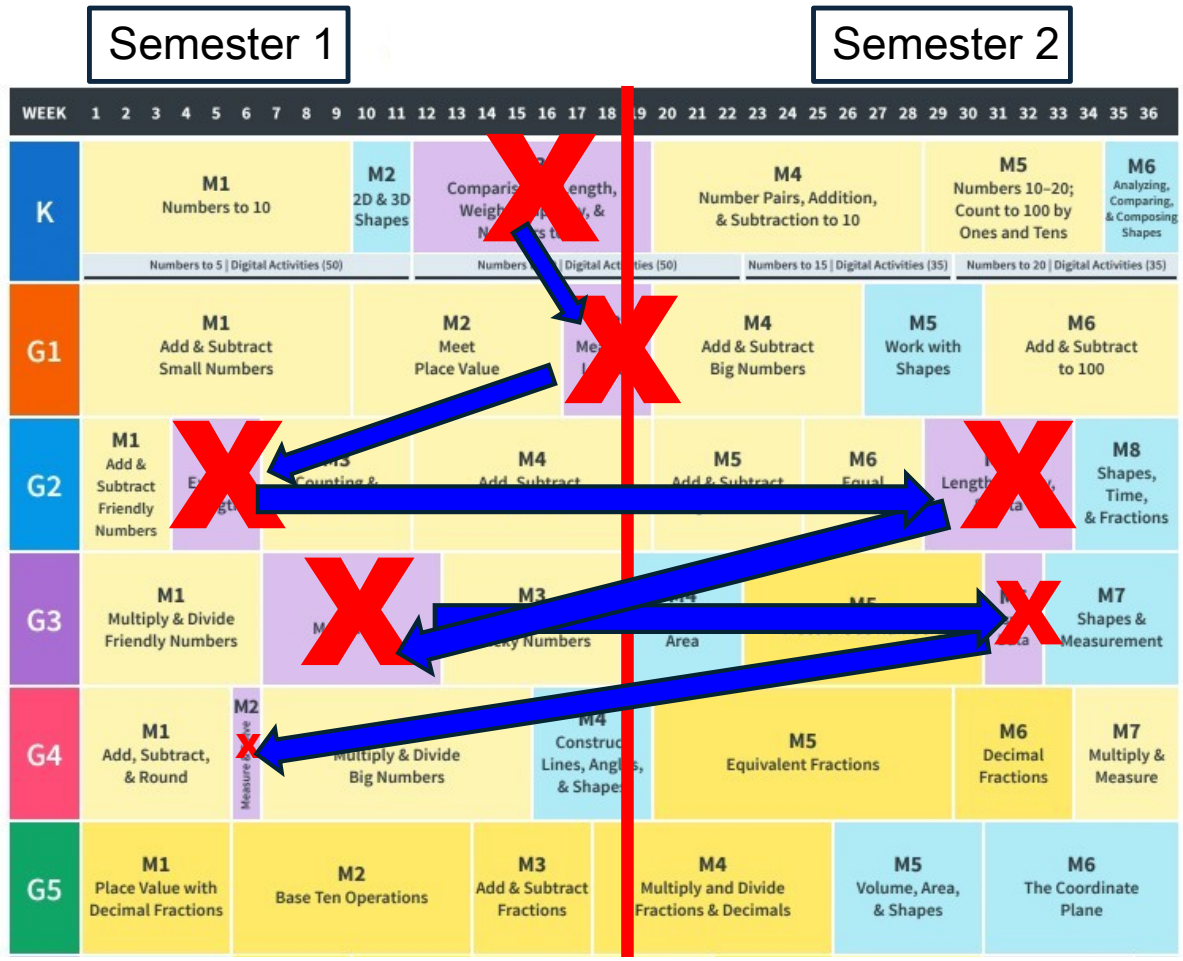


Figure 6. Geometric Modular Deficiency, G6–G8

Examples of disproportionate losses in mathematical curricular complexity as a result of modular deficiency propagation within the sixth grade (G6) through eighth grade (G8) are shown in Table 3. The original, intact sixth

grade through eighth grade (G6 – G8) curricula encompassed 108 total credits and had a total mathematical curricular complexity of 278 (see Figure 4).

Table 3. Changes in Curricular Complexity, G6–G8

Math Learning Area Deficiency Cavity	Number of Credits Remaining	Loss of Credits (weeks), %	Resulting Complexity	Loss of Complexity (Learning Loss), %
Geometry (6^{th} – 8^{th} Grade)	80	25.9	182	34.5
Probability and Statistics (6^{th} – 8^{th} Grade)	94	13.0	257	7.6
Whole Number Operations to Decimals, Fractions (6^{th} – 8^{th} Grade)	86	20.4	215	22.7
Algebra (6^{th} – 8^{th} Grade)	69	36.1	184	33.8
Fractions & Algebra (6^{th} – 8^{th} Grade)	47	56.5	125	45.0

Discussion

Curricular Analytics Modeling of Modular Deficiency

In all the kindergarten through eighth grade (K-8) test cases of modular learning deficiencies, the decrease in mathematical curricular complexity was proposed to be proportional to learning loss. *Non*-mastery was recorded herein as a zero (0%) toward overall learning whereas partial mastery may, in fact, prove otherwise within a longitudinal K-8 mathematics classroom study. In kindergarten through fifth grade, the percent (%) loss in mathematical curricular complexity from propagation tended to be far greater than merely the percent (%) loss in academic credits (classroom weeks) as shown in Table 2. This effect was observed to generally increase with increasing deficiency level (see Table 2). Hence, in primary school mathematics timely learning was shown to be highly important as the effect of modular deficiencies leading to mathematically deficient cavity areas was highly disproportionate when compared to classroom time (academic credits).

Nevertheless, learning accuracy is also vital because if a foundational mathematics concept is learned inaccurately, misunderstandings can carry into more advanced mathematical subject areas (Newcastle University, 2026; Elliott *et al.*, 2026). Nevertheless, in sixth grade through eighth grade the percent (%) loss in mathematical curricular complexity from propagation tended to be comparable to the percent (%) loss in academic credits (classroom weeks) due primarily to a lower number of prerequisite modules in G6-G8 compared to the kindergarten through fifth grade portion of the Zearn® mathematics curriculum (Table 33). Herein, we modeled kindergarten through fifth grade mathematics and sixth grade through eighth grade mathematical learning as separate entities whereas they are, in fact, academically linked. It should also be noted that remedial work, although cumbersome, can correct deficiencies in mathematical learning depending on an individual student's own intellectual persistence and total mathematical learning effort.

Conclusions

Propagation of Modular Learning Deficiencies

Learning breaks encompassing merely not accomplishing homework, absences, health issues, dishonesty, *etc*. by students result in modular deficiencies (*i.e.* lack of mathematical modular mastery in Figure 1). Mathematical modular deficiencies are proposed herein to propagate into mathematical deficiency fissures and finally into mathematical learning area cavities much like tooth decay. Further experimental work within the mathematics classroom is needed to prove disproportionality to learning loss and mathematical curricular complexity whereas herein we show merely a proof of concept. Due directly to prerequisite subject matter entanglements, mathematical curricular complexity (learning) is potentially lost at a far greater rate than merely proportional to classroom time. Curricular analytics flowcharts and the mathematical curricular complexity can indeed be used to quantitatively model a wholly *pre*-college mathematics curricula as a proof of concept.

Learning Diffusion Model

Many mathematical modules of different classroom time lengths lie between beginning kindergarten and ending eighth grade. By extracting (removing) any particularly critical barrier learning module *via non*-mastery, the diffusional pathway becomes considerably less tortuous. A decreased helical learning tortuosity and smaller helical learning persistence pathlength indicate deficiency whereas true mathematical learning rigor can only be achieved by diffusing through (mastering) all of the barrier K-8 mathematical learning modules.

Recommendations

The mathematical curricular complexity can be used to plan optimal kindergarten through twelfth grade mathematics curricula, which are more student learner-friendly in terms of order, timing, and consistent intensity (Wigdahl, 2014; Heileman et al., 2018). Our trajectory is to further analyze the modular deficiency losses in learning for the higher secondary school (ninth grade through twelfth grade) mathematics curriculum *via* the mathematical curricular complexity toward STEM college majors. A mathematics classroom longitudinal study is proposed herein to follow the curricular analytics modeling proof of concept demonstrated here. The curricular analytics modeling effectively lays down a theoretical groundwork. We are proposing a longitudinal study with research design that repeatedly observes both the same individual students over an extended period of time as well as following student cohorts in order to track long-term trends, changes, developments, and causal relationships. A longitudinal study would greatly help prove the modular deficiency initiation-propagation and spillover effect learning deficiency theories proposed herein.

Statements and Declarations

Acknowledgments and Notes: The authors wish to thank the Department of Foundational Sciences within the College of Marine Sciences and Maritime Studies at the Texas A&M University Galveston campus for their generous financial support of this research and its remote presentation at the 2026 International Conference on

Education in Mathematics, Science, and Technology (ICEMST) of the International Society for Technology, Education, and Science (ISTES) organization in Cappadocia (Nevsehir), Turkiye.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Aronson, C. L. & Aronson, S. J. (2025). A Mathematical Curricular Complexity Analysis of U.S. Data Science and Data Engineering Bachelor's Degree Programs. In M. Shelley & O. T. Ozturk (Eds.), *Proceedings of ICONEST 2025-- International Conference on Engineering, Science and Technology*, Los Angeles, California, USA International Society for Technology, Education, and Science (ISTES) Organization, pp. 1 - 20.

Aronson, C. L. & Aronson, S. J. (2026). An Economic Mathematical Curricular Complexity Analysis of Data Science Bachelor's Degree Programs at U.S. Public Universities. Presented at *ICEMST 2026 – International Conference on Education in Mathematics, Science and Technology*, Cappadocia (Nevsehir), Turkiye, International Society for Technology, Education, and Science (ISTES) Organization, May 6, 2026. Paper # 19190957.

Aronson, C. L., Ravi, R., Black, K. R., & Aronson, S. J. (2025). The Effect of Mathematical Curricular Complexity on College Major Stratification. In M. Shelley & O. T. Ozturk (Eds.), *Proceedings of ICEMST 2025-- International Conference on Education in Mathematics, Science and Technology*, Antalya, Turkiye. International Society for Technology, Education, and Science (ISTES) Organization, pp. 34 - 53.

Elliott, M. C., Johnson, E. C., Gregory, K. P. & Hughes, C. E. (May 14, 2026). *Rethinking the Nature and Extent of Inductive Effects in Organic Compounds,* J. Chem. Ed., Washington, DC: American Chemical Society. Retrieved from https://pubs.acs.org/doi/10.1021/acs.jchemed.6c00141

Heileman, G. L., Abdallah, C. T., Slim, A., & Hickman, M. (2018). *Curricular Analytics: A Framework for Quantifying the Impact of Curricular Reforms and Pedagogical Innovations, arXiv, arXiv:1811.09676*.

Retrieved from and. https://arxiv.org/pdf/1811.09676 Retrieved from https://arxiv.org/pdf/1811.09676 and https://curricularanalytics.org/help

Morrison, J. R.; Wolf, B., Ross, S. M., Risman, K. L., & McLemore, C. C. (2019). *Efficacy Study of Zearn Math in a Large Urban School District*. Center for Research and Reform in Education, Johns Hopkins University School of Education, Baltimore, MD: Johns Hopkins University. Retrieved from:. https://jscholarship.library.jhu.edu/server/api/core/bitstreams/efb681c4-4020-4a40-b7a2-851a52b27007/content

Mollenkamp, D., (December 12, 2022). *What Is Zearn - the Math Platform the Gates Foundation Is Betting Big On? EdSurge, Instructional Trends*. Retrieved from https://www.edsurge.com/news/2022-12-12-what-is-zearn-the-math-platform-the-gates-foundation-is-betting-big-on

National Center for Education Statistics (2020). *Adult Numeracy in the United States*. https://nces.ed.gov/pubs2020/2020025/index.asp

Newcastle University. (May 18, 2026). Debunking a core chemistry concept taught in classrooms everywhere, *Phys. Org*. Retrieved from https://phys.org/news/2026-05-debunking-core-chemistry-concept-taught.html

O'Connor, J., Peterson, E. JAG Gomes, J. A. G., Moreno, H., & Cocke, D. L. (2006). *Helical Learning Model*. Modeling in Physics and Physics Education, 20-25. Retrieved from. https://www.researchgate.net/publication/268273823_Helical_Learning_Model

Wigdahl, J., Heileman, G. L., Slim, A, & Abdallah, C. T., (June 15-18, 2014). Curricular Efficiency: What Role Does It Play in Student Success? *Proceedings of the 2014 American Society for Engineering Education*, Indianapolis, Indiana, Paper ID# 9609.

Zearn Curricular Pacing Map (2022). *New York City, NY: Zearn*. Retrieved from and. https://webassets.zearn.org/resources/ImplementationPacingMap.pdf Retrieved from https://webassets.zearn.org/resources/ImplementationPacingMap.pdf, https://about.zearn.org/curriculum

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Multimedia Presentation Modalities and Mathematical Problem Solving: An Experimental Study Within Mayer's Cognitive Theory of Multimedia Learning

Enise Ozcankaya [1], Muhammet Demirbilek [2], Cem Turan [3]

[1] Suleyman Demirel University, Turkey, eniseozcankaya7@gmail.com, 0009-0008-3672-2343
[2] Suleyman Demirel University, Turkey, mdbilek@gmail.com, 0000-0001-7448-9206
[3] İstanbul Sabahattin Zaim University, Tukey, cem.turan@izu.edu.tr, 0000-0002-0887-0954
* Corresponding author: Derar Serhan (icemstoffice@gmail.com)

Article Info

Article History
Received:
10 February 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Multimedia learning
Modality effect
Cognitive load
Problem solving
Ctml

Abstract

This study examined the effects of different multimedia presentation modalities on fifth-grade students' mathematical problem-solving performance within the framework of Mayer's Cognitive Theory of Multimedia Learning (CTML). The research compared three instructional conditions: text-only, audio-only, and audio-supported animation. A quantitative repeated-measures experimental design was employed with 27 fifth-grade students enrolled in a public middle school in Türkiye. To reduce order effects, counterbalanced presentation sequences were implemented across participant groups. Students solved mathematics word problems presented in all three modalities, and their responses were coded dichotomously as correct or incorrect. Both descriptive and inferential statistical analyses revealed statistically significant differences across conditions. Students achieved the highest accuracy in the text-only condition, followed by the audio-supported animation condition, whereas the audio-only condition produced the lowest performance. The findings suggest that text-based presentation may provide more effective cognitive support by allowing self-paced processing and reducing working memory demands during mathematical reasoning tasks. The study contributes to multimedia learning literature by providing experimental evidence regarding modality effects in mathematics education and offers practical implications for the design of cognitively efficient digital learning materials.

Citation: Ozcankaya, E., Demirbilek, M., & Turan, C. (2026). Multimedia Presentation Modalities and Mathematical Problem Solving: An Experimental Study Within Mayer's Cognitive Theory of Multimedia Learning. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 100–113), Cappadocia, Nevsehir / Türkiye. ISTES.

CC BY NC SA

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

OPEN ACCESS

Introduction

The integration of multimedia technologies into educational environments has fundamentally transformed instructional practices in contemporary mathematics education. Digital learning environments increasingly incorporate text, narration, animation, and interactive visual materials to improve learner engagement and facilitate conceptual understanding. Particularly in elementary and middle school education, multimedia-supported instructional materials are frequently considered more effective and motivating than traditional teaching methods. However, despite the widespread implementation of multimedia technologies in classrooms, the extent to which different presentation modalities enhance mathematical problem-solving performance remains an important issue in educational research.

The theoretical framework of the present study is based on Mayer's Cognitive Theory of Multimedia Learning (CTML), which explains learning as an active cognitive process involving the integration of verbal and visual information (Mayer, 2002). According to Mayer (2002), meaningful learning occurs when learners actively select relevant information, organize it into coherent mental structures, and integrate these structures with prior knowledge. CTML is grounded in three major assumptions: the dual-channel assumption, the limited-capacity assumption, and the active-processing assumption. The dual-channel assumption proposes that humans process auditory/verbal and visual/pictorial information through separate cognitive systems. The limited-capacity assumption suggests that each channel possesses restricted working memory resources, while the active-processing assumption emphasizes that effective learning requires active cognitive engagement rather than passive reception of information.

Within CTML, instructional effectiveness depends not on the quantity of multimedia elements but on how efficiently instructional materials support cognitive processing and reduce unnecessary mental load. Mayer (2002) argues that multimedia instruction should facilitate meaningful schema construction while minimizing extraneous cognitive processing. One of the most influential principles derived from CTML is the modality principle, which suggests that distributing information across auditory and visual channels may enhance learning efficiency by preventing overload within a single processing system. However, the theory also emphasizes that poorly organized multimedia environments may increase cognitive load and negatively affect learning performance.

This issue becomes particularly important in mathematics education because mathematical problem solving requires intensive cognitive processing, including working memory coordination, attentional control, logical sequencing, and mental representation skills. Mathematical tasks often require learners to process multiple pieces of information simultaneously while mentally maintaining intermediate solution steps. Therefore, the modality through which information is presented may substantially influence learners' comprehension processes and problem-solving performance.

The relationship between instructional design and cognitive processing has also been extensively discussed within Cognitive Load Theory (Sweller, 1988). Cognitive Load Theory proposes that instructional materials exceeding learners' working memory capacity may hinder learning efficiency and academic performance (Sweller et al.,

2019). Younger learners are particularly vulnerable to cognitive overload because executive functioning and self-regulation skills continue developing during middle school years. Consequently, multimedia environments that appear visually rich or technologically advanced may not necessarily improve learning outcomes if they impose excessive cognitive demands on learners.

Previous studies have demonstrated that multimedia-supported instructional environments may positively influence students' engagement, motivation, and conceptual understanding in mathematics education (Moreno & Mayer, 2007). Similarly, animation-supported instructional materials have been associated with improved conceptual visualization and learner attention in digital learning environments (Schnotz & Rasch, 2005). Nevertheless, empirical findings regarding modality effectiveness remain inconsistent. While some studies report that multimedia integration enhances learning outcomes, others indicate that excessive auditory or visual stimulation may create cognitive interference and negatively affect comprehension processes.

One important explanation for these inconsistent findings involves the transient nature of auditory information. Spoken information disappears immediately after presentation, forcing learners to retain information mentally while simultaneously processing mathematical operations (Baddeley, 2012). In contrast, text-based presentation allows rereading, self-paced processing, and sequential information organization, potentially reducing working memory burden during mathematical reasoning tasks. From the perspective of CTML, learner-paced processing may therefore play a critical role in mathematical comprehension and problem-solving accuracy.

Although multimedia learning literature has expanded substantially during the past two decades, much of the existing research has primarily focused on the general effectiveness of multimedia instruction rather than directly comparing specific presentation modalities within equivalent mathematical tasks. Earlier studies grounded in Cognitive Theory of Multimedia Learning (CTML) and the modality principle demonstrated that distributing information across auditory and visual channels may reduce cognitive load and improve learning efficiency (Mousavi et al., 1995; Tindall-Ford et al., 1997). Similarly, meta-analytic studies examining multimedia learning environments reported that modality effects may vary according to instructional design, learner characteristics, and task complexity (Ginns, 2005; Höffler & Leutner, 2007). More recent findings additionally suggest that modality effectiveness may depend on learner age, mathematical difficulty level, and cognitive processing requirements (Paas & Sweller, 2014). Nevertheless, relatively limited research has examined text-only, audio-only, and audio-supported animation conditions together within the same repeated-measures experimental structure among fifth-grade students in mathematics education.

Accordingly, the purpose of this study is to investigate the effects of different multimedia presentation modalities on fifth-grade students' mathematical problem-solving performance within the framework of Mayer's Cognitive Theory of Multimedia Learning. Specifically, the study compares students' performances across text-only, audio-only, and audio-supported animation conditions in order to better understand how modality differences may influence mathematical comprehension and problem-solving accuracy.

Method

Research Design

This study employed a quantitative repeated-measures experimental design to investigate the effects of different multimedia presentation modalities on fifth-grade students' mathematical problem-solving performance. In repeated-measures designs, the same participants are exposed to all experimental conditions, allowing direct comparison of performance across modalities while minimizing individual variability and between-subject differences. The independent variable of the study was presentation modality, consisting of three instructional conditions: text-only, audio-only, and audio-supported animation. The dependent variable was students' mathematical problem-solving performance.

To control for potential order effects, a counterbalanced presentation structure was implemented throughout the experimental procedure. Participants were divided into three sequencing groups and exposed to the instructional modalities in different orders in order to reduce the influence of fatigue, learning, and practice effects on performance outcomes.

Participants

The participants consisted of 27 fifth-grade students enrolled in a public middle school in Türkiye. All participants were native Turkish speakers and belonged to a relatively homogeneous age group in order to maintain developmental consistency during the experimental process. Students were selected randomly and distributed homogeneously across different academic achievement levels to minimize achievement-related bias. Participation in the study was voluntary.

Instructional Materials

The instructional materials consisted of mathematics word problems developed according to the fifth-grade mathematics curriculum. The problems focused on mathematical topics that students had previously encountered during classroom instruction in order to minimize the influence of prior content unfamiliarity on performance.

A total of nine mathematical word problems were used in the study. Each presentation modality included three equivalent mathematical problems with similar levels of cognitive difficulty. The problems were designed at a moderate difficulty level appropriate for fifth-grade learners and required basic mathematical reasoning and problem-solving skills.

Each mathematical problem was presented in one of the following modalities:

1. **Text-Only Condition:**
   Problems were presented in written textual format without auditory or animated support.
2. **Audio-Only Condition:**

Problems were presented through narration without written text or visual representation.

3. **Audio-Supported Animation Condition:**
   Problems were presented through synchronized animation and audio narration.

All instructional materials contained equivalent mathematical content across conditions. Students responded individually in written form immediately after each presentation.

Procedure

The study was conducted in a controlled classroom environment during a single experimental session. Each participant completed all three modality conditions. To minimize order effects, participants were assigned to three counterbalanced sequencing groups.

The first group of students received the modalities in the sequence of text–audio–audio-supported animation. The second group received the sequence of audio-supported animation–audio–text, whereas the third group experienced the sequence of audio–text–audio-supported animation.

Each mathematical problem was presented individually, and students were given approximately two minutes to solve each problem. Including instructions, transitions, and response collection, the total duration of the experimental session ranged between 20 and 25 minutes.

Student responses were coded dichotomously as correct or incorrect. Correct responses were coded as "1," whereas incorrect responses were coded as "0" for statistical analysis.

Data Analysis

The collected data were analyzed using both descriptive and inferential statistical methods. Frequencies, percentages, means, and standard deviations were calculated to compare students' mathematical problem-solving performances across presentation modalities.

Because the study employed a repeated-measures design involving three related conditions and the data did not meet the assumptions required for parametric repeated-measures analysis, a Friedman test was conducted to determine whether the observed differences across modalities were statistically significant.

The results of the Friedman test revealed a statistically significant difference between presentation modalities, $\chi^2(2) = 10.38$, $p = .006$. Descriptive findings indicated that students achieved the highest performance in the text-only condition (62.9%), followed by the audio-supported animation condition (37.0%), whereas the audio-only condition produced the lowest performance level (18.5%). These findings suggest that presentation modality had a significant effect on fifth-grade students' mathematical problem-solving performance within the framework of Mayer's Cognitive Theory of Multimedia Learning (CTML).

Findings

The findings revealed statistically significant differences in students' mathematical problem-solving performances across the three presentation modalities. A Friedman test was conducted to compare students' performances in the text-only, audio-only, and audio-supported animation conditions. The analysis demonstrated a significant effect of presentation modality on mathematical problem-solving performance, $\chi^2(2) = 10.38$, $p = .006$.

The mean rank scores obtained from the Friedman analysis indicated that the text-only condition produced the highest performance level among the three modalities, followed by the audio-supported animation condition and the audio-only condition. Pairwise post-hoc comparisons were conducted using Wilcoxon signed-rank tests with Bonferroni adjustment. The results demonstrated a statistically significant difference between the text-only and audio-only conditions ($Z = -2.83$, $p = .005$), as well as between the text-only and audio-supported animation conditions ($Z = -2.11$, $p = .035$). However, the difference between the audio-only and audio-supported animation conditions was not statistically significant ($Z = -1.39$, $p = .166$). These findings suggest that text-based presentation provided significantly stronger support for mathematical problem-solving performance compared with auditory instructional formats.

Table 1. The Descriptive Performance Results Across the Three Modality Conditions

Presentation Modality	**Correct Response Rate (%)**	**Mean Rank**	**Performance Rank**
Text Only	62.9	1.65	1
Audio - Supported Animation	37.0	2.04	2
Audio Only	18.5	2.31	3

As presented in Table 1, students achieved the highest level of accuracy in the text-only condition, whereas the audio-only condition produced the lowest performance outcomes. The audio-supported animation condition resulted in moderate performance levels but did not surpass the text-based condition. Lower mean rank values indicate higher performance levels.

The findings suggest that text-based presentation may provide more effective cognitive support for fifth-grade learners during mathematical problem-solving tasks. One possible explanation is that textual presentation allows learners to regulate the pace of information processing through rereading and sequential cognitive organization. In contrast, the relatively low performance observed in the audio-only condition may be associated with the transient nature of auditory information, which can increase working memory demands during mathematical reasoning processes.

Although the audio-supported animation condition produced higher performance levels than the audio-only condition, the results indicate that multimedia enrichment alone may not necessarily improve mathematical problem-solving performance. Rather, the effectiveness of multimedia environments may depend on the extent to which instructional materials align with learners' cognitive processing capacities and support manageable information processing.

Additionally, the consistency of performance patterns across counterbalanced sequencing groups suggests that the observed differences were associated primarily with modality characteristics rather than presentation order effects. Overall, the findings indicate that fifth-grade students performed more successfully when mathematical information was presented in a self-paced textual format compared with transient auditory instructional conditions.

Discussion

The present study investigated the effects of different multimedia presentation modalities on fifth-grade students' mathematical problem-solving performance within the framework of Mayer's Cognitive Theory of Multimedia Learning (CTML). The findings revealed statistically significant differences among the three modality conditions, $\chi^2(2) = 10.38$, $p = .006$. Students achieved the highest level of performance in the text-only condition (62.9%), followed by the audio-supported animation condition (37.0%), whereas the audio-only condition produced the lowest performance level (18.5%). These findings indicate that presentation modality significantly influences mathematical comprehension and problem-solving performance among elementary-level learners.

One of the most important contributions of the present study is that it problematizes the widespread assumption that technologically enriched multimedia environments inherently improve learning outcomes. Although multimedia-supported instructional systems are frequently associated with increased motivation and learner engagement, the present findings suggest that multimedia richness alone does not guarantee improved academic performance. Despite the inclusion of synchronized animation and narration, students demonstrated significantly stronger performance in the self-paced text-only condition. This result suggests that instructional effectiveness may depend less on technological complexity and more on the extent to which instructional materials support cognitively manageable information processing.

The findings provide empirical support for the central assumptions of Cognitive Theory of Multimedia Learning and Cognitive Load Theory. CTML proposes that learning occurs most effectively when learners can actively process information across limited-capacity visual and auditory channels without experiencing excessive cognitive overload (Mayer, 2002). Similarly, Cognitive Load Theory emphasizes that instructional materials imposing unnecessary processing demands may reduce learning efficiency and interfere with schema construction (Sweller, 1988). In the present study, the text-based condition may have facilitated more efficient cognitive processing by allowing learners to regulate the pace of information intake, revisit mathematical relationships, and organize problem structures sequentially during reasoning tasks.

The superiority of the text-only condition is particularly important within the context of elementary mathematics education. Mathematical problem solving requires learners to coordinate numerical information, maintain intermediate reasoning steps mentally, and process linguistic instructions simultaneously. Under such cognitively demanding conditions, transient auditory input may place additional pressure on working memory systems. This interpretation is consistent with Baddeley (2012) working memory framework, which suggests that spoken information may increase cognitive burden because auditory input disappears immediately after presentation and cannot be revisited easily by learners.

The relatively weak performance observed in the audio-only condition further supports previous modality-effect literature. Earlier studies demonstrated that purely auditory instructional formats may overload working memory when learners are unable to control the pacing of information presentation (Tabbers et al., 2004). Similarly, Mayer and Moreno (2003) argued that rapidly disappearing auditory information may interfere with learners' ability to organize and integrate information during cognitively complex tasks. The current study extends these findings specifically to elementary-level mathematical problem solving and provides experimental evidence suggesting that auditory-only presentation may be insufficient for supporting mathematical reasoning among younger learners.

Another important finding concerns the limited effectiveness of the audio-supported animation condition. Although this condition produced higher performance levels than the audio-only modality, it did not outperform the text-based presentation. This finding suggests that animation and narration do not necessarily reduce cognitive demands automatically. Rather, multimedia effectiveness appears to depend on whether multimedia elements are synchronized appropriately with learners' cognitive processing capacities and attentional resources. Previous multimedia learning research similarly reported that animations may facilitate learning only when instructional elements reduce rather than increase extraneous cognitive load (Schnotz & Rasch, 2005). Otherwise, simultaneous visual and auditory stimuli may divide attention and create additional processing demands, particularly for younger learners whose executive functioning and attentional regulation abilities are still developing.

The present study also contributes methodologically to multimedia learning research. While previous studies have frequently examined multimedia effectiveness in general instructional contexts, relatively limited research has directly compared text-only, audio-only, and audio-supported animation modalities within the same repeated-measures experimental structure among fifth-grade students. Furthermore, the implementation of counterbalanced sequencing procedures strengthened the internal validity of the study by minimizing order-related biases such as practice and fatigue effects. The inclusion of inferential statistical analysis through the Friedman test additionally strengthened the interpretability of the findings beyond descriptive percentage comparisons alone.

Despite these contributions, several limitations should be acknowledged. First, the study was conducted with a relatively small sample of 27 students from a single public middle school in Türkiye, limiting the generalizability of the findings. Second, the study focused exclusively on short-term mathematical problem-solving performance and did not examine long-term retention, transfer of learning, or conceptual permanence. Third, the research investigated a limited number of mathematical tasks and presentation modalities. Future research may therefore

examine modality effects across different mathematical domains, larger participant samples, and diverse developmental levels. Additional studies may also investigate the influence of learner-related variables such as prior mathematical knowledge, digital literacy, attentional control, reading comprehension, and working memory capacity on multimedia learning performance.

In conclusion, the findings of the present study suggest that self-paced textual presentation may provide more effective cognitive support for fifth-grade students during mathematical problem-solving tasks than audio-only or audio-supported animation conditions. More importantly, the results demonstrate that the effectiveness of multimedia learning environments depends not simply on technological richness but on how successfully instructional materials align with learners' cognitive architecture and working memory limitations. Consequently, multimedia instructional design in mathematics education should prioritize cognitive clarity, learner-paced processing, instructional simplicity, and manageable information flow rather than excessive multimedia complexity.

Implications

The findings of the present study have important theoretical, pedagogical, and practical implications for multimedia instructional design and mathematics education. Most importantly, the results demonstrate that instructional effectiveness in digital learning environments should not be evaluated solely in terms of technological richness, visual attractiveness, or the quantity of multimedia elements embedded within instructional materials. Rather, the findings suggest that effective multimedia instruction depends primarily on the extent to which instructional materials align with learners' cognitive processing capacities and support manageable working memory demands. From the perspective of Mayer's Cognitive Theory of Multimedia Learning (CTML), multimedia environments become effective not because they contain more technological components, but because they facilitate cognitively efficient information processing (Mayer, 2002).

A particularly significant implication of the present study concerns the role of learner-paced processing in elementary mathematics education. The statistically superior performance observed in the text-only condition suggests that younger learners may benefit substantially from instructional formats that allow them to regulate the pace and sequence of information processing independently. Mathematical problem solving requires learners to coordinate numerical operations, maintain intermediate reasoning steps mentally, and process linguistic information simultaneously. Text-based instructional formats may support these processes more effectively by enabling rereading, sequential organization, and repeated cognitive rehearsal during problem solving. Consequently, instructional designers and classroom teachers should consider incorporating self-paced textual scaffolding into digital mathematics materials, particularly for elementary and middle school learners whose executive functioning and attentional regulation skills are still developing.

The findings additionally raise important concerns regarding the instructional use of purely auditory presentation formats in mathematics education. The significantly lower performance observed in the audio-only condition suggests that transient auditory information may impose additional cognitive demands on younger learners during

mathematical reasoning tasks. Because spoken information disappears immediately after presentation, learners may experience difficulty retaining numerical relationships and problem structures within working memory. Therefore, auditory instructional materials should be designed cautiously and supported with complementary textual or visual scaffolding to reduce transient information effects and support cognitive organization processes.

Another important implication concerns the instructional use of multimedia animations in mathematics education. Although the audio-supported animation condition produced higher performance levels than the audio-only condition, it did not outperform the text-based condition. This finding suggests that animation and narration alone are insufficient to guarantee improved learning outcomes. Instead, multimedia effectiveness appears to depend on whether multimedia elements reduce or increase extraneous cognitive load during learning. Educational software developers and instructional designers should therefore avoid excessive multimedia stimulation and prioritize instructional clarity, cognitive simplicity, and coherent information flow when developing digital mathematics learning environments.

The study also contributes theoretically to the broader multimedia learning literature by providing empirical evidence that modality effects vary according to learner characteristics, cognitive processing demands, and task structure. Much of the previous multimedia learning literature has emphasized the motivational advantages of technologically enriched learning environments. However, the present findings demonstrate that increased engagement or visual attractiveness does not necessarily translate into improved academic performance. In this respect, the study supports a growing body of research emphasizing that multimedia learning environments should be designed according to evidence-based cognitive principles rather than technological assumptions alone.

From a practical perspective, the findings may provide guidance for curriculum developers, educational technology designers, classroom teachers, and developers of digital mathematics platforms. Digital instructional systems targeting younger learners should prioritize cognitively manageable presentation formats that facilitate information organization and reduce unnecessary attentional division. Particularly in mathematics education, instructional materials should support learners' reasoning processes rather than overload working memory through simultaneous competing stimuli.

Finally, the present study contributes methodologically to modality-effect research by directly comparing text-only, audio-only, and audio-supported animation conditions within the same repeated-measures experimental structure. The statistically significant differences identified through inferential analysis strengthen the evidence regarding the influence of presentation modality on mathematical problem-solving performance among fifth-grade students. These findings may therefore serve as a foundation for future research examining cognitive processing differences across multimedia learning environments, developmental levels, and mathematical domains.

Limitations and Future Research

Despite its contributions to multimedia learning research, the present study has several limitations that should be

considered when interpreting the findings. First, the study was conducted with a relatively small sample consisting of 27 fifth-grade students from a single public middle school. Although the participants were selected randomly and distributed homogeneously across different achievement levels, the limited sample size restricts the generalizability of the findings to broader educational populations.

Second, the study focused specifically on short-term mathematical problem-solving performance. Long-term learning outcomes such as retention, transfer of knowledge, and conceptual permanence were not examined. Since multimedia learning environments may influence immediate performance differently from long-term cognitive development, future studies should investigate whether modality effects remain stable over extended instructional periods.

Another limitation concerns the restricted number of instructional modalities examined in the study. The research compared only text-only, audio-only, and audio-supported animation conditions. Other multimedia combinations such as text-plus-visual, interactive simulations, adaptive digital systems, or student-controlled multimedia environments were not included. Future research may therefore examine a wider variety of multimedia instructional formats in mathematics education.

Additionally, the study primarily evaluated performance outcomes through correct and incorrect responses. Although this approach provided clear descriptive comparisons across modalities, the study did not directly measure learners' cognitive load, attentional processes, or emotional responses during mathematical problem solving. Future studies may incorporate cognitive load scales, eye-tracking systems, think-aloud protocols, or learning analytics methods to examine the cognitive mechanisms underlying modality differences more comprehensively.

The developmental characteristics of learners should also be considered in future research. Since working memory capacity, attentional regulation, and self-paced learning abilities change across age groups, modality effects may differ substantially between younger learners, adolescents, and adults. Therefore, future studies should compare modality effects across different educational levels and developmental stages.

Another important area for future research involves individual learner differences. Variables such as prior mathematical knowledge, reading comprehension ability, digital literacy, working memory capacity, and learning preferences may significantly influence how learners process multimedia instructional materials. Investigating these learner-related factors may contribute to a more personalized understanding of multimedia learning environments.

Finally, future research may investigate modality effects within different mathematical domains and task structures. The cognitive demands of arithmetic operations, geometry, algebraic reasoning, and word problems may vary considerably, potentially influencing how learners respond to multimedia instruction. Examining modality effects across diverse mathematical contexts may therefore provide a more comprehensive understanding of multimedia learning processes in mathematics education.

Overall, despite these limitations, the present study contributes valuable evidence regarding the relationship between multimedia presentation modalities and mathematical problem-solving performance among fifth-grade students. The findings provide a foundation for future research examining how instructional design principles derived from Mayer's Cognitive Theory of Multimedia Learning may support more cognitively efficient and learner-centered mathematics instruction.

Conclusion

The present study examined the effects of text-only, audio-only, and audio-supported animation modalities on fifth-grade students' mathematical problem-solving performance within the framework of Mayer's Cognitive Theory of Multimedia Learning (CTML). The findings revealed statistically significant differences across presentation modalities, with students demonstrating the highest performance in the text-only condition and the lowest performance in the audio-only condition.

The results suggest that self-paced textual presentation may provide more effective cognitive support for younger learners during mathematical reasoning tasks by facilitating sequential information processing and reducing working memory demands. In contrast, transient auditory information may increase cognitive load and hinder mathematical comprehension when learners cannot regulate the pace of information processing independently.

Another important finding is that multimedia enrichment alone does not necessarily improve learning outcomes. Although the audio-supported animation condition produced better performance than the audio-only condition, it did not surpass the effectiveness of text-based presentation. These findings indicate that instructional effectiveness depends not merely on technological richness, but on the extent to which instructional materials align with learners' cognitive architecture and support manageable cognitive processing.

Overall, the study emphasizes that multimedia instructional design in mathematics education should prioritize cognitive clarity, learner-paced processing, and instructional simplicity rather than excessive multimedia complexity.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). Cambridge University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313–331.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6), 722–738.

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *Psychology of learning and motivation* (pp. 85–139). Academic Press. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334.

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed, pp. 27–42). Cambridge University Press.

Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development*, *53*(3), 47–58. https://doi.org/10.1007/BF02504797

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). The split-attention effect. In *Cognitive load theory* (pp. 111–128). Springer. https://doi.org/10.1007/978-1-4419-8126-4_9

Tabbers, H. K., Martens, R. L., & van Merriënboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*(1), 71–81. https://doi.org/10.1348/000709904322848824

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*(4), 257–287.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177. https://doi.org/10.1007/s10648-005-3951-0

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

AI-Enhanced Pedagogical Transformation in Teacher Education: A Systematic Framework

Eyüp Yurt

Bursa Uludağ University, Turkiye, 0000-0003-4732-6879
Corresponding author: Eyüp Yurt (icemstoffice@gmail.com)

Article Info

Article History
Received:
7 February 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
AI-enhanced learning
AI in education
Pedagogical transformation
Teacher education

Abstract

The rapid proliferation of generative AI tools presents teacher education with a critical choice: restrict AI use or prepare preservice teachers for responsible integration. This study argues that systematically integrating AI literacy as a central competency equips preservice teachers to use AI effectively and responsibly while teaching students about its potential and limitations. Rather than treating AI as a threat, teacher education must embrace transformative integration, positioning AI literacy alongside content and pedagogical knowledge. Empirical research reveals consistent findings: when teacher educators model responsible AI use and critical evaluation, when AI learning is grounded in authentic pedagogical tasks, and when competency development addresses technical skills, critical evaluation, ethical reasoning, and student guidance capacity, preservice teachers emerge prepared for AI-augmented classrooms. The study proposes a transformation framework integrating TPACK theory with multi-dimensional AI competency, grounded in scholarship from Shulman, Darling-Hammond, Cochran-Smith, and situated cognition theory. It articulates five essential dimensions: conceptual AI knowledge, pedagogical AI application, critical evaluation, ethical and equity understanding, and student guidance competency. This approach prepares educators to help all students thrive in AI-influenced contexts.

Citation: Yurt, E. (2026). AI-Enhanced Pedagogical Transformation in Teacher Education: A Systematic Framework. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 114–128), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

OPEN ACCESS

Introduction

As ChatGPT and similar tools rapidly proliferate in educational contexts, teacher educators face a critical question: Should we restrict AI use or prepare teachers for responsible integration? The traditional response has often been restrictive: ban AI, emphasize plagiarism risks, prohibit student engagement. However, recent empirical research points in a different direction. Trust et al. (2023) advocate for "responsible integration" rather than AI restriction in teacher education. Wen and Wen (2024) demonstrate that preservice teachers can engage in sophisticated AI evaluation. Moorhouse et al. (2024) report that structured AI training significantly develops preservice teachers' AI competency. These investigations suggest that teacher education must systematically integrate AI rather than restrict it, equipping future educators with competencies essential for their evolving professional contexts. This transformation requires grounding in foundational teacher education theory while extending established frameworks to address contemporary realities.

Theoretical Foundations: Establishing Conceptual Framework

Pedagogical Content Knowledge and Teacher Expertise

Understanding how to prepare teachers for AI integration requires grounding in foundational teacher education theory. Lee Shulman (1986, 1987), in seminal work distinguishing teacher knowledge from other professional expertise, introduced the concept of "pedagogical content knowledge" (PCK)—the distinctive knowledge teachers develop about teaching particular content to students. Shulman argues that effective teaching requires far more than subject matter expertise; teachers must understand how to represent content, anticipate student misconceptions, employ subject-specific teaching strategies, and adapt instruction to diverse learners. This foundational insight—that teaching requires integrated understanding of both content and pedagogy—remains central to understanding how teachers should approach any new tool or practice, including AI.

Grossman (1990) extended Shulman's framework, identifying that PCK encompasses: knowledge of representations and explanations for teaching particular topics; understanding of students' common misconceptions and learning difficulties; knowledge of curricular materials and instructional strategies most effective for teaching specific content; and understanding of how to sequence instruction for particular student populations. This comprehensive conception of teacher knowledge emphasizes that effective teaching results from deep, integrated understanding developed through sustained engagement with content and students, not from technical skill with particular tools. When considering AI integration, this theoretical foundation suggests that teachers cannot simply be trained to use AI tools; rather, they must develop sophisticated understanding of how AI relates to disciplinary content, pedagogical strategies, and specific student needs.

TPACK: Extending PCK to Technology Integration

Mishra and Koehler (2006), recognizing that technology increasingly shapes contemporary teaching, extended Shulman's PCK framework into TPACK (Technological Pedagogical Content Knowledge). Koehler and Mishra (2009) explain that effective technology integration requires understanding the complex relationships among

content, pedagogy, and technology within specific contexts. The TPACK framework encompasses seven knowledge domains: Content Knowledge (CK)—deep understanding of disciplinary content; Pedagogical Knowledge (PK)—understanding of general teaching and learning strategies; Technological Knowledge (TK)—understanding of various technologies and their capabilities; Pedagogical Content Knowledge (PCK)—understanding of how to teach particular content effectively; Technological Content Knowledge (TCK)—understanding of how technologies can represent and present particular content; Technological Pedagogical Knowledge (TPK)—understanding of how technologies can support particular pedagogical approaches; and TPACK proper—the integrated understanding of how content, pedagogy, and technology interact within specific contexts.

Critically, Koehler and Mishra (2009) emphasize that the intersections among these domains are not merely additive. Technology integration is not content knowledge plus technology knowledge. Rather, effective technology integration requires understanding how specific technologies interact with particular content and pedagogical approaches in ways that cannot be fully understood by examining components separately. A calculator changes how we teach mathematics; a microscope changes how we teach biology; email changes how we communicate about learning. Similarly, AI changes how teachers approach lesson planning, assessment, differentiation, and student engagement. Understanding these transformations requires TPACK—integrated understanding of how AI specifically interacts with their disciplinary content and pedagogical approaches within their particular student contexts.

Teacher Education Excellence: Foundations and Principles

Darling-Hammond (2000, 2010), in comprehensive research on teacher preparation effectiveness, identifies key characteristics distinguishing high-quality teacher education programs. She argues that effective programs share several features: coherent, research-based curriculum grounded in learning science and subject matter knowledge; extensive clinical practice with feedback and mentoring; integration of theory and practice through design of authentic teaching tasks; reflection and examination of teaching practice; collaborative learning communities supporting ongoing development; and attention to equity and diverse learners. Importantly, Darling-Hammond (2000, p. 166) emphasizes that "teacher education must be thought of as a coherent whole, not as a collection of discrete courses," suggesting that programs must be systematically organized around core principles rather than assembled from isolated units.

Building on this foundation, Cochran-Smith (2005) argues that teacher education has responsibility to prepare teachers for social justice—to understand how teaching can either reproduce or challenge educational inequities. She emphasizes that teacher preparation must develop both technical competence and critical consciousness regarding whose interests are served by particular educational practices. This perspective is particularly relevant for AI integration, where concerns about bias, equity, and access must be central rather than peripheral considerations. Teacher education that prepares teachers merely to use AI tools without developing critical consciousness regarding AI's potential to perpetuate biases or exacerbate inequities fails its social justice responsibility.

Teacher Learning and Professional Development

Understanding how to effectively integrate AI into teacher preparation also requires understanding how teachers learn. Guskey (2002) and Guskey and Yoon (2009) synthesize research on effective professional development, finding that learning is most effective when: it focuses on content and pedagogy rather than generic teaching skills; it involves active engagement with content rather than passive reception; it provides adequate time for learning and practice; it includes collaborative engagement with colleagues; and it connects to teachers' existing knowledge and concerns. Additionally, Guskey and Yoon (2009) emphasize that professional development should be evaluated not merely by participants' satisfaction but by actual changes in teacher practice and student learning outcomes.

Joyce and Showers (1980, 2002), in influential work on teacher learning and transfer, identify that coaching and peer observation significantly enhance teachers' capacity to implement new practices. They argue that merely presenting information about new practices is insufficient; teachers require opportunity to practice in supportive environments with feedback from peers and mentors. This research suggests that effective AI integration in teacher preparation requires not merely presenting information about AI but rather providing preservice teachers with authentic opportunities to engage with AI in teaching tasks, supported by feedback and peer dialogue.

Teacher Educators as Models

An important principle grounded in substantial research holds that teacher educators' instructional practices profoundly influence preservice teachers' developing conceptions of teaching. Wideen et al. (1998) find that what teacher educators do in their own teaching communicates more powerfully to preservice teachers than explicit instruction about teaching methods. Lortie (1975), in his foundational work on teacher socialization, argues that preservice teachers bring substantial "apprenticeship of observation"—years of watching teachers teach—that shapes their expectations about what teaching should look like. Teacher educators cannot simply tell preservice teachers how to teach differently; rather, by modeling different approaches, they demonstrate what alternative teaching looks like and communicate that change is possible.

Feiman-Nemser (2001) extends this insight, arguing that "teacher educators need to model the kind of teaching they hope preservice teachers will emulate" and that "the teaching practices of teacher educators significantly shape preservice teachers' images of good teaching." This principle is particularly relevant for AI integration. If teacher educators incorporate AI thoughtfully into their instruction, explicitly addressing both affordances and limitations while demonstrating critical evaluation, they model responsible AI integration. Conversely, if teacher educators restrict or ignore AI while preservice teachers encounter it elsewhere, they fail to help preservice teachers develop coherent understanding of how AI should be approached.

Authentic Learning Tasks and Situated Cognition

Lave and Wenger (1991), in foundational work on situated cognition, argue that meaningful learning occurs

through engagement with authentic practices within communities of practice. Knowledge is not abstract, decontextualized, and transferable; rather, it is situated—embedded in authentic contexts and activities. Apprentices learning a trade don't first learn abstract principles in isolation then apply them to real work; rather, they learn through engagement in real work, progressively taking on more responsibility with guidance from experienced practitioners.

This theoretical perspective suggests that preservice teachers should learn about AI integration not through abstract instruction but through engagement in authentic teaching tasks. Rather than discussing how ChatGPT might be used in lesson planning, preservice teachers should actually engage in lesson planning using ChatGPT, experiencing both its affordances and limitations directly. This authentic task engagement, grounded in real pedagogical work they will perform in future classrooms, generates more meaningful and transferable learning than abstract discussion. Merriam and Bierema (2014) extend this perspective, arguing that adults learn most effectively when they understand how learning relates to their professional identities and goals. Preservice teachers are motivated by recognition that AI literacy will directly affect their effectiveness as teachers. Learning about AI within authentic teaching contexts—designing actual lessons, creating assessments, planning differentiation—connects to preservice teachers' emerging professional identities and goals in ways that abstract training cannot.

Empirical Evidence: Current Research on AI in Teacher Education

Editorial Guidance: Integration Over Restriction

Trust et al. (2023), writing as editors of a leading teacher education journal, explicitly advocate for reframing how teacher education approaches AI. Rather than implementing restrictive policies banning AI use, they recommend a "model and guide" approach where teacher educators incorporate AI into their own instruction while simultaneously modeling critical evaluation and ethical use. This editorial guidance emerges from recognition that restriction alone proves insufficient—students will encounter AI regardless of institutional policies, and teachers unprepared to engage with AI critically may integrate it poorly in future classrooms. Trust, Whalen, and Mouza recommend several concrete practices: establishing explicit guidelines for appropriate AI use rather than blanket bans; engaging preservice teachers in critical examination of AI output, comparing AI-generated materials with human-designed alternatives; discussing ethical implications throughout coursework rather than treating ethics as peripheral; and positioning AI as a pedagogical enhancement that extends teacher capacity rather than as either magic solution or dangerous threat. This balanced perspective acknowledges genuine concerns while recognizing that responsible integration serves preservice teachers better than prohibition. Importantly, they emphasize that teacher educators themselves must model this integration, not merely teach about it abstractly—consistent with Feiman-Nemser (2001) argument that teacher educators' practices shape preservice teachers' conceptions.

Qualitative Research: Preservice Teachers' Critical Evaluation Processes

Wen and Wen (2024) conducted an illuminating qualitative study documenting how 16 preservice teachers used

ChatGPT when designing curriculum units for a literacy methods course. Over eight weeks, preservice teachers worked in pairs, iteratively prompting ChatGPT to develop five-day lesson plans. The researchers identified a consistent four-stage design cycle: First, the inquiry and prompting stage involved formulating questions to ChatGPT, with sophistication of prompts increasing over time from generic requests to highly specific, strategically designed queries incorporating district standards and student population characteristics. Second, the examination and evaluation stage revealed that preservice teachers did not passively accept AI suggestions but rather critically assessed them against their understanding of specific students, pedagogical approaches, and curricular requirements.

Third, the modification and tinkering stage showed preservice teachers engaging in iterative refinement, either reformulating prompts to be more specific or manually revising AI-generated content to better align with their pedagogical vision. Fourth, the synthesis and organization stage involved integrating various elements—some AI-generated, some human-created—into coherent lesson plans reflecting pedagogical judgment. Wen and Wen (2024) characterize this as "thinking partnership"—preservice teachers using AI not as authoritative source but as collaborative thinking partner generating ideas, examples, and initial drafts that preservice teachers evaluate and refine. This mirrors what Vygotsky (1978) described as scaffolding within the zone of proximal development: AI provides support enabling preservice teachers to accomplish more sophisticated planning than they could independently.

Critically, Wen and Wen found that preservice teachers with stronger TPACK—integrated understanding of content, pedagogy, and context—more effectively evaluated AI output and adapted suggestions appropriately. Those with weaker pedagogical content knowledge struggled more to assess whether AI suggestions were suitable. This finding, grounded in Shulman (1986) conception of PCK and extended through TPACK framework, reveals that AI competency cannot be developed in isolation from disciplinary expertise and pedagogical knowledge; rather, critical AI use fundamentally depends on strong TPACK. Additionally, preservice teachers spontaneously developed what might be called critical awareness, expressing concern about accuracy (and conducting fact-checking using Google), recognizing that AI-generated materials were generic requiring contextual customization, and raising ethical questions about appropriate AI use. These concerns emerged organically from engagement with authentic tasks rather than from external mandate, supporting Lave and Wenger (1991) situated cognition perspective that meaningful learning emerges through authentic activity.

Intervention Research: Structured AI Training and Competency Development

Moorhouse and colleagues (2024) conducted a mixed-method intervention study with 54 graduate preservice language teachers in Hong Kong, implementing an 11-week structured training course on generative AI integration. The curriculum was explicitly designed around five competency dimensions, progressing from foundational AI literacy through pedagogical application toward ethical considerations and student guidance. Following Darling-Hammond (2000) principles, activities were grounded in authentic language teaching tasks: preservice teachers used ChatGPT to generate grammar explanations then reflected on pedagogical appropriateness; they created differentiated reading comprehension materials then critically analyzed whether

differentiation effectively served diverse learners.

Pre-post quantitative measurement revealed significant gains across competency dimensions: pedagogical application increased with effect size d=0.78, critical understanding d=0.82, AI literacy knowledge d=0.91, and ethical understanding d=0.71. However, confidence in guiding students in AI use showed more modest improvement (d=0.42), a gap Moorhouse et al. attributed to the course emphasizing teachers' own AI use more heavily than strategies for teaching students about AI. This finding highlights the importance of comprehensive curriculum addressing all competency dimensions, not merely some.

Qualitative interviews revealed that preservice teachers had developed balanced perspective recognizing AI as enhancement rather than replacement. Participants articulated recognition that AI-generated materials required customization, stating that "ChatGPT gives me a starting point, but I have to think about my actual students and my pedagogical goals to make it work." Importantly, ethical concerns emerged spontaneously, with participants raising considerations about AI potentially perpetuating biases, environmental costs of training large systems, and equitable access—concerns consistent with Cochran-Smith (2005) emphasis on social justice in teacher education. These findings provide empirical validation that structured, multi-dimensional AI training enhances preservice teachers' competency across dimensions and cultivates the critical, ethical reasoning essential for responsible integration.

Systematic Transformation Framework

Foundational Principle: AI Literacy as Core TPACK

The framework begins with fundamental reconceptualization grounded in Shulman (1986), Koehler and Mishra (2009), and TPACK research: AI literacy is not an optional add-on or peripheral topic but rather a central competency within teacher preparation, representing a contemporary extension of TPACK for the age of generative AI. This means AI should be integrated throughout programs rather than isolated in separate "AI in Education" courses. When AI is addressed only in isolated units while core methodology courses remain unchanged, preservice teachers may view AI as a specialized topic disconnected from their actual teaching. Instead, systematic integration throughout the curriculum communicates that AI is integral to contemporary teaching, consistent with Darling-Hammond (2000) emphasis on coherent, integrated teacher education. This requires deliberate pedagogical decision-making: How does AI relate to literacy instruction? To mathematics pedagogy? To science inquiry? How might AI serve diverse learners in your discipline? Grounding AI within disciplinary contexts ensures preservice teachers understand AI not as generic tool but as having specific affordances and limitations within their subject matter. This approach aligns with Shulman (1986) insight that teaching requires integrated understanding of content and pedagogy—AI integration similarly requires integrated understanding of how AI interacts with specific content and pedagogical approaches.

Five-Dimensional Competency Development

Effective teacher preparation programs should systematically develop all five competency dimensions throughout

coursework, ensuring comprehensive preparation rather than addressing only isolated aspects. The first dimension, conceptual AI knowledge, involves helping preservice teachers understand how generative language models function at a conceptual level appropriate for K-12 educators—not requiring deep technical expertise but providing sufficient understanding to make informed pedagogical decisions. Drawing on Cochran-Smith (2005) social justice perspective and research on technology and equity (Selwyn, 2019), this dimension should include critical understanding of how AI systems are trained, what biases might be embedded in training data, and how these technical features translate into pedagogical implications.

The second dimension, pedagogical AI application, involves helping teachers recognize opportunities to use AI for authentic teaching purposes within their discipline, consistent with situated cognition theory (Lave & Wenger, 1991). Teachers should generate lesson ideas, create differentiated materials, develop formative assessments using AI. Importantly, this learning should be discipline-specific, grounded in actual content standards and pedagogical approaches used in their subject areas. Following Grossman (1990) conception of PCK, preservice teachers should understand not merely that AI exists but specifically how AI can represent content, address student misconceptions, and support particular pedagogical strategies within their discipline.

The third dimension, critical evaluation, represents perhaps the most essential competency and most directly requires strong PCK grounding. Preservice teachers must develop sophisticated capacity to assess AI output, recognizing when AI provides accurate, appropriate, high-quality content and when it falls short. This requires grounding in disciplinary expertise: a history teacher must be able to evaluate whether AI-generated historical documents are accurate and appropriately contextualized; a mathematics teacher must recognize whether AI-generated explanations are mathematically sound; a literature teacher must assess whether AI characterizations of texts are defensible. This dimension fundamentally depends on strong pedagogical content knowledge—it cannot be developed through abstract training but requires engagement with actual disciplinary content and pedagogical purposes.

The fourth dimension, ethical and equity reasoning, involves helping teachers think carefully about moral dimensions of AI integration. Drawing on Cochran-Smith (2005) social justice framework and recent work on AI and bias (Noble, 2018; O'Neill, 2016), this dimension addresses how teachers should handle plagiarism concerns if students use AI, how to ensure AI use supports rather than undermines equity, what privacy and data security concerns teachers should understand, and what biases particular AI tools might embed. This ethical reasoning should pervade the entire program rather than being contained in a single ethics unit—consistent with integrative approaches to teacher education ethics (Hansen, 2001).

The fifth dimension, guiding students in AI use, prepares teachers to teach students about AI responsibly. This includes establishing clear expectations for appropriate AI use, teaching students to evaluate AI output critically, discussing ethical concerns, and helping students understand when AI is appropriate versus inappropriate as a learning tool. This dimension, Moorhouse et al. (2024) note, remains significantly underdeveloped in current teacher preparation despite its importance and urgency.

Pedagogical Design Principles for Integration

Integration throughout teacher preparation should follow several pedagogical design principles grounded in foundational teacher education research. First, embed AI throughout courses rather than isolating it, consistent with Darling-Hammond (2000) emphasis on coherent, integrated curricula. When a literacy methods course addresses AI, discussion should be integrated throughout the semester, not compressed into single units. Preservice teachers should see AI applications when discussing comprehension instruction, writing pedagogy, assessment, and differentiation—recognizing AI's relevance across content.

Second, ground AI learning in authentic teaching tasks, directly applying Lave and Wenger (1991) situated cognition theory. Rather than abstract discussion of AI capabilities, preservice teachers should engage in real lesson design, assessment development, and curriculum planning using AI. This authenticity, validated by Wen & Wen's research and supported by Merriam and Bierema (2014) work on adult learning, creates meaningful learning grounded in actual work teachers will perform.

Third, model and explicitly teach critical evaluation, applying Feiman-Nemser (2001) principle that teacher educators must model desired practices. Teacher educators should pause during instruction to critically examine AI output, asking aloud: Is this accurate? Is this appropriate for these students? What might be missing? What assumptions underlie this suggestion? Additionally, provide explicit frameworks and heuristics supporting preservice teachers' evaluation, consistent with cognitive apprenticeship approaches (Collins et al., 1989).

Fourth, explicitly address ethics rather than treating it as peripheral, applying Hansen (2001) perspective that ethics should be integrated throughout teacher education. Throughout coursework, address ethical concerns: plagiarism implications if students use AI, potential for embedded biases, privacy and data security, environmental costs, equity considerations. When discussing lesson design with AI, simultaneously discuss ethical implications. This integration communicates that ethical reasoning is central to responsible teaching.

Fifth, progress instruction from concrete to abstract engagement, applying constructivist learning principles (Piaget, 1954; von Glasersfeld, 1989). Early coursework might engage preservice teachers in concrete AI tasks: generating and evaluating AI-generated lesson ideas, creating differentiated materials using AI, adapting AI suggestions for specific students. As competence develops, engagement becomes more abstract and critical: evaluating AI's theoretical assumptions, considering alternative frameworks AI might miss, discussing AI's role in broader educational ecosystems.

Sixth, create collaborative learning communities where preservice teachers discuss AI challenges together, share strategies, and learn from peers' experiences, applying communities of practice theory (Wenger, 1998). Wen & Wen found that pairs of preservice teachers developed richer understanding than individuals working alone; collaborative dialogue exposed alternative approaches and prompted deeper questioning. Seventh, scaffold progressively toward independence, applying Wood et al. (1976) scaffolding theory. Early AI engagement involves highly structured prompts and guided evaluation; progressively, preservice teachers should formulate

their own prompts, conduct independent evaluation, and make autonomous decisions about AI integration with diminishing scaffolding.

Implementation Example: English Language Arts Methods Course

Consider how these principles might be integrated throughout an English language arts methods course, grounded in PCK and TPACK frameworks. Early in the course, teacher educators might introduce AI writing tools and have preservice teachers use ChatGPT to generate writing prompts. Rather than passively accepting generated prompts, preservice teachers would critically evaluate whether prompts are appropriate for their target students and grade levels, whether they align with standards, whether they would generate interesting student writing, and what modifications might improve them. This grounds AI introduction in authentic task while immediately cultivating critical evaluation, consistent with situated cognition principles.

Mid-course, as preservice teachers develop stronger pedagogical content knowledge about writing instruction, they might design complete writing units leveraging AI productively. For example, a unit where students use AI to generate initial drafts then engage in metacognitive revision, learning through comparison between AI output and their own refined writing. Or a unit where students critique AI-generated writing, developing critical thinking while learning standards for quality writing. Throughout, preservice teachers discuss ethical guidelines: How should they communicate about AI to students? How should they handle concerns about fairness or integrity? What expectations should they establish? These discussions apply Cochran-Smith (2005) social justice perspective, helping preservice teachers think critically about how AI integration affects diverse learners.

Late in the course, preservice teachers develop capstone unit plans integrating all learning: AI-enhanced writing instruction demonstrating how AI supports their pedagogical vision while maintaining writing instruction integrity, including explicit plans for teaching students responsible AI use, establishing ethical expectations, and assessing whether students are developing genuine writing competence. This progression—from introductory engagement through critical evaluation through pedagogical design through implementation planning—follows Darling-Hammond (2000) principles for coherent teacher education and Lave and Wenger (1991) theories of situated learning.

Assessment of AI Competency Development

Assessing AI competency requires moving beyond merely testing whether preservice teachers can operate tools toward evaluating integrated understanding specified in the five-dimensional framework. Portfolio assessment proves particularly valuable, applying constructivist assessment principles (Wiggins & McTighe, 2005): preservice teachers compile collections of AI-integrated lesson designs including the lessons themselves, artifacts from design process showing critical evaluation and iteration, reflections on pedagogical decisions, and evidence of student learning when lessons are implemented. Such portfolios reveal authentic evidence of integrated TPACK including AI literacy.

Critical analysis tasks provide another assessment approach: present AI-generated lesson plans, assessments, or materials, then ask preservice teachers to identify strengths and limitations, explain necessary modifications for specific students, discuss potential biases, and articulate ethical concerns. These open-ended tasks, consistent with authentic assessment approaches (Merriam & Bierema, 2014), reveal depth of critical understanding. Pedagogical decision-making scenarios present complex situations requiring judgments about AI integration, assessing quality of reasoning. Teaching practice observation provides direct evidence: observe preservice teachers' actual teaching, evaluating how they integrate AI, how they address student concerns, how they maintain pedagogical integrity while leveraging AI, consistent with clinical approaches to teacher education (Darling-Hammond, 2010). Reflection and metacognition assessment examines preservice teachers' written or spoken reflections on their AI use: Do they demonstrate understanding of AI's role in their teaching? Can they identify areas for growth? Deep reflection reveals metacognitive awareness associated with genuine competency (Mezirow, 1997).

Institutional Implementation and Professional Development

Systematic transformation requires institutional commitment beyond individual faculty innovation. Following Darling-Hammond (2000) research on program coherence, program leadership must clearly communicate that AI literacy is a priority, allocating resources and time to support integration. This involves developing or adopting comprehensive AI competency frameworks guiding curricular redesign and assessment. Perhaps most critically, it requires substantial faculty professional development grounded in adult learning principles (Merriam & Bierema, 2014).

Not all teacher educators have deep understanding of AI, experience using AI tools, or clear vision for integration. Professional development should help faculty develop foundational AI understanding through hands-on experience with tools, encouraging experimentation with different prompts and observation of AI capabilities and limitations. Following Joyce and Showers (2002) research on effective professional development, faculty need collaborative opportunities to practice, receive feedback, and refine their implementation. Faculty should examine how AI specifically applies to their courses: What authentic tasks might involve AI? How can AI enhance rather than replace learning? Faculty should develop understanding of how to assess competency development, determining what artifacts provide evidence and what tasks reveal critical thinking. Faculty need support in addressing their own concerns about AI: questions about academic integrity, worries about student thinking development, concerns about equity, anxieties about technological competency.

Professional development should be ongoing and collaborative rather than one-shot workshops, ideally organized as learning communities where faculty examine student work together, share teaching strategies, and support each other's learning, consistent with communities of practice theory (Wenger, 1998). Programs should also consider whether dedicated expertise is needed—perhaps a faculty member with deep AI knowledge who can lead professional development and support integration across the program. Equitable access represents another institutional imperative, applying Cochran-Smith (2005) social justice perspective: all preservice teachers should have access to AI tools regardless of socioeconomic status. Without equitable access, AI competency development becomes privileged for already-advantaged students.

Addressing Common Concerns

Several predictable concerns may arise as programs begin transformation. Rather than dismissing these concerns, thoughtful engagement proves more productive. Some faculty worry that AI will replace teachers; however, well-designed integration positions AI as tool extending teacher capacity by handling routine tasks, freeing teacher time and cognitive resources for what requires human judgment: understanding individual students, making nuanced pedagogical decisions, providing emotional support and mentorship. This aligns with Shulman (1986) conception of teacher expertise as requiring judgment that cannot be automated.

Some worry students will use AI to cheat; while this is genuine concern rooted in legitimate integrity questions, restriction alone addresses it insufficiently. Better to design assignments where AI use is transparent and constructive—where students must explain learning, where AI becomes thinking tool rather than shortcut. Some faculty feel insufficiently knowledgeable about AI; they need not be experts, merely willing to learn alongside preservice teachers and model intellectual engagement with new knowledge, consistent with reflective practice approaches (Schön, 1983). Others worry about adding content to already-full curricula; integration rather than addition addresses this by reframing existing content through AI lens rather than creating separate units. Finally, some express concerns about equity and access, worrying that AI might exacerbate existing inequities; these concerns should be addressed seriously through explicit attention to how AI integration can be designed for equity rather than dismissing equity concerns.

Conclusion

Empirical evidence, theoretical frameworks spanning decades of teacher education research, and leadership guidance converge on clear direction: Teacher education must transform through thoughtful AI integration rather than defensive restriction. Grounded in Shulman (1986) foundational work on pedagogical content knowledge, extended through TPACK frameworks (Koehler & Mishra, 2009), supported by Darling-Hammond (2000) research on program coherence, and informed by situated cognition theory (Lave & Wenger, 1991), this transformation represents not radical departure but rather thoughtful evolution of established teacher education principles for contemporary contexts.

Research demonstrates that when teacher educators model responsible AI use, ground learning in authentic pedagogical tasks consistent with situated learning theory, develop multi-dimensional competencies grounded in TPACK, and attend to both technical and ethical dimensions of AI integration, preservice teachers emerge prepared to integrate AI responsibly in future classrooms. The choice facing teacher education is not whether AI will shape future teaching—it inevitably will—but rather whether teacher educators will deliberately prepare teachers for this reality through coherent, well-grounded programs or leave them unprepared. Systematic transformation toward comprehensive AI literacy development, grounded in foundational teacher education theory while extending frameworks to address contemporary realities, represents the professional responsibility of teacher education for the contemporary era.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Cochran-Smith, M. (2005). The new teacher education: For better or for worse? *Educational Researcher*, *34*(7), 3–17. https://doi.org/10.3102/0013189X034007003

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Lawrence Erlbaum.

Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, *51*(3), 166–173. https://doi.org/10.1177/0022487100051003002

Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, *61*(1-2), 35–47. https://doi.org/10.1177/0022487109348024

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, *103*(6), 1013–1055. https://doi.org/10.1177/016146810110300601

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3), 381–391. https://doi.org/10.1080/135406002100000512

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, *90*(7), 495–500. https://doi.org/10.1177/003172170909000709

Hansen, D. T. (2001). *Exploring the moral heart of teaching: Toward a teacher's creed*. Teachers College Press.

Joyce, B., & Showers, B. (1980). Improving in-service training: The messages of research. *Educational Leadership*, *37*(5), 379–385.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development (3rd ed.)*. Association for Supervision and Curriculum Development.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.

Merriam, S. B., & Bierema, L. L. (2014). *Adult learning: Linking theory and practice*. Jossey-Bass.

Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult and Continuing Education*, *74*, 5–12.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Moorhouse, B. L., Wan, Y., Wu, C., Kohnke, L., Ho, T. Y., & Kwong, T. (2024). Developing language teachers' professional generative AI competence: An intervention study in an initial language teacher education course. *System*, *125*, 103399. https://doi.org/10.1016/j.system.2024.103399

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

O'Neill, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishers.

Piaget, J. (1954). *The construction of reality in the child*. Basic Books.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Selwyn, N. (2019). *Should robots replace teachers?* AI and the future of education. Polity Press.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.1177/002205741319300302

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22.

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education, 23*(1), 1–23. Society for Information Technology & Teacher Education, Waynesville, NC USA.

von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese, 80*(1), 121–140. https://doi.org/10.1007/BF00869951

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

Wen, W., & Wen, H. (2024). Bridging generative AI technology and teacher education: Understanding preservice teachers' processes of unit design with ChatGPT. *Contemporary Issues in Technology and Teacher Education*, *24*(4), 582–611. https://www.learntechlib.org/primary/p/223935/

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Wideen, M. F., Mayer-Smith, J., & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research*, *68*(2), 130–178. https://doi.org/10.3102/00346543068002130

Wiggins, G., & McTighe, J. (2005). *Understanding by design (2nd ed.)*. Association for Supervision and Curriculum Development.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Application of Machine Learning Techniques to Support Professional Track Selection in Medical Education

Nurgul Abutalipova [1*], Gulnaz Nabiyeva [2], Bakytgul Kalenova [3], Zhasulan Orazbekov [4]

[1] Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[2] Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan
[3] Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan
[4] Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
* Corresponding author: Nurgul Abutalipova (icemstoffice@gmail.com)

Article Info

Article History
Received:
18 December 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Machine learning
Medical education
Career prediction
Nlp
Smote

Abstract

Specialization assignment in medical universities commonly relies on GPA ranking, a practice that overlooks individual talent, motivation, and career aptitude. This study develops and evaluates machine learning (ML) models for predicting professional trajectories of General Medicine students based on aptitude survey data. A dataset of 518 fifth-year students from Asfendiyarov Kazakh National Medical University was collected via a Gallup-based survey. Natural language processing (NLP) preprocessing, SMOTE class balancing, and 5-fold cross-validation were applied. Eight ML and deep learning algorithms (SVM, LSTM, MLP, LR, CNN, NB, KNN, DT) were compared using accuracy, recall, and F1-score. Results: SVM and LSTM achieved the highest accuracy of 88.64%. SMOTE substantially improved prediction for minority classes. ML-based aptitude modelling offers a viable, evidence-based complement to GPA-only specialisation systems, with direct applicability in educational counselling platforms.

Citation: Abutalipova, N., Nabiyeva, G., Kalenova, B., & Orazbekov, Z. (2026). Application of Machine Learning Techniques to Support Professional Track Selection in Medical Education. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 129–138), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

Specialization choice in medicine is a high-stakes decision that directly affects a physician's professional effectiveness, job satisfaction, and ultimately patient outcomes. Research consistently shows that physicians placed in specializations misaligned with their aptitude and motivation are more susceptible to burnout and early career attrition. Despite this evidence, many medical universities, including those in Kazakhstan, continue to assign students to specializations primarily on the basis of grade point average (GPA) — a metric that captures academic performance but not personality, talent, or intrinsic motivation.

The Global Health Workforce Alliance has identified the uneven distribution of medical personnel as a critical public health challenge of the twenty-first century (World Health Organization, 2016). Balancing physician supply across specializations and geographic regions remains an ongoing concern even in high-income countries (Schweizerische Ärztezeitung, 2022). In Kazakhstan, shortages of general practitioners (GPs) in rural areas coexist with concentration of surgical and narrow specialists in major cities, partly reflecting a GPA-driven allocation system in which surgery attracts the highest-ranked students and GP tracks receive those with the lowest scores — irrespective of individual suitability (Querido et al., 2016).

Existing research has focused predominantly on identifying factors that influence career preference rather than on building predictive models (Scott et al., 2011). Classical statistical approaches such as logistic regression have been applied, but the potential of machine learning (ML) for predicting medical students' career trajectories remains underexplored (Petek Šter et al., 2015). This study addresses that gap by systematically evaluating seven ML and deep learning classifiers on an aptitude-based dataset collected from final-year medical students.

The survey instrument used in this study was developed by Tanu AI on the basis of the Gallup StrengthsFinder framework, which identifies natural talent themes and professional aptitude. The study therefore positions ML-based aptitude modelling as a data-driven complement — not a replacement — to existing GPA-based allocation, with the aim of improving the quality and sustainability of specialisation decisions in medical education.

The specific objectives of this study were:

- to collect and preprocess aptitude survey data from 518 fifth-year medical students;
- to address class imbalance using the Synthetic Minority Over-sampling Technique (SMOTE);
- to train and compare multiple ML and deep learning algorithms;
- to assess the best-performing model's suitability for integration into a career counselling platform.

Literature Review

Career aptitude prediction has attracted growing attention across educational data mining and ML research. Ade and Deshmukh (Ade & Deshmukh, 2015) achieved 89.6% accuracy by combining Naive Bayes, K-Star, and SVM classifiers using an ensemble voting approach, concluding that integrating student interests, talent, and growth

potential substantially improves specialisation guidance. Lent et al. (Lent et al., 2008) examined the role of self-efficacy, outcome expectations, and social support in career choice among 1,208 students, finding that non-academic personal factors exert significant influence on professional trajectory.

Rangnekar et al. (Rangnekar et al., 2012) proposed a career mapping system based on competency profiles across science, mathematics, language, and logical reasoning, testing five ML algorithms (KNN, SGD, LR, Random Forest, AdaBoost) on 2,000 samples. KNN and SGD achieved the highest predictive accuracy for career path classification. A summary of prior studies is presented in Table

The content summarized as Knowledge extraction process is presented in Figure 1.

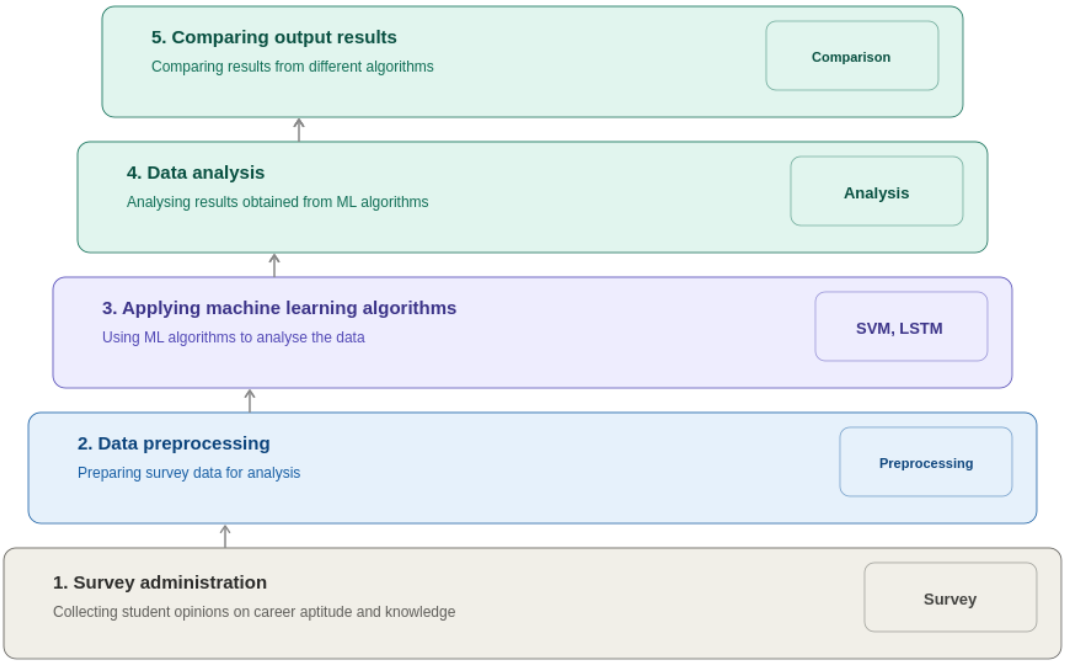


Figure 1. Knowledge Extraction Process

Mandalapu and Gong (Mandalapu & Gong, 2022) reported 98.83% accuracy using gradient boosted trees (GBT), while Kabakus and Senturk (Kabakus & Senturk, 2020) reached 100% using Bayesian Networks and J48 on academic achievement data. Sobnath et al. (Sobnath et al., 2020) obtained 96% accuracy with logistic regression and decision tree classifiers. Almarabeh (Almarabeh, 2017) evaluated multiple data mining classifiers for student performance analysis, providing the confusion-matrix-based evaluation framework applied in the present study.

Importantly, most prior work has targeted general or vocational education settings; very few studies have applied ML to career-track prediction within medical education specifically, where class distributions are highly skewed and the stakes of misclassification are particularly high. This study contributes to filling that gap.

The content summarized as Comparison with prior studies is presented in Table 1.

Table 1. Comparison with Prior Studies

Author(s)	Algorithm	Best accuracy	Dataset	Domain
Ade & Deshmukh	NB, K-Star, SVM	89.6%	Student records	General education
Mandalapu & Gong	GBT	98.83%	Career data	Career guidance
Kabakus & Senturk	BN, J48	100%	Academic data	Student achievement
Sobnath et al.	LR, DT	96%	Career data	Career guidance
Present study	**SVM, LSTM, MLP, et al.**	**88.64%**	**518 medical students**	**Medical education**

Methods

Participants and Ethical Approval

Data were collected from 518 fifth-year students at Asfendiyarov Kazakh National Medical University (Almaty, Kazakhstan) during November–December 2023. Of 580 students invited, 518 participated (response rate: 89.3%); 39 declined citing time constraints and 23 reported no interest in the topic. The sample comprised 67% female and 33% male students, with a mean GPA of 3.85 ± 0.42. The study was approved by the university ethics committee, and all participants provided written informed consent. Participant anonymity and data confidentiality were maintained throughout. Specialisation distribution reflected GPA-based university allocation: Surgery (n = 269; 51.9%), Internal Medicine (n = 190; 36.7%), Obstetrics & Gynaecology (n = 41; 7.9%), and General Practice/GP (n = 18; 3.5%).

Survey Instrument

Aptitude data were collected via a structured online questionnaire (Google Forms) developed by Tanu AI on the basis of the Gallup StrengthsFinder framework. The instrument comprised eight thematic blocks: current academic year and status; preferred future specialisation; intended subspecialty for research; planned residency track; key foundational subjects; critical professional competencies; prior clinical or volunteer experience; and perceived influence of experience on specialisation preference. Responses were exported to Microsoft Excel for subsequent processing.

Data Preprocessing and NLP Pipeline

The preprocessing pipeline comprised five sequential stages. First, raw survey responses were collected via Google Forms. Second, missing values were imputed using mode substitution and contradictory responses were

removed (text cleaning). Third, categorical variables — including specialisation labels and competency preferences — were encoded as numeric indices using Scikit-learn's LabelEncoder (tokenization and encoding). Fourth, SMOTE was applied to the training set to address pronounced class imbalance (Surgery: 51.9%, GP: 3.5%); the test set was kept in its original distribution to ensure unbiased evaluation. Fifth, all features were standardized using StandardScaler (zero mean, unit variance) prior to model training.

The content summarized as Data processing pipeline is presented in Figure 2.

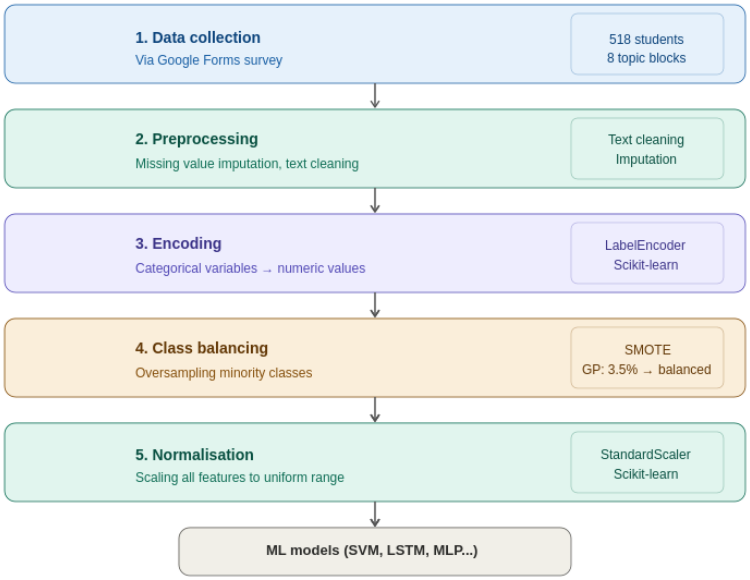


Figure 2. Data Processing Pipeline

The content summarized as LabelEncoder implementation in Python is presented in Figure 3.

```python
from sklearn.preprocessing importLabelEncoder
importpandas as pd

# Load data exported from Google Excel
df = pd .read_excel( 'kaznmu_students_data.xlsx'      )

# Initialise the encoder
le= LabelEncoder()

# Encode target column 'Future_Specialty' as numeric
df['Specialty_Encoded' ]= le.fit_transform(df[Future_Specialty'])

# Verify mapping: e.g. Surgery→0, Therapy→1
```

Figure 3. LabelEncoder Implementation in Python

Machine Learning Models

Seven classifiers were trained and evaluated: Support Vector Machine (SVM, RBF kernel, C = 1.0), Logistic Regression (LR, lbfgs solver, max_iter = 100), K-Nearest Neighbours (KNN, k = 3), Naive Bayes (NB, var_smoothing = 1e−9), Multilayer Perceptron (MLP, 4 hidden layers, 50 epochs), Long Short-Term Memory network (LSTM, 64 units, 50 epochs), and Convolutional Neural Network (CNN, 32 filters, dropout = 0.5). Hyperparameters were set according to Scikit-learn defaults and established recommendations in the literature (Table 2).

Table 2. Model Hyperparameters

Model	Key parameter(s)	Value	Rationale
SVM	kernel, C	RBF, 1.0	Scikit-learn default
LR	solver, max_iter	lbfgs, 100	Multiclass classification
KNN	n_neighbors	k = 3	Small dataset
NB	var_smoothing	1e−9	Scikit-learn default
MLP	hidden_layers, epochs	4 layers, 50	Empirical selection
LSTM	units, epochs	64, 50	NLP standard
CNN	filters, dropout	32, 0.5	Overfitting prevention

Evaluation

Model performance was assessed using accuracy, recall, and F1-score derived from the confusion matrix (Almarabeh, 2017). The dataset was split 80% training / 20% test. To account for the small sample size (n = 518), 5-fold cross-validation was applied to estimate generalisation performance. A majority-class baseline (predicting Surgery for all instances) yielded 51.9% accuracy, providing a lower-bound reference. Statistical significance of pairwise differences was assessed using McNemar's test.

Performance metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F\text{-}measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

Results

All seven classifiers substantially outperformed the 51.9% majority-class baseline, confirming that the models captured meaningful patterns beyond the dominant class. SVM achieved the highest accuracy - 88.64%, recall - 88.64%, and F1-score - 88.64%, followed closely by LSTM accuracy: 88.64%; F1: 88.01%. Decision Tree and KNN yielded the lowest performance - 77.27%. Full results are reported in Table 3.

Table 3. Model Performance on the Held-Out Test Set

Model	Accuracy (%)	Recall (%)	F1-score (%)	Rating
SVM	**88.64**	**88.64**	**88.64**	**Best**
LSTM	88.64	87.89	88.01	High
MLP	85.23	84.97	85.10	High
LR	83.41	83.12	83.26	High
CNN	79.55	79.23	79.38	Moderate
NB	79.55	79.10	79.32	Moderate
KNN	77.27	76.98	77.12	Low
DT	77.27	77.01	77.14	Low

SMOTE substantially improved minority-class prediction: GP track F1-score increased from 45.3% to 71.8%, and Obstetrics & Gynaecology from 58.2% to 76.4%, confirming the efficacy of synthetic oversampling for imbalanced medical datasets.

Cross-validation confirmed model stability: SVM mean CV accuracy 87.9% ± 1.2%, LSTM 87.5% ± 1.8%. The low standard deviations indicate consistent performance across data folds. The difference between SVM and LSTM was not statistically significant ($p = 0.73$), whereas SVM significantly outperformed DT ($p = 0.003$), ruling out a chance advantage. SVM's superior performance is attributable to the RBF kernel's ability to construct non-linear decision boundaries in the feature space, effectively capturing complex aptitude patterns within a relatively small dataset. LSTM's competitive accuracy reflects its strength in modelling sequential dependencies in student response patterns — the ordered survey blocks provided a natural sequential structure amenable to recurrent architectures. MLP's performance (85.23%) was constrained by overfitting risk in the $n = 518$ setting, mitigated by Dropout regularization. CNN (79.55%) and NB (79.55%) performed moderately, with NB limited by its feature-independence assumption.

Discussion

The present study demonstrates that ML classifiers trained on aptitude survey data can predict medical specialization preferences with accuracy substantially above chance (51.9% baseline) and comparable to prior career-prediction studies in general education contexts. SVM's 88.64% accuracy is close to the 89.6% reported by Ade and Deshmukh and considerably below the 96% reported by Sobnath et al. using simpler tabular career datasets; the gap likely reflects the greater complexity and higher class imbalance of the medical-education setting.

From a practical standpoint, the SVM model is particularly attractive for deployment: it trains in seconds on $n = 518$ samples, produces interpretable support vectors, and generalizes stably across cross-validation folds. Integration into a student-facing counselling portal — in which SVM predictions are presented alongside GPA-based recommendations — could provide advisers with a richer, dual-evidence basis for guiding specialization decisions, without discarding the existing institutional framework.

SMOTE's impact on minority classes (GP and Obstetrics) is especially noteworthy from a health-workforce perspective. If predictive models systematically under-serve students suited to under-subscribed tracks, they would replicate and potentially amplify existing allocation biases. The post-SMOTE improvements (GP F1: 45.3% → 71.8%) suggest that aptitude-informed guidance could, over time, help correct the structural imbalance between supply and demand in primary care and regional medicine.

Several limitations must be acknowledged. First, the dataset was collected from a single university, which may limit generalizability to other institutions or national contexts with different GPA allocation thresholds or specialization structures. Second, the sample size, while adequate for the present analysis, is modest; larger multi-site studies are needed to validate findings. Third, the survey instrument has not yet undergone formal psychometric validation (e.g., Cronbach's alpha estimation), which would strengthen confidence in the construct validity of the aptitude measures. Fourth, the 2023 data collection period should be replicated longitudinally to assess whether aptitude-based predictions correspond to actual career satisfaction and retention.

Future research should: (1) collect data from multiple universities across Kazakhstan and neighboring countries; (2) expand the sample to $n \geq 1000$ and apply more robust evaluation methods such as stratified k-fold cross-validation; (3) conduct formal psychometric validation of the Tanu AI survey instrument; (4) compare ML predictions against GPA-based outcomes longitudinally; and (5) explore explainable AI (XAI) methods to improve transparency of individual predictions for students and advisers.

Conclusion

This study evaluated seven ML algorithms for predicting professional track preferences of 518 final-year General Medicine students, using aptitude survey data processed through an NLP pipeline. SVM and LSTM achieved the highest accuracy (88.64%), substantially outperforming the GPA-only baseline. SMOTE effectively addressed class imbalance, improving minority-class prediction by up to 26 percentage points. Five-fold cross-validation

confirmed the stability and generalisability of the results.

The findings establish an evidence base for integrating ML-driven aptitude modelling into medical career counselling, complementing — rather than replacing — existing GPA-based allocation systems. Such a hybrid approach has the potential to reduce career mismatch, mitigate burnout risk, and contribute to a more equitable distribution of medical professionals across specialisations and regions. Validation at scale across multiple institutions remains the critical next step.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Ade, R., & Deshmukh, P. R. (2015). Methods for incremental learning: A survey. *International Journal of Data Mining & Knowledge Management Process*, *5*(4), 119–135. https://doi.org/10.5121/ijdkp.2015.5409

Almarabeh, H. (2017). Analysis of students' performance by using different data mining classifiers. *International Journal of Modern Education and Computer Science*, *9*(8), 9–15. https://doi.org/10.5815/ijmecs.2017.08.02

Kabakus, A. T., & Senturk, U. (2020). Examining the performances of machine learning algorithms for the classification and prediction of students' academic achievements. *Journal of Educational Technology & Online Learning*, *3*(2), 1–16. https://doi.org/10.31681/jetol.706095

Lent, R. W., Sheu, H. B., Singley, D., Castiriano, J. A., & Lopez, A. M. (2008). Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of*

Vocational Behavior, *73*(2), 328–335. https://doi.org/10.1016/j.jvb.2008.07.005

Mandalapu, V., & Gong, Y. (2022). Predicting career paths using gradient boosting trees. *International Journal of Advanced Computer Science and Applications*, *13*(5), 412–419. https://doi.org/10.14569/IJACSA.2022.0130548

Petek Šter, M., Švab, I., & Šter, B. (2015). Prediction of intended career choice in family medicine using artificial neural networks. *European Journal of General Practice*, *21*(1), 63–69. https://doi.org/10.3101/13814788.2014.905615

Querido, S. J., Vergouw, D., Wigersma, L., Batenburg, R. S., De Rond, M. E. J., & Ten Cate, O. T. J. (2016). Dynamics of career choice among students in undergraduate medical courses. *Medical Teacher*, *38*(1), 18–29. https://doi.org/10.3109/0142159X.2015.1074990

Rangnekar, S., Bhola, S. S., & Gupta, R. (2012). Career path of students: A survey of professional students. *International Journal of Research in Commerce & Management*, *3*(6), 115–122.

Schweizerische Ärztezeitung (2022). La qualité des soins dépend d'une relève suffisante et qualifiée. *Bulletin des médecins suisses*, *103*(11), 350–353. https://doi.org/10.4414/bms.2022.20632

Scott, I., Gowans, M., Wright, B., Brenneis, F., Banner, S., & Boone, J. (2011). Determinants of choosing a career in family medicine. *CMAJ*, *183*(1), E1–E8. https://doi.org/10.1503/cmaj.100293

Sobnath, D., Ali, N., Jaafar, J., Alsaawy, Y., & Khusro, S. (2020). Recommendation of suitable machine learning approach for career path prediction. *PLOS ONE*, *15*(12), e0243432. https://doi.org/10.1371/journal.pone.0243432

World Health Organization (2016). *Global strategy on human resources for health: Workforce 2030*. https://www.who.int/publications/i/item/9789241511131

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

A Culturally Responsive Android Learning Media to Promote Differentiated Deep Learning and Cultural Literacy

Firdaus [1 *], Syahruddin [2], Nukhrawi Nawir [3], Latang [4]

[1] Makassar State University, Indonesia, 0009-0003-6902-1504
[2] Makassar State University, Indonesia, 0000-0002-1667-2405
[3] Makassar State University, Indonesia, 0000-0003-4085-1371
[4] Makassar State University, Indonesia, 0009-0008-4333-076X
* Corresponding author: Firdaus (icemstoffice@gmail.com)

Article Info

Article History
Received:
18 December 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Deep learning
Android learning media
Cultural literacy
Learning styles
Differentiated instruction

Abstract

The transformation of Indonesian education toward Golden Indonesia 2045 requires strengthening human resources and digitalization while maintaining cultural values and deep learning aligned with students' diverse needs. However, the use of diagnostic assessments based on learning types remains limited, teachers face time constraints, and literacy achievement—particularly in vocational schools—remains moderate. In physical education (PJOK), integrating ethnosport, such as traditional games, offers potential to enhance learning experiences and cultural literacy. Preliminary findings in vocational schools across Bantaeng Regency indicate high smartphone use for gaming, highlighting the need for adaptive Android-based learning media. This study aims to develop and evaluate the effectiveness of Android-based PJOK learning media integrating ethnosport to support differentiated learning and improve cultural literacy. The study employed a Research and Development approach using the ADDIE model across 11 vocational schools (September–December 2025), including expert validation, small and large group trials, and a one-group pretest–posttest design analyzed using paired sample t-test (SPSS 23). Results show a very high level of teacher need (84.42%) and strong student demand (70.27%), with readiness for digital learning. The developed SportEdunesia media, created using MIT App Inventor, includes differentiated materials, cultural literacy quizzes, and traditional games assessment. It achieved very high validity (Aiken's V=0.935) and demonstrated strong practicality. Effectiveness testing revealed a significant improvement in cultural literacy scores from 4.90 to 7.13 (p=0.000), with a large effect size (Cohen's d=0.93). These findings indicate that Android-based PJOK media integrating learning style differentiation and ethnosport is effective, meaningful, and relevant for enhancing cultural literacy in 21st-century learning contexts.

Citation: Firdaus, Syahruddin, Nawir, N., & Latang (2026). A Culturally Responsive Android Learning Media to Promote Differentiated Deep Learning and Cultural Literacy. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 139–147), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

Education, as a strategic axis in preparing the next generation to continue the future, needs to adapt and transform to prepare Indonesians for welcoming Golden Indonesia 2045. This is expected to result in the emergence of teachers who can translate the mission of education into learning activities that equip students for life—teachers who are innovative, critical, and use curricula that accommodate the diverse needs of learners from various sources.

Technological advancements in education are in line with the fourth goal of Indonesia's President (Asta Cita), which emphasizes strengthening human resource development, science, technology, education, health, sports achievement, gender equality, and the roles of women, youth, and persons with disabilities. As stated by Sabarudin & Purwadi (2024), education is a critical issue in the ASTA CITA program, which includes building integrated excellent schools in every regency and renovating schools in need of improvement to strengthen science, technology, and digitalization. Education is key to making Indonesia an advanced nation by improving individual, community, and national quality of life.

Therefore, efforts to develop educational services through technological innovation integrated with cultural values —as a national identity—are of paramount importance. Such development should align with deep learning policies that can be fostered through learning based on students' learning styles. The learning process should not solely focus on mastery of material but also aim at character development, deepening understanding, and creating meaningful and enjoyable learning experiences for students. Hence, the design of learning experiences needs to consistently prioritize the evolving and dynamic needs, potentials, and diversity of students.

Literature review findings (Firdaus, 2025) show that only about 20% of studies report learning activities that utilize the results of diagnostic assessments, especially students' learning styles, in instructional implementation. If this continues, there is potential for malpractice in learning. The research further reveals that limited time is the main obstacle for teachers; hence, it is recommended to develop learning media that help teachers deliver adaptive learning by systematically integrating diagnostic assessment results into differentiated instructional strategies.

Another classic problem in our education is the issue of literacy. The National Education Report indicates that literacy scores at all levels of education remain in the medium category. Specifically, the literacy rate of vocational high school (SMK) students this year shows that only 49.26% have literacy competency above the minimum, a decrease of 4.59% from 2021 (53.85%) (Ministry of Education, Culture, and Research, 2023).

Literacy is not limited to one's ability to read and write. As Asraf (2021) states, literacy has a broader definition: it is not just about reading, but also about obtaining information from a variety of sources, including digital ones.

Cultural literacy is one of the essential 21st-century competencies. This literacy needs to be instilled early within families, schools, and communities because it plays a role in preserving national culture and shaping Indonesia's identity in the global community. As Pratiwi and Asyarotin (2019) noted, the implementation of cultural and civic

literacy can be a solution to disinformation among millennials, helping them avoid and tackle this phenomenon, and ensuring that millennials continue to love and preserve culture as part of Indonesia's national identity.

As one of the subjects, Physical, Health, and Sports Education (PJOK) should contribute to realizing the goals of National Education. In its implementation, PJOK should not be limited to physical activities, but carried out by integrating Indonesian cultural roots. This aligns with (Kemendikbud, 2024), which states that Indonesian national education fosters progress by adhering to and considering the Indonesian context, especially cultural roots.

The integration of Ethnosport, such as traditional games, in PJOK learning can improve physical fitness and foster cultural literacy through an understanding of cultural diversity. This is supported by Farida Noor Isnaini (2021), who argues that physical education not only enhances the physical and mental strength of students, but also boosts their literacy and cognitive abilities.

A preliminary study in vocational schools (SMK) in Bantaeng Regency found high use of smartphones among students for online gaming. Therefore, Android-based learning media is needed as a counterbalance. Interviews with PJOK teachers also confirm that although such media can boost learning motivation, the effectiveness of material delivery must be adapted to students' individual learning styles.

These arguments and empirical facts provide a strong rationale for researchers to conduct research on the Development of Android-Based Learning Media Assisted by Ethnosport to Support Differentiated Deep Learning for Enhancing Cultural Literacy.

Several studies on Android-based learning media have already been conducted, but they have yet to adapt to students' learning styles. One such study (Indra et al., 2024) found that the use of interactive ThingLink-based learning media can optimize students' concentration, including those with inattentive ADHD, but the developed media focused only on learning concentration. Another study (Lestari et al., 2024) suggests that the InnovARte Learning module can be expanded with more interactive and varied content, and equipped with new features such as more complex AR simulations and a variety of gamification elements. Previous studies have mainly focused on teaching modules and have yet to accommodate differences in students' learning styles. Therefore, this research will develop Android-based learning media with features tailored to each student's learning style needs.

The research problems formulated are: How is the design of Android-based Physical, Health, and Sports Education (PJOK) learning media assisted by Ethnosport to support differentiated deep learning for enhancing cultural literacy, and how effective is this Android-based PJOK learning media assisted by Ethnosport in supporting differentiated deep learning for enhancing cultural literacy?

Method

This research is a Research and Development (R&D) study, adopting the ADDIE (Analysis, Design, Development, Implementation, and Evaluation) development model. The research site was 11 vocational high schools

(SMK) in Bantaeng Regency, South Sulawesi Province, Indonesia.

Research Timeline: This research was conducted from September 4 to December 4, 2025. The preliminary study (field study and literature review) and proposal development were conducted from July to August 2025. Treatment preparation, namely the preparation of learning media designs and cultural literacy instruments, was conducted from September 2025. Preparation for the experimental research included sample selection, treatment administration, and data analysis. Target group: Grade XI students, semester III.

Media Validation, Evaluation, and Revision: 1) Expert Judgment, a design validation activity conducted by product and instrument experts. The designed product will then be reviewed by experts to describe and/or explain the shortcomings of the developed product for subsequent improvement. 2) Product Trial, after the Android-based PJOK learning media is revised by experts, then displayed and introduced, then a small group trial is conducted, the results of the small group trial become the basis for improvement, before conducting a field trial on a larger number of respondents. 3) Usage Trial, after revisions from the small group trial, then a trial is conducted with a larger number of subjects, the conclusion from the results of this large group trial is the final basis for improving and perfecting the Android-based PJOK learning media before the product is implemented. 4) Product Implementation, the next step is an effectiveness trial to find out; a) Whether this Android-based PJOK learning media has been implemented well, b) Whether the Android-based PJOK learning media is effective in improving cultural literacy. Quantitative descriptive statistics are used to find out whether the media design has been implemented well and correctly, as well as to test the effectiveness of the product through a pre-experimental research design in the form of The Two group pretest-posttest design. Meanwhile, the cultural literacy test is a test used to measure the level of cultural literacy after going through a validation process

The content summarized as One group pretest-posttest experimental research design is presented in Figure 1.

$$O_1 \quad \text{x} \quad O_2$$

$O_1 \quad \text{x} \quad O_2$

Figure 1. One Group Pretest-Posttest Experimental Research Design

The procedures implemented in this study were as follows: (1) determining a single group of research subjects; (2) administering a pre-test (O1) to measure students' initial abilities; (3) providing the treatment (P) in the form of implementing the developed learning media; (4) administering a post-test (O2) to assess students' final abilities after the treatment; (5) calculating and comparing the mean scores of the pre-test and post-test; and (6) analyzing

the difference between the two means using a statistical test (paired sample t-test) to determine whether there is a significant effect of the use of the learning media. Data were analyzed using the t-test procedure with the assistance of SPSS 23 software

Results

This research aims to develop an Android-based SportEdunesia learning media for Physical Education (PJOK) integrated with traditional games and a differentiated learning approach to improve cultural literacy in 10th-grade vocational high school students. The research results are presented based on the development stages, including needs analysis, design, development, implementation, and product evaluation.

During the needs analysis stage, the need for Android-based learning media was found to be very high from the teachers' perspective. This is reflected in the 84.42% need percentage, indicating that teachers view digital media that can support more innovative PJOK learning as crucial. Teachers also emphasized the importance of media that can accommodate students' different learning styles, integrate traditional games, and provide practical and systematic user guides. These findings indicate a gap between the demands of curriculum-based learning that prioritizes in-depth learning and the relatively conventional learning practices in the field. Therefore, developing learning media is not only a technical requirement but also a strategic pedagogical need.

From the students' perspective, the analysis shows that the majority of students already have Android devices and internet access, making them infrastructurally ready to support digital-based learning. Students' need for Android-based learning media fell into the "need" category, with a percentage of 70.27%. Furthermore, the majority of students indicated a preference for interactive media, allowing freedom in choosing learning methods, and integrating cultural elements through traditional games. This indicates that students need not only digital media but also media that adapts to their learning characteristics and preferences, and is able to connect learning materials to relevant cultural contexts.

Based on the results of the needs analysis, an Android-based learning media called *SportEduNesia* was developed to accommodate differentiated learning styles and enhance students' cultural literacy. The media was created using the MIT App Inventor platform and features differentiated instructional materials, interactive quizzes based on cultural literacy, and an assessment system integrated with traditional games. Expert validation results indicated that the developed media achieved a very high level of validity, with an Aiken's V score of 0.935. This demonstrates that the media meets feasibility criteria in terms of content, language, presentation, and instructional usefulness. The validation process also generated several suggestions for improvement, including the addition of login features, learning style assessments, instructional modules, and discussion features. These revisions strengthened the quality of the media both technically and pedagogically.

During the implementation phase, trial results showed that the developed learning media possessed a very high

level of practicality. Students were able to use the application independently and responded positively to its attractiveness and ease of use. Teachers also provided positive feedback, noting that the media facilitated the systematic integration of traditional games into learning and increased student engagement. This indicates that the media is not only theoretically feasible but also practically applicable in real classroom settings.

The effectiveness of the learning media was measured by comparing students' cultural literacy pretest and posttest scores. The analysis revealed an increase in the mean score from 4.90 in the pretest to 7.13 in the posttest. Statistical testing using a paired sample t-test showed a significance value of 0.000 ($p < 0.05$), indicating a significant difference between pre- and post-intervention results. Furthermore, the effect size (Cohen's d) of 0.93 suggests a large effect, indicating that the media not only achieves statistical effectiveness but also produces a strong practical impact on improving students' cultural literacy.

Overall, the findings demonstrate that the development of Android-based PJOK learning media integrating differentiated instruction and traditional games is a strategic approach to improving learning quality. *SportEduNesia* functions not only as an instructional aid but also as a medium that connects learning content with students' cultural context, thereby creating more meaningful learning experiences. Therefore, this media can serve as an alternative solution to address the challenges of 21st-century learning, which demand the integration of technology, differentiated instruction, and the strengthening of cultural literacy.

Discussion

This discussion focuses on the development of the Android-based PJOK learning media, *SportEduNesia*, to enhance students' cultural literacy through the aspects of validity, practicality, and effectiveness. The validity test results indicate that the media falls within the "very valid" category, reflecting its alignment with the curriculum, student characteristics, and principles of digital instructional design. This condition is consistent with Nurmalisa et al. (2023), who emphasize that adapting instruction to students' needs highlights the importance of developing digital media capable of accommodating diverse learning style preferences.

The instructional materials were systematically organized by integrating cognitive, affective, and psychomotor aspects, supported by engaging visual design and communicative language. Revisions based on expert feedback strengthened the quality of the media without altering the primary focus of the study. This finding is in line with Balakrishnan and Gan (2016), who argue that students with different learning styles adopt different learning approaches.

The effectiveness of the media is evidenced by statistical test results showing a significant improvement in students' cultural literacy after using *SportEduNesia*. These findings confirm that the media functions not merely as an instructional aid, but as a pedagogical instrument capable of fostering conceptual understanding while promoting the internalization of cultural values in PJOK learning. Theoretically, this result aligns with Hattie and O'Leary (2025), who assert that individuals taught through modalities or styles aligned with their preferences tend

to achieve better learning outcomes than those taught through less preferred modalities. The differentiated learning features embedded in *SportEduNesia* enable students to access materials and traditional game activities according to their learning characteristics, thereby enhancing both cognitive and affective engagement.

Furthermore, these findings reinforce the study by Yuliansih et al. (2021), which states that instructional media have a significant partial effect on student learning outcomes. The integration of Android-based technology with local cultural contexts demonstrates that systematically designed media can make a substantial contribution to improving cultural literacy achievement. In synthesis, the effectiveness of *SportEduNesia* arises from the combination of adaptive media design responsive to learning styles and the instructional power of media as a significant determinant of learning outcomes, positioning it as an innovative, contextual, and meaningful solution in PJOK learning

Conclusion

The conclusions of this study are formulated based on two main research questions, namely the design and effectiveness of the developed learning media. First, from the design perspective, this study successfully produced an Android-based PJOK learning media, *SportEduNesia*, developed using the ADDIE model by integrating an ethnosport approach and differentiated learning. The media includes adaptive instructional materials based on learning styles, interactive cultural literacy quizzes, and contextual traditional games. Expert validation results indicate that the media falls into the "very valid" category (Aiken's V = 0.935), confirming its feasibility as a learning tool that meets content and design standards while aligning with pedagogical needs and student characteristics.

Second, from the effectiveness perspective, empirical testing demonstrates that the use of *SportEduNesia* significantly improves students' cultural literacy. This is evidenced by a significant difference between pretest and posttest results ($p < 0.05$). These findings indicate that the media functions not only as an instructional aid but also as a tool that promotes deep understanding and the internalization of cultural values through contextual and adaptive learning experiences.

In conclusion, this study affirms that the development of Android-based learning media integrating differentiated instruction and local cultural elements represents an innovative and effective solution for enhancing the quality of PJOK learning and students' cultural literacy.

Recommendations

Future research is recommended to further develop the *SportEduNesia* media so that it can be used on the iOS operating system, thereby expanding its user reach. In addition, it is necessary to incorporate digital communication features that enable interaction between teachers and students as well as among students, in order to support collaborative learning and enhance student engagement.

Content enrichment is also essential by adding a wider variety of instructional materials and traditional games that are more diverse and contextually aligned with local cultures across different regions. Furthermore, expanding the research scope to include primary and higher education levels is necessary to examine the consistency of the media's effectiveness across diverse learner characteristics and to strengthen the generalizability of the findings.

Statements and Declarations

Acknowledgments and Notes: The author extends sincere gratitude to all parties who have provided support in the implementation of this research. Special appreciation is addressed to the promoter and co-promoter, the expert validators, as well as the teachers and students who participated in the trial process and offered constructive feedback for the refinement of the developed learning media. It is hoped that the findings of this study will make a meaningful contribution to the development of innovative, contextual, and impactful PJOK learning.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: The author also expresses profound appreciation to the Beasiswa Pendidikan Indonesia (BPI), Ministry of Primary and Secondary Education of the Republic of Indonesia (Kemdikdasmen), through PPAPT (Center for Higher Education Financing and Assessment), under the auspices of the Ministry of Higher Education, Science, and Technology (Kemdiktisaintek) of the Republic of Indonesia, as the sponsoring institution that provided financial support, enabling this research to be conducted successfully.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Asraf, A. (2021). The Implementation of Interview Techniques to Improve Students' Speaking Skills in Conditional Sentence Materials. *Indonesian Journal of Educational Development*, *1*(4), 586–593. https://doi.org/10.5281/zenodo.4559856

Balakrishnan, V., & Gan, C. L. (2016). Students' learning styles and their effects on the use of social media technology for learning. *Telematics and Informatics*, *33*(3). https://doi.org/10.1016/j.tele.2015.12.004

Farida Noor Isnaini (2021). The Importance of Physical Education in Enhancing Literacy and Cognitive Skills of Elementary School Students in the Era of the Industrial Revolution 4.0. *Jurnal Syntax Transformation*, *2*(01), 75–85. https://doi.org/10.46799/jst.v2i1.202

Firdaus, S. N. N. L. (2025). A Pedagogical Critique of Learning Practices That Ignore Diagnostic Assessment Results for Student Learning Types. *PPSDP International Journal of Education*, *4*(2), 730–748. https://doi.org/10.59175/pijed.v4i2.787

Hattie, J., & O'Leary, T. (2025). Learning Styles, Preferences, or Strategies? An Explanation for the Resurgence of Styles Across Many Meta-analyses. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10002-w

Indra, S., Maryani, N., Amril, L. O., Latifah, Z. K., & Adri, H. T. (2024). The Influence of Interactive Learning Media Based on the Thinglink Application on Students' Learning Concentration. *TADBIR MUWAHHID, 8*(2), 259–267. https://doi.org/10.30997/jtm.v8i2.15861

Kemendikbud (2024). *Curriculum for Early Childhood Education, Elementary Education, and Secondary Education*. Minister of Education, Culture, Research, and Technology Regulation Number 12 of 2024.

Lestari, W. S., Ulina, M., & Gaol, M. L. (2024). Innovarte Learning*:* Augmented Reality-Based Learning Media for Students with Disabilities Innovarte Learning: Augmented Reality-Based Learning Media for Students with Disabilities. *4*(140), 2084–2098.

Ministry of Education, Culture, Research, & Technology (2023). *Indonesia's Education Report Card 2023*. Freedom to Learn.

Nurmalisa, Y., Sunyono, S., Yulianti, D., & Sinaga, R. M. (2023). An Integrative Review: Application of Digital Learning Media to Developing Learning Styles Preference. *International Journal of Information and Education Technology*, *13*(1). https://doi.org/10.18178/ijiet.2023.13.1.1795

Pratiwi, A., & Asyarotin, E. N. K. (2019). Implementing Cultural and Civic Literacy as a Solution to Disinformation Among the Millennial Generation in Indonesia. *Journal of Information and Library Studies*, *7*(1), 65–80. https://doi.org/10.24198/jkip.v7i1.20066

Sabarudin, D., & Purwadi. (2024). Presidential Candidate Prabowo Subianto's Economy in the 2024-2029 Best Results Fast Program. *Indonesian National Journal*, *1*(1), 1–13.

Yuliansih, E., Arafat, Y., & Wahidy, A. (2021). The influence of learning media and learning interests on student learning outcomes. *JPGI (Indonesian Teacher Research Journal)*, *6*(2). https://doi.org/10.29210/021064jpgi0005

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Artificial Intelligence Integration into AS9100-Based Quality Management Systems: A Literature Review in the Context of Defense Technologies

Yasin Çıbuk [1], Feyzi Kaysi [2]

[1] Istanbul University – Cerrahpaşa, Vocational School of Technical Sciences, Turkey
[2] Istanbul University – Cerrahpaşa, Vocational School of Technical Sciences, Turkey, 0000-0001-6681-4574
* Corresponding author: Yasin Çıbuk (icemstoffice@gmail.com)

Article Info

Article History
Received:
6 January 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
AS9100 Rev D
AI-enabled quality management
Large language models
Defense industry
Vocational higher education

Abstract

This study addresses a critical gap at the intersection of aerospace and defense quality management and artificial intelligence, providing a theoretical foundation for a modular, AI-enabled quality management system tailored to AS9100 Rev D. Conducted within Turkey's defense and aerospace industry context, the study presents a comprehensive literature review organized around four thematic axes: quality management systems and digital transformation; artificial intelligence and risk management integration; large language models and document intelligence; and quality management in the defense and aerospace sector. The review shows that the sector-specific requirements introduced by AS9100 beyond ISO 9001—configuration management, traceability, counterfeit part control, and risk management—have been systematically overlooked in existing AI integration research. This gap appears across three dimensions: standard specificity, architectural coherence, and scalability. The study develops a theoretical framework to close this gap by positioning Quality 4.0, large language models, Retrieval-Augmented Generation (RAG), and cloud-based serverless architectures within Turkey's national defense industry ecosystem and standardization policy context. The findings also indicate an implication for vocational higher education, where prospective quality engineers are rarely prepared to evaluate or design AI-integrated QMS solutions compliant with AS9100.

Citation: Çıbuk, Y., & Kaysi, F. (2026). Artificial Intelligence Integration into AS9100-Based Quality Management Systems: A Literature Review in the Context of Defense Technologies. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 148–160), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

The defense, aviation, and space industry is a critical sector that inherently demands zero error tolerance, strict traceability requirements, and high security standards. Organizations operating in this sector must fully comply with the AS9100 standard—built upon the ISO 9001 foundation and incorporating sector-specific stringent requirements—in order to establish a sustainable quality infrastructure (Annaç Göv, 2018). Conventional quality management systems generally struggle to meet the intensive documentation, configuration management, and supply chain risk controls demanded by AS9100; manual management of processes leads to human-induced errors and significant efficiency losses. IAQG (2016a) obligates organizations under AS9100 Rev D to manage approximately one hundred additional requirements, placing a disproportionate burden on the quality departments of mid-sized defense suppliers. This compliance load is well documented in industry analyses, with small and medium-sized suppliers in particular being known to allocate disproportionate resources to meeting configuration management and traceability requirements (Dale, 2010).

Industry 4.0 and digital transformation dynamics now make new technological approaches indispensable for managing these complex quality processes. In Turkey, the reflections of this transformation are particularly evident in defense industry policy; the AI-focused academic programs and university-industry collaborations of the Presidency of Defence Industries (SSB) demonstrate that this need has been recognized at the institutional level (SSB/YÖK, 2024). On the international front, the integration of ISO-based quality management systems with information systems is gaining increasing importance, a trend that has triggered the emergence of next-generation AI management standards such as ISO/IEC 42001 (Gueorguiev, 2025). The transformative impact of AI applications on quality audits, process automation, document management, and data-driven decision support systems is moving quality management beyond a static ruleset toward a dynamic and intelligent structure (Minglana et al., 2021). The European Defence Agency's (EDA) AI strategy (2022) further confirms that data-driven quality assurance in defense supply chains has become a decisive factor for competitiveness, and that NATO member countries are increasingly demanding digital quality management infrastructure from defense suppliers.

The educational dimension of this gap is equally significant and has received no scholarly attention. Vocational higher education programs in defense and aerospace engineering — in Turkey as elsewhere — produce graduates who enter quality management roles in AS9100-certified organizations without any structured exposure to the intersection of AI technologies and sector-specific compliance requirements. This curricular absence mirrors the research gap: just as the AI+QMS literature ignores AS9100 specificity, engineering education programs rarely engage with AI-driven QMS as a domain of professional competency. The present literature review therefore serves a dual purpose: it maps the research gaps for scholars, and it simultaneously diagnoses the curricular blind spots that leave prospective quality engineers unprepared for the AI-transformation of their professional domain.

This article presents a systematic literature review across three axes—the AS9100 standard, artificial intelligence technologies, and cloud-based modular architecture—mapping existing research gaps at their intersection and proposing a theoretical framework to close these gaps. The following sections are structured as: the historical

evolution of quality and the distinctive requirements of AS9100 beyond ISO 9001 (Section 2); the role of AI in quality management and the Quality 4.0 literature (Section 3); recent work on large language models and document intelligence (Section 4); quality management literature in the defense and aerospace industry (Section 5); identified literature gaps (Section 6); and conclusion (Section 7).

From ISO 9001 to AS9100: The Evolution of Quality and Sector-Specific Requirements

Historical Foundations of Quality Management

The scientific foundations of quality management rest on the statistical process control (SPC) work developed by Walter Shewhart at Bell Laboratories in the 1920s. Shewhart classified variability in production processes as either common cause or special cause, and pioneered the use of control charts to monitor process deviation in real time. W. Edwards Deming transformed this work into a systematic management philosophy and applied it in Japan during the 1950s, triggering a quality revolution in Japanese manufacturing through the Plan-Do-Check-Act (PDCA) cycle.

While Joseph Juran positioned quality as a strategic management function, Philip Crosby's 'zero defects' philosophy and Armand Feigenbaum's Total Quality Control (TQC) concept paved the way for quality to evolve into an organization-wide responsibility. The synthesis of these pioneers' work was conceptualized in the late 1980s as Total Quality Management (TQM) and spread on a global scale. The need to internationally standardize these developments was formalized with the ISO 9000 series published by ISO in 1987.

The current ISO 9001:2015 revision elevated the standard beyond a purely procedural framework by adding strategic dimensions such as 'risk-based thinking,' 'leadership,' and 'understanding the organization's context,' repositioning it as a management system standard (ISO, 2015). The global diffusion of ISO 9001 has also reached remarkable proportions: according to the ISO Survey 2023, over one million valid ISO 9001 certificates exist in more than one hundred and twenty-five countries worldwide, with certificate numbers continuing to grow in defense- and aviation-heavy economies (ISO, 2023). This scale demonstrates that ISO 9001 serves not only as a technical foundation for upper-tier standards such as AS9100, but also as the common language of global supply chain assurance.

This historical trajectory of quality management has gained a new dimension through the concept of Quality 4.0. Chiarini and Kumar (2022), in their mixed-methods study of Italian manufacturing firms, found that Quality 4.0 still lacks a clear definitional framework, identifying the integration of digital technologies—artificial intelligence, big data, IoT, digital twins—into quality management practices as the core of this concept. Chiarini (2020), in a systematic review of the Industry 4.0 and TQM literature, emphasized that the integration of these two fields is still in its infancy and that significant research gaps exist, particularly in the areas of standardization, data quality, and organizational change management. Maganga and Taifa (2023) demonstrate how standardized quality systems are evolving into dynamic, data-driven structures through digital transformation. Chiarini and Cherrafi (2023) presented a PDCA-based implementation framework integrating ISO 9001 with Industry 4.0 for the manufacturing sector; however, the scope of the study remained limited to a general manufacturing context and

did not address sector-specific standards such as AS9100. Antony et al. (2024), in their qualitative study investigating critical failure factors in Quality 4.0 implementations, identified the neglect of standard specificity and the disregard for sectoral context as key obstacles. The common blind spot in this literature is also clear: Quality 4.0 studies predominantly focus on a general manufacturing context and do not systematically address heavily requirement-laden standards such as AS9100. Aquilani et al. (2017), in their comprehensive systematic review of TQM critical success factors, also found that existing measurement and evaluation frameworks fail to fully capture organizational transformation.

AS9100: Sector-Specific Requirements Beyond ISO 9001

The historical roots of AS9100 trace back to the late 1990s when the aviation industry began building sector-specific requirements on top of ISO 9001. IAQG published the first version of AS9100 in 1999, and the standard has since undergone four major revisions. The most recent version, AS9100 Rev D (2016), was prepared in full alignment with the High Level Structure of ISO 9001:2015 and incorporates approximately one hundred aviation- and defense-specific requirements (IAQG, 2016a). AS9100 certification is registered in the OASIS database operated by IAQG and is sought as a prerequisite across global supply chains.

The most critical points of divergence between the two standards are evident along the axes of traceability, counterfeit part control, and configuration management. While supplier and product control in a standard ISO 9001 implementation is addressed at a basic level, AS9100 makes the detection, reporting, and quarantine of unauthorized, life-expired, or counterfeit parts that may infiltrate the supply chain a legal obligation (AS9100 Cl. 8.1.4). The concept of 'traceability' in AS9100 is not limited to the final product; it has been expanded to cover the entire lifecycle, from the design phase to the production line, from raw material certificates to personnel competency approvals. Configuration management (AS9100 Cl. 8.1.2), which requires the immediate recording and retrospective querying of every hardware and software change in a defense system, is among the heaviest procedural requirements distinguishing AS9100 from ISO 9001 (Annaç Göv, 2018).

How critical these requirements are is clearly reflected in the sector's audit standards. The IAQG's AS9101F audit standard (2016) mandates that traceability records and configuration management be subjected to separate effectiveness evaluations in all audits conducted under AS9100. This requirement formally confirms the weight of these processes within AS9100 at the audit framework level and directly points to the process areas where AI-driven automation can deliver the highest compliance value. Early-period studies examining the distinctive requirements of quality management systems in the defense aviation sector also confirm the criticality of these requirements, demonstrating that the traceability and configuration management processes within AS9100 impose an operationally heavy burden that cannot be met by conventional QMS infrastructure (Thomas, 2006).

In the Turkish context, Annaç Göv (2018) examined the AS9100 compliance processes of leading firms such as TAI and THY Teknik, documenting with concrete examples the operational burden of the requirement gap between the two standards. The prominent finding of the study is that domestic defense and aviation suppliers expend significant resources and time, particularly in documenting configuration management and traceability

requirements. A recent study examining total quality management perception in Turkish civil aviation finds a positive but measured attitude toward TQM practices among sector employees, indicating that advanced quality standards such as AS9100 require organizational culture and human factors to be taken into account as well (Karaman, 2024).

The Role of AI in Quality Management and the Quality 4.0 Literature

Historical Development of AI and its Implications for the Defense Sector

The foundations of artificial intelligence were laid with Alan Turing's work in 1950 and at the Dartmouth Conference in 1956. Early research concentrated on symbolic reasoning and expert system development but could not be scaled to industrial levels due to limited computing power and data constraints. The 'AI winter' of the 1980s resulted in a marked reduction in research funding following exaggerated expectations.

The drivers of today's AI wave are the processability of large datasets, the dramatic increase in GPU-based parallel computing capacity, and the revolutionary impact of the Transformer architecture proposed by Google researchers in 2017 on natural language processing. With the emergence of BERT (2018) and GPT-series models, large language models (LLMs) have evolved into systems capable of processing unstructured text data with an analytical depth comparable to human intelligence (Minglana et al., 2021). This development has dramatically expanded the use cases of AI in document-heavy processes—quality procedures, compliance audits, technical reports, and the like.

The strategic role of AI in defense technologies has also deepened with this technological evolution. In defense, AI is employed across a wide spectrum, from autonomous systems management and threat detection to predictive maintenance and supply chain security. The AI programs of DARPA, the EDA's AI strategy, and NATO's AI principles are paving the way for AI to be adopted in defense not only operationally but also in management and quality process dimensions. In Turkey, the AI R&D activities of leading defense industry companies such as ASELSAN, HAVELSAN, and ROKETSAN reveal the potential of these technologies within the domestic defense ecosystem. The Turkish Informatics Association's (TBD, 2025) technical report on AI standards documents the Turkish reflections of international AI standards such as ISO/IEC 5338, ISO/IEC TS 4213, and ISO/IEC 42001, bringing to the fore the need for alignment with these standards in national policy development processes. In this context, the adoption of AI in the defense industry should be regarded not merely as a technical choice but as a strategic step aligned with national standardization policy. From an organizational readiness perspective, Gökalp et al. (2021), examining the digital transformation maturity levels of defense sector organizations in Turkey, found that the majority of firms are positioned in the middle-to-lower maturity bands in terms of data management infrastructure and digital process integration. This finding places the organizational readiness factor alongside technical feasibility as a design constraint for an AI-driven QMS solution.

AI and Quality Management Integration

The effects of AI in the context of quality management have been studied with growing systematicity in recent

years. Minglana et al. (2021), in their bibliometric mapping of AI applications across QMS processes, found that non-conformance detection, process automation, and predictive quality control are the most studied areas. This finding indicates that AI is treated as a technical rather than a strategic layer within QMS. The absence of any AS9100-specific study in this bibliometric mapping provides quantitative evidence of the primary literature gap addressed in this work.

Ispas et al. (2025), in their comprehensive review of AI applications in risk management processes within integrated management systems, emphasize that machine learning and NLP-based approaches yield significant results in proactive risk detection and error prevention. The effectiveness of predictive analytics in anticipating audit findings and prioritizing corrective actions offers a natural alignment with AS9100's risk-based approach. These findings indicate that AI in the defense industry should now be treated not merely as a 'tool' but as a 'system component' that is traceable and auditable.

The growing need to integrate AI into quality and risk management systems has also mobilized international standardization bodies, providing the groundwork for the emergence of next-generation standards such as ISO/IEC 42001 AI Management System. Gueorguiev (2025), in his clause-by-clause comparison of the integration points between ISO/IEC 42001 and ISO 9001, found that the two standards have a complementary character along the axes of Clause 4.1 (organizational context) and Clause 4.2 (needs and expectations of interested parties) and that a concrete roadmap for integration can be provided. This situation strengthens the theoretical foundation for positioning AI as a manageable component within a multi-layered quality management ecosystem that also encompasses AS9100. Siqing et al. (2013), examining the impact of QMS on knowledge creation processes in Chinese aviation firms, argued that QMS should be evaluated not merely as a procedural obligation but as an infrastructure that feeds organizational learning and innovation capacity. This finding provides a conceptual basis for an AI-enabled QMS to serve not only compliance functions but also organizational knowledge management.

Large Language Models, RAG, and Document Intelligence

Integration of LLMs into Organizational Processes

The integration of large language models into organizational processes constitutes a rapidly expanding area of research. Ke et al. (2025), analyzing approximately three hundred studies, systematically mapped the current state of LLMs in document understanding, information extraction, and RAG. This survey demonstrates that the technical maturity for automatically classifying unstructured text data within hierarchical relational frameworks has been reached and that applicability is high in specialized document ecosystems such as quality procedure repositories.

Fahland et al. (2025), in their study investigating the capacity of LLMs to explain and understand business processes, demonstrated the potential of LLMs to interpret workflow documents and serve as autonomous organizational assistants. The study also found that the quality of process explanation, evaluated from the user's perspective, is sensitive to the domain specificity of the model's training dataset, and drew attention to the need

for domain adaptation in technical fields such as quality documents. This finding supports, at a conceptual level, the possibility that quality management modules can evolve from static document repositories into structures that actively analyze processes.

The recent review by Knott et al. (2026) evaluates the contribution of LLMs to compliance monitoring and process optimization from a European perspective, emphasizing that the automation potential in document management and regulatory compliance processes is sector-independent. The complex regulatory compliance and traceability requirements of the healthcare sector bear structural similarities to the AS9100 compliance process in the defense industry; LLM applications in the healthcare context therefore provide a valuable analog for the defense QMS context.

RAG Technology and Quality Management Applications

The Retrieval-Augmented Generation (RAG) approach enriches response generation with an instantly accessed document repository in order to reduce the hallucination problem of LLMs. Ke et al.'s (2025) comprehensive survey demonstrates that the increase in long-context processing capacity opens the door to analyzing complex quality procedures and configuration records in their entirety within the model context. This feature directly aligns with the accuracy and traceability requirements demanded by AS9100.

Nevertheless, the limitations of LLM and RAG technologies in the quality management context should not be overlooked. Huang et al. (2023), in their study examining the performance of LLMs on technical and regulatory documents, showed that response accuracy drops markedly when models encounter highly domain-specific terminology, and that the risk of misinterpreting numerical tolerances and procedural conditions in guidance documents increases. This finding confirms that domain adaptation and human oversight must be incorporated into the system architecture as indispensable components in the use of LLMs in a standard requiring technical precision such as AS9100.

Cloud-Based Quality Management Systems and Existing Software Solutions

Conventional quality management software is generally designed as a monolithic structure and operates dependent on organizations' own on-premise servers. SAP QM has served as the digital backbone of quality processes for large industrial organizations for more than a decade; however, high licensing costs, complex customization requirements, and lengthy update cycles make these platforms difficult to access, particularly for mid-sized defense suppliers. More recent cloud QMS platforms such as ETQ Reliance and Qualio improve the user experience but fall short of fully meeting the deep traceability and configuration management requirements demanded by AS9100.

Managing standards such as AS9100, which require intensive data traffic, real-time updates, and complex cross-references, demands modern architectures that are flexible, scalable, and integrable with external systems rather than static software. At this point, the concept of cloud-based serverless architecture comes to the fore. Serverless

architecture is a software paradigm in which infrastructure management, server maintenance, and auto-scaling are managed by cloud service providers, allowing developers to focus solely on business logic and function code. The Function-as-a-Service (FaaS) model enables each quality process to be configured as an independent, triggerable, and manageable service. Suh et al. (2023) demonstrated that quality management systems built on serverless architecture significantly reduce maintenance costs compared to conventional on-premise structures and increase the flexibility to adapt to standard updates.

The security dimension emerges as a critical parameter in the evaluation of cloud-based QMS in the defense industry context. Quality documents that are not classified but contain sensitive defense data require that access controls, data encryption, and audit trail mechanisms applied in cloud infrastructure go beyond general-purpose cloud services to meet defense-specific security requirements. FedRAMP certification in the United States provides a framework to fill this gap, while SSB's information security regulatory framework in Turkey serves a similar function.

For the integration of AI into quality processes, event-driven architecture approaches and API-first infrastructure are prominent in the literature. Modular architecture approaches in the quality management context increase the system's adaptability to different standards (ISO 9001, AS9100, IATF 16949) by separating the data layer, business logic layer, and presentation layer from each other. From a scalability perspective, Katona (2021) study —focused on the statistical validation of risk-based quality control techniques in the automotive sector—provides a strong empirical basis for the adaptability of QMS methodologies to different sectoral contexts. Sumranwong (2011), who addressed one of the first integration attempts of AS9100 with digital systems, presented an early approach to combining standard requirements with e-learning modules; while this study represents a pioneering effort demonstrating the feasibility of supporting AS9100 with digital tools, it falls short of the level of automation and intelligence provided by today's AI technologies.

Identified Literature Gaps

When the intersection set of the four thematic axes presented above is analyzed, the existing literature is found to contain three structural gaps.

The first gap concerns standard specificity. The approximately one hundred sector-specific requirements that AS9100 adds to ISO 9001 (IAQG, 2016a) are not addressed in general QMS research. The bibliometric mapping conducted by Minglana et al. (2021) reveals that the most studied areas at the AI+QMS intersection are non-conformance detection and process automation; yet notably, none of the studies reviewed makes a direct reference to the AS9100 standard. Quality 4.0-oriented studies (Antony et al., 2024; Chiarini & Cherrafi, 2023; Chiarini & Kumar, 2022; Chiarini, 2020; Maganga & Taifa, 2023) also do not directly address the AS9100 context. No system design aimed at automating these requirements through AI exists in the literature.

The second gap concerns architectural coherence. Studies examining AI-enabled quality audit applications generally focus on a single process and do not offer a modular and integrated architectural proposal that addresses

process management, document linking, and risk control together (Ispas et al., 2025). Existing QMS software—SAP QM, ETQ Reliance, Qualio, MasterControl, AssurX—only partially complies with AS9100; none simultaneously offers LLM/RAG integration and multi-standard compatibility.

The third gap concerns scalability. Design-oriented research on modular quality systems that are compliant with sector-specific standards (AS9100, IATF 16949) while also being adaptable to different standards is virtually nonexistent. Katona (2021) provides valuable contributions on risk-based quality control in the automotive sector, and Sumranwong (2011) demonstrates the feasibility of supporting AS9100 with digital tools. However, none of these studies addresses a system that hosts AS9100 and IATF 16949 on the same modular core with LLM integration and FaaS-based architecture. The emergence of next-generation AI management standards such as ISO/IEC 42001 (Gueorguiev, 2025) renders the closing of this gap both possible and necessary.

All three dimensions of this gap require quality management, AI, and the defense sector to be examined not merely in binary combinations but from an integrated system design perspective within the framework of the AS9100 standard. Table 1 presents a systematic comparison of these gaps against the key sources in the existing literature.

Table 1. Source-Based Comparison of Identified Literature Gaps

Source	**Quality 4.0**	**AI + QMS**	**AS9100 Specificity**	**Architectural Coherence**	**Scalability**
Chiarini (2020)	Yes	Partial	No	No	No
Chiarini and Kumar (2022)	Yes	Partial	No	No	No
Maganga and Taifa (2023)	Yes	Partial	No	No	No
Minglana et al. (2021)	No	Yes	No	No	No
Ispas et al. (2025)	No	Yes	No	Partial	No
Gueorguiev (2025)	No	Yes	Partial	No	No
Annaç Göv (2018)	No	No	Yes	No	No
Katona (2021)	No	No	No	No	Yes
Suh et al. (2023)	No	Partial	No	Yes	Partial
This study	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Conclusion

This literature review has demonstrated that existing studies on the integration of quality management systems with artificial intelligence—when systematically examined under the headings of Quality 4.0, AI+QMS, large language models, and quality management in the defense industry—contain three critical gaps: standard specificity, architectural coherence, and scalability.

On the dimension of standard specificity, the approximately one hundred sector-specific requirements added by AS9100 Rev D to ISO 9001 are systematically overlooked in the AI+QMS literature. The absence of any AS9100-specific study in the bibliometric mapping conducted by Minglana et al. (2021) constitutes the quantitative evidence of this gap. On the dimension of architectural coherence, the single-process-oriented structure of existing AI-enabled quality audit studies reveals the absence of proposals that address configuration management, traceability, and document intelligence within an integrated architectural framework. On the dimension of scalability, it has been established that design-oriented studies hosting both AS9100 and IATF 16949 on the same modular core are virtually nonexistent.

These gaps justify—both theoretically and practically—the development of a modular system that integrates AI and cloud-based architecture into quality management processes with a focus on AS9100 requirements, structured within the Design Science Research (DSR) methodology (Hevner et al., 2004). Turkey's defense industry ecosystem—SSB's AI academic programs, STM Think Tech (2024) technology agenda, and Turkish Informatics Association (TBD) (2025) AI standards work—further confirms this need at the level of national strategic priorities.

The QualityAiMate system to be designed within the scope of this study aims to close the three-dimensional literature gap mapped above along the axes of AS9100's standard specificity, the architectural coherence of LLM/RAG-powered document intelligence, and the scalability of FaaS-based serverless architecture. The design output is expected to both provide the quality management literature with a body of reusable design knowledge and equip organizations in the defense industry with a directly applicable digital quality infrastructure. Beyond its contribution to the QMS and AI literatures, this study carries a specific implication for vocational higher education. The three-dimensional literature gap identified here — standard specificity, architectural coherence, and scalability — corresponds precisely to the competency gaps that AS9100-relevant engineering curricula must address if graduates are to function effectively in AI-transformed quality management environments. Universities and vocational schools preparing engineers for the defense and aerospace sector should treat AI-integrated QMS design not as a peripheral elective topic but as a core professional competency domain, one that requires structured exposure to sector-specific standards, LLM-based document intelligence, and cloud architecture principles in combination.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: The ideas, conceptual framework, theoretical contributions, and all original intellectual content presented in this study belong entirely to the authors. Generative AI tools — specifically Google Gemini, OpenAI ChatGPT, and Anthropic Claude — were used solely to support editorial processes, including linguistic fluency, structural organization, grammar review, and preliminary manuscript feedback. No analytical decisions, gap identifications, or interpretive conclusions were generated by or delegated to these tools. All scholarly responsibility for the content of this manuscript rests exclusively with the authors.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: This study is a systematic literature review based entirely on publicly available published sources. No human subjects, primary data collection, experiments, or personal data processing were involved. The study therefore falls outside the scope of research requiring ethics committee approval under BERA (2018) guidelines and KVKK (Law No. 6698). All sources referenced are properly cited in accordance with APA conventions, and no data were collected or processed beyond what is available in the open academic literature.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Annaç Göv, S. (2018). Havacılık Sektöründe AS9100 Kalite Standardı: ISO 9001 ile Karşılaştırma [AS9100 quality standard in the aviation sector: A comparison with ISO 9001]. *İktisadi İdari ve Siyasal Araştırmalar Dergisi*, *3*(7), 144–155.

Antony, J., Kaul, A., Bhat, S., Sony, M., Kaul, V., Zulfiqar, M., & McDermott, O. (2024). Critical failure factors for Quality 4.0: An exploratory qualitative study. *International Journal of Quality and Reliability Management*, *41*(4), 1044–1062.

Aquilani, B., Silvestri, C., Ruggieri, A., & Gatti, C. (2017). A systematic literature review on total quality management critical success factors and the identification of new avenues of research. *The TQM Journal*, *29*(1), 184–213. https://doi.org/10.1108/TQM-01-2016-0003

Chiarini, A. (2020). Industry 4.0, quality management and TQM world: A systematic literature review and a proposed agenda for further research. *The TQM Journal*, *32*(4), 603–626.

Chiarini, A., & Cherrafi, A. (2023). Integrating ISO 9001 and Industry 4.0: An implementation guideline and PDCA model for manufacturing sector. *Total Quality Management & Business Excellence*, *34*(13–14), 1629–1654. https://doi.org/10.1080/14783363.2023.2192916

Chiarini, A., & Kumar, M. (2022). What is Quality 4.0? *An exploratory sequential mixed methods study of Italian manufacturing companies. International Journal of Production Research*, *61*(5), 1–19.

Dale, G. K. (2010). AS9100 aerospace requirements: Above and beyond. *Quality Progress*, *43*(1), 60–63.

European Defence Agency (EDA) (2022). *Artificial intelligence in defence: Strategic framework for European defence supply chains*. EDA.

Fahland, D., Fournier, F., Limonad, L., Skarbovsky, I., & Swevels, A. J. E. (2025). How well can a large language model explain business processes as perceived by users? *Data and Knowledge Engineering*, *157*, Article 102416.

Gökalp, E., Gökalp, M. O., & Çoban, S. (2021). Digital transformation capability maturity model enabling the assessment of industrial manufacturers. *Computers in Industry*, *132*, 103522.

Gueorguiev, T. (2025). An approach to integrate artificial intelligence in ISO 9001-based quality management systems. *Measurement: Sensors*, *38*, 101787.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Tang, H., & Deng, Y. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Computing Surveys*, *57*(3), Article 63.

International Aerospace Quality Group (IAQG) (2016a). *AS9100 Rev D: Quality management systems—Requirements for aviation, space, and defense organizations*. SAE International.

International Aerospace Quality Group (IAQG) (2016b). *AS9101F: Quality management systems—Audit requirements for aviation, space, and defense organizations.* SAE International.

International Organization for Standardization (ISO) (2015). *ISO 9001:2015—Quality management systems—Requirements*. ISO.

International Organization for Standardization (ISO) (2023). *ISO survey of certifications 2023: Results*. ISO.

Ispas, L., Mironeasa, C., Severin, T.-L., Cerlincă, D.-A., & Mironeasa, S. (2025). Artificial intelligence applications in risk management within integrated management systems: A review. *Systems*, *13*(1), 967.

Karaman, F. (2024). Türk sivil havacılığında toplam kalite yönetimi algısı ve çalışan performansının incelenmesi [Examining total quality management perception and employee performance in Turkish civil aviation]. *Alanya Akademik Bakış*, *8*(2), 421–438.

Katona, A. I. (2021). Validation of risk-based quality control techniques: A case study from the automotive industry. *Journal of Applied Statistics*, *49*(12), 3008–3025.

Ke, W., Zheng, Y., Li, Y., Xu, H., Nie, D., Wang, P., & He, Y. (2025). Large language models in document intelligence: A comprehensive survey, recent advances, challenges, and future trends. *ACM Transactions on Information Systems*, *44*(1), 1–64.

Knott, M., Krebs, M., & Kerscher, A. (2026). Large language models in healthcare quality management: A European perspective on process automation and compliance. *Frontiers in Digital Health*, *8*, 1–11.

Maganga, D. P., & Taifa, I. W. R. (2023). Quality 4.0 conceptualization: An emerging quality management concept for manufacturing industries. *The TQM Journal*, *35*(4), 986–1013.

Minglana, J., Tobias, R. R., & Roxas, R. E. (2021). *Artificial intelligence applications in quality management system: A bibliometric study*. Proceedings of TENCON 2021—IEEE Region 10 Conference.

Presidency of Defence Industries (SSB) / Council of Higher Education (YÖK) (2024). *Yapay Zekâ Akademik Tez Programı (ATP) Duyurusu [Artificial Intelligence Academic Thesis Program (ATP) announcement]*. SSB.

Siqing, S., Qiuhong, Z., & Fan, H. (2013). Impact of quality management practices on the knowledge creation process: The Chinese aviation firm perspective. *Computers & Industrial Engineering*, *64*(1), 211–223.

STM ThinkTech (2024). *Üretken Yapay Zekâ (Generative AI) Teknolojisi: Savunma Sanayi Etki Analizi [Generative AI technology: Defense industry impact analysis]*. STM Savunma Teknolojileri Mühendislik.

Suh, Y., Park, Y., & Kang, D. (2023). Serverless cloud architecture for quality management system integration: A cost and agility analysis. *Journal of Manufacturing Systems*, *67*, 112–124.

Sumranwong, D. (2011). *An eLearning model application for AS 9100 standard*. Proceedings of ICIMIE 2011.

Thomas, K. T. (2006). Quality management system for defence aeronautical industry. *Defence Science Journal*, *56*(3), 333–348.

Turkish Informatics Association (TBD) (2025). *Yapay Zekâ Standartları Raporu: ISO/IEC 5338, ISO/IEC TS 4213 ve ISO/IEC 42001 Kapsamında Türkiye Değerlendirmesi [AI standards report: Turkey assessment within ISO/IEC 5338, ISO/IEC TS 4213 and ISO/IEC 42001]*. TBD.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Cultural Connotation and Contemporary Educational Value of Microcosmization in Traditional Chinese Medicine

Shengxue Li [1 *], Guangren Sun [2]

[1] Research Center of Jilin Tiansanqi Pharmaceutical Co., Ltd., China
[2] Research Center of Jilin Tiansanqi Pharmaceutical Co., Ltd., China
* Corresponding author: Shengxue Li (icemstoffice@gmail.com)

Article Info

Article History
Received:
9 February 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
TCM microcosmization
Cultural connotation
Contemporary educational value
TCM Qi
Yin-yang balance

Abstract

TCM microcosmization represents the integration of traditional TCM theory with modern microscopic science, serving as a key path for TCM modernization and cultural inheritance. This study examines its core connotation, cultural implications, and educational value, providing scientific support for its application. Departing from the macro-holistic view of TCM, this approach reinterprets core concepts microscopically, proposing that "TCM Qi" essentially reflects invisible microbial communities and their metabolic products, forming the material basis of human life. Research shows the ratio of human somatic to microbial cells is 1:1, with intestinal microorganisms comprising one-third of fecal dry matter and forming an integral human microecosystem. This aligns with the TCM concept of Yin-Yang balance: microbial imbalance causing absorption disorders corresponds to "dysfunction of the spleen in transporting and transforming". Regulating microbial communities at the microscopic level can achieve Yin-Yang balance, corroborating TCM theory. Culturally, this approach preserves TCM's "unity of man and nature" concept while incorporating modern scientific thinking, strengthening cultural confidence. Educationally, it addresses gaps in traditional TCM training, fostering talents with both scientific and cultural literacy. Overall, TCM microcosmization represents an inevitable modernization trend, supporting cultural inheritance and educational reform, warranting further research and promotion.

Citation: Li, S., & Sun, G. (2026). Cultural Connotation and Contemporary Educational Value of Microcosmization in Traditional Chinese Medicine. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 161–169), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

OPEN ACCESS

Introduction

Traditional Chinese medicine (TCM) has historically been grounded in a macroscopic, holistic, and relational understanding of the human body. Its theoretical system emphasizes the coordinated operation of Qi and blood, the functional interdependence of the zang-fu organs, the circulation of meridians, and the dynamic equilibrium of Yin and Yang. Within contemporary scientific discourse, however, TCM is often considered difficult to interpret because many of its core concepts are expressed in classical philosophical language rather than modern biomedical terminology. This epistemic gap has become a major challenge both for the modernization of TCM and for the design of TCM education in contemporary academic settings (Sender et al., 2016; Xu & Yang, 2013; Yang et al., 2022).

Against this background, the concept of TCM microcosmization has emerged as a potentially productive pathway for theoretical translation. In the present study, microcosmization does not mean reducing TCM to laboratory biology or replacing classical theory with microbiology. Rather, it refers to an interpretive strategy that seeks microscopic correlates for selected TCM concepts, thereby making them more accessible to modern scientific reasoning while preserving the systemic and relational logic of TCM. Among the currently available scientific interfaces, the host–microbiota system is especially relevant because it connects invisible biological processes with whole-body regulation, environmental responsiveness, and adaptive balance (Kasarello et al., 2023; Li et al., 2023; Lyte, 2014; Thursby & Juge, 2017; Yang et al., 2022; Zhang et al., 2020).

One of the most pedagogically significant propositions within this framework is that the classical concept of Qi may be partially interpreted through the functional activity of microscopic living systems and their metabolites. This proposition is educationally valuable because it allows students to relate an abstract classical term to observable biological processes. Estimates indicating an approximately 1:1 ratio between bacterial and human cells in the body have further reinforced the scientific importance of the microbiome as a constitutive component of human physiology rather than a merely external adjunct (Sender et al., 2016).

At the same time, research on intestinal microbiota has become central to contemporary explanations of metabolism, immunity, neuroendocrine signaling, and inter-organ communication. These findings resonate with several TCM propositions, including the importance of transformation and transportation, the systemic consequences of internal imbalance, and the maintenance of health through dynamic equilibrium rather than static normality (Kasarello et al., 2023; Li et al., 2023; Lyte, 2014; Thursby & Juge, 2017; Yang et al., 2022; Zhang et al., 2020). Such convergences do not constitute literal proof of classical theory; however, they do provide a meaningful conceptual bridge between TCM and modern science and, importantly, a useful pathway for pedagogical innovation in TCM education.

Accordingly, the present study aims to clarify the cultural connotation and contemporary educational value of TCM microcosmization. Specifically, it seeks to: (1) define the conceptual scope of microcosmization in TCM; (2) interpret its cultural significance in relation to the inheritance and modernization of TCM knowledge; (3) construct visual models that explain the relationships among microecology, systemic regulation, and educational

translation; and (4) discuss how these models may support curriculum innovation and interdisciplinary talent cultivation.

Methods

This study adopted a narrative review combined with conceptual modeling. Rather than evaluating a specific clinical intervention, it sought to construct an interpretive framework for understanding the cultural connotation and educational value of microcosmization in Traditional Chinese Medicine (TCM). Relevant evidence was drawn from open-access literature on gut microbiota, host–microbiota interaction, neuro-endocrine-immune regulation, TCM modernization, and TCM education.

Literature selection was guided by three criteria: conceptual relevance, public accessibility for academic verification, and applicability to educational interpretation. Priority was given to studies that could help connect modern microscopic research with core TCM concepts, including Qi, Yin–Yang balance, spleen transformation and transportation, and the unity of humans and nature.

The analysis was conducted in three steps. First, representative scientific findings related to intestinal microecology and microbiota-associated regulation were identified and summarized. Second, these findings were interpreted in relation to classical TCM theory in order to establish a cross-disciplinary conceptual framework. Third, the framework was translated into visual models to support both theoretical explanation and educational application, including a systems model, a regulatory framework, and an educational visualization model.

This study should therefore be understood as a theory-oriented and education-oriented synthesis rather than as an experimental or clinical investigation. It does not assume that microbiology can fully explain all dimensions of TCM, nor does it claim that classical TCM concepts have direct one-to-one equivalents in modern science. Instead, it explores whether microcosmization may serve as a meaningful interface for connecting classical TCM theory with contemporary scientific discourse and educational practice.

Results

Core Connotation of TCM Microcosmization

The first major finding of this study is a clearer definition of TCM microcosmization. In the present context, microcosmization refers to the translation of macro-level TCM constructs into a microscopic explanatory framework grounded in microorganisms, metabolites, secretions, and host regulatory signaling. This perspective does not abandon the systemic logic of TCM; rather, it reformulates selected concepts in a biological language that is more compatible with contemporary teaching, academic communication, and interdisciplinary discussion.

The second finding is the identification of intestinal microecology as a particularly effective anchor for this translation. The gut microbiota is scientifically relevant because it participates in nutrient transformation, immune programming, barrier maintenance, signaling to the nervous and endocrine systems, and metabolic homeostasis

(Kasarello et al., 2023; Li et al., 2023; Lyte, 2014; Thursby & Juge, 2017; Yang et al., 2022; Zhang et al., 2020). Educationally, it is equally valuable because it offers students a visible and evidence-based entry point for understanding otherwise abstract concepts such as internal balance, dynamic regulation, and multi-organ crosstalk.

The third finding is a more cautious reformulation of the relationship between Qi and microscopic life. Instead of treating Qi as a direct synonym for microbes, this study proposes that microecological activity may serve as one plausible material and functional substrate through which selected dimensions of Qi can be interpreted in modern pedagogy. This formulation is more suitable for SCI-style writing because it acknowledges both conceptual continuity and explanatory limitation.

Construction of the Human Yin–Yang System Model

To clarify the systemic logic of TCM microcosmization, a Human Yin–Yang System Model was constructed, as shown in Figure 1.

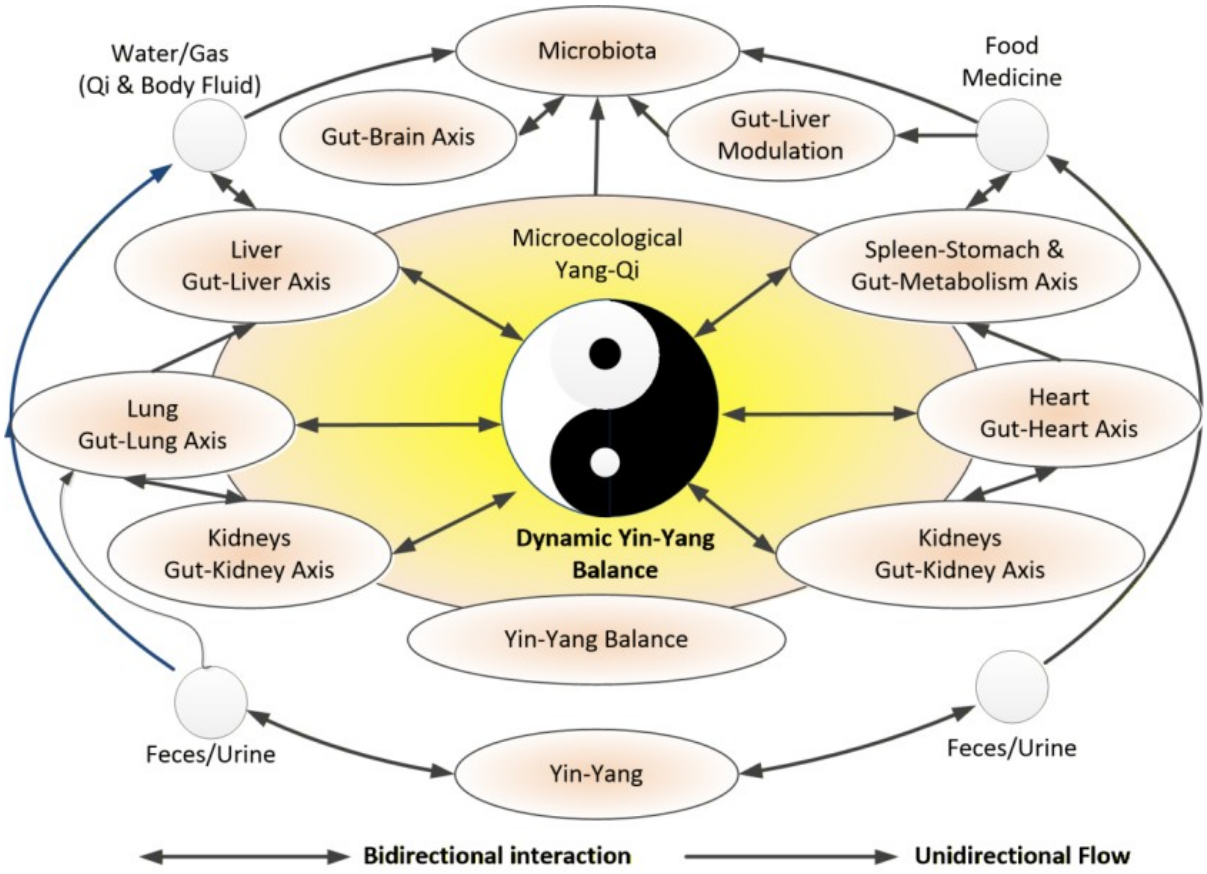


Figure 1. Human Yin–Yang System Model

[Notes and abbreviations for Figure 1: TCM, traditional Chinese medicine; Qi, a classical TCM concept referring to functional vitality or dynamic activity; GBA, gut–brain axis; GLA, gut–liver axis; GLoA, gut–lung axis; GKA, gut–kidney axis; GHA, gut–heart axis; GMA, gut–metabolic axis. "Microecological Yang-Qi" is used here as an interpretive term rather than a standardized biomedical entity.]

[Notes and abbreviations for Figure 1: TCM, traditional Chinese medicine; Qi, a classical TCM concept referring to functional vitality or dynamic activity; GBA, gut–brain axis; GLA, gut–liver axis; GLoA, gut–lung axis; GKA, gut–kidney axis; GHA, gut–heart axis; GMA, gut–metabolic axis. "Microecological Yang-Qi" is used here as an interpretive term rather than a standardized biomedical entity.]

In Figure 1, A conceptual schematic illustrating the proposed integration of intestinal microecology, microecological Yang-Qi, and six major gut-associated regulatory axes within a dynamic Yin–Yang balance framework. The model includes the gut–brain, gut–liver, gut–lung, gut–kidney, gut–heart, and gut–metabolic axes and depicts their functional relationships with major organ systems. Environmental inputs (food, medicine, water, and gas exchange) and excretory outputs (feces and urine) are included to emphasize the open and adaptive nature of host–microecology interactions.

This model positions microecological Yang-Qi at the center and arranges six major gut-associated axes around it, thereby presenting microecology as a whole-body regulatory hub rather than a phenomenon confined to the intestine. From a writing perspective, the model strengthens the argument that TCM microcosmization should be understood in systemic terms: it visually aligns the TCM idea of dynamic Yin–Yang balance with contemporary concepts of bidirectional signaling, homeostatic regulation, and multi-organ integration.

Construction of the human regulatory framework

To further explain the regulatory pathway implied by microcosmization, a Human Regulatory Framework was constructed, as shown in Figure 2. In Figure 2, a schematic regulatory framework showing microecology as a central neuro-endocrine-immune hub linking the human body, mental cognition, treatment principles, treatment techniques, therapeutic products, food decomposition, probiotics, and microecological modulation. The figure illustrates how therapeutic and physiological pathways may converge on microecological regulation and thereby contribute to systemic adaptation and functional balance.

Compared with Figure 1, this model is more explicitly intervention-oriented. It visualizes how treatment principles, dietary processing, microecological modulation, and host adaptation converge on a shared microecological core, thereby translating an abstract theoretical proposition into a more operational framework. In educational terms, the figure supports the view that TCM microcosmization can be taught not only as a conceptual bridge, but also as a mechanism-oriented approach to regulation and intervention.

Construction of the Educational Visualization Model of TCM Microcosmization

To translate the preceding theoretical discussion into pedagogy, an Educational Visualization Model of TCM Microcosmization was constructed, as shown in Figure 1.

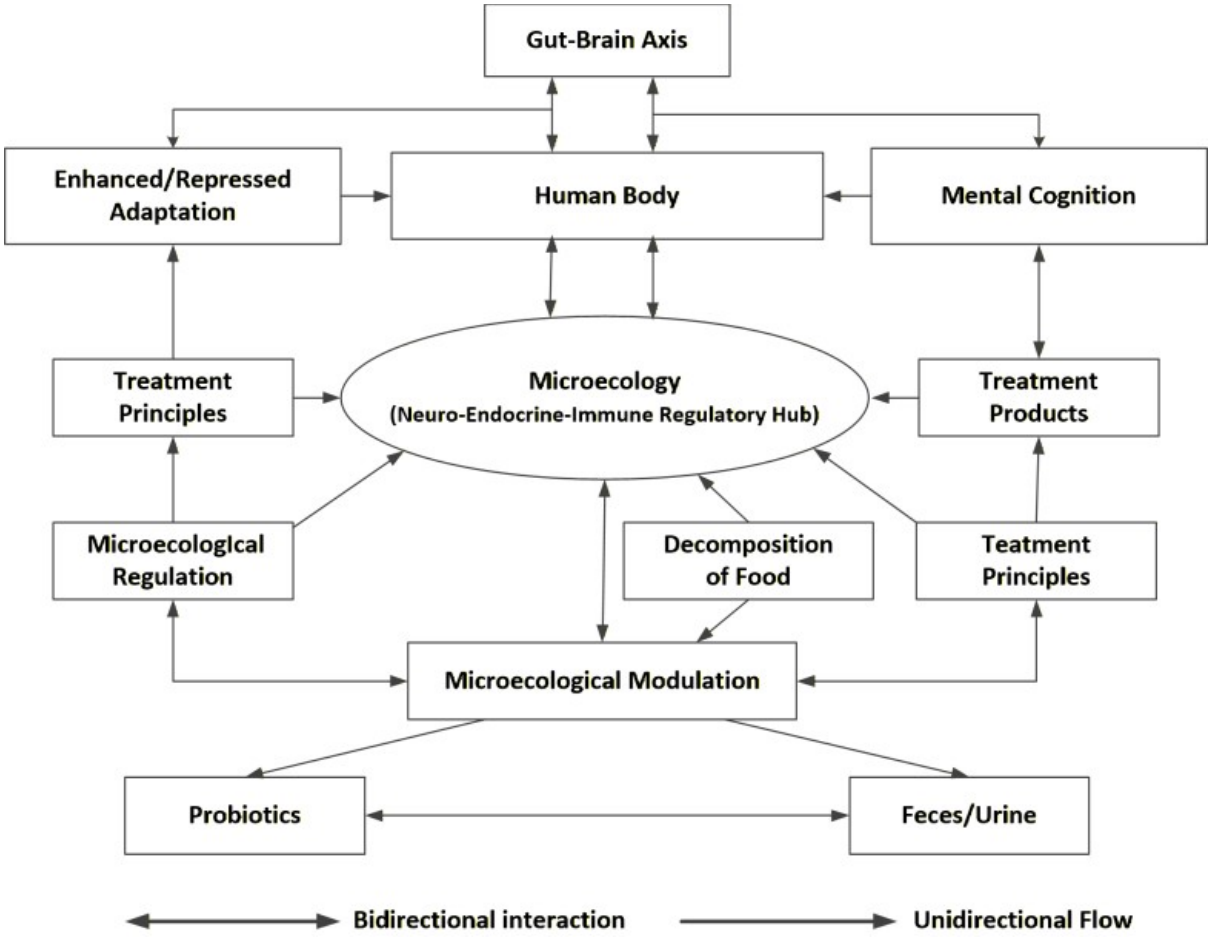


Figure 2. Human Regulatory Framework

[Notes and abbreviations for Figure 2: NEI, neuro-endocrine-immune; TCM, traditional Chinese medicine. "Microecology" refers here to the host-associated microbial ecosystem and its metabolites. "Treatment principles," "treatment technique," and "therapeutic products" are conceptual educational categories and should not be interpreted as validated efficacy claims for any specific intervention.]

Discussion

The findings of the present study suggest that TCM microcosmization has significance at both the cultural and educational levels. Culturally, it enables classical TCM concepts to be reinterpreted in terms that are more intelligible to contemporary academic audiences without abandoning the foundational TCM emphasis on relation, process, and dynamic equilibrium. In this sense, microcosmization may be regarded as a strategy of inheritance through translation rather than replacement (Xu & Yang, 2013; Yang et al., 2022).

From the perspective of scientific communication, the microbiota provides a particularly suitable interface for this translation. It is invisible to unaided observation, widely distributed in and on the body, highly responsive to environment and diet, and deeply implicated in digestive, immune, neuroendocrine, and metabolic regulation (Kasarello et al., 2023; Li et al., 2023; Lyte, 2014; Sender et al., 2016; Thursby & Juge, 2017; Yang et al., 2022; Zhang et al., 2020). These characteristics make it both a persuasive educational metaphor and a plausible partial

explanatory scaffold for teaching TCM concepts such as internal balance, systemic interdependence, and the continuity between the human body and its environment.

The educational implications are equally significant. TCM education has long faced the tension between fidelity to classical textual traditions and the demand for modern scientific explainability. By incorporating microbiome science, microecological models, and visual frameworks, educators may be better positioned to support students from interdisciplinary backgrounds, particularly those trained primarily in conventional biomedicine (Qu et al., 2021; Yang et al., 2022). At the same time, such an approach may strengthen cultural confidence by showing that TCM can participate in contemporary scientific discussion without losing its historical identity.

Nevertheless, several cautions are necessary. First, the microcosmization approach should not be overstated as a complete or exclusive explanation of TCM, because classical TCM is philosophically richer than any single biological model. Second, many of the proposed correspondences between TCM concepts and modern microecological processes remain heuristic rather than experimentally definitive. Third, pedagogical simplification must avoid creating a false equivalence between metaphorical language and mechanistic proof. For these reasons, the framework proposed here should be treated as a rigorous educational hypothesis and conceptual bridge rather than as a final ontology of TCM.

Future research may strengthen this field in at least three directions. First, empirical teaching studies are needed to evaluate whether microcosmization-based pedagogy improves student comprehension, acceptance, and cross-disciplinary communication. Second, deeper interdisciplinary scholarship should connect microbiome science, systems biology, philosophy of medicine, and TCM hermeneutics. Third, more refined educational resources are required, including bilingual glossaries, figure-based teaching modules, and case-centered curriculum design.

Conclusion

TCM microcosmization represents an important contemporary pathway for connecting the cultural inheritance of traditional Chinese medicine with the explanatory demands of modern science and education. By using intestinal microecology as a conceptual anchor, selected TCM ideas—especially Qi, Yin–Yang balance, and spleen-related transformation and transportation—can be rendered more accessible, discussable, and pedagogically usable within contemporary academic settings.

The three models developed in this study provide complementary support for this argument. Figure 1 offers a systems model of dynamic balance, Figure 2 presents an intervention-oriented regulatory framework, and Figure 3 translates the same logic into educational design. Together, these models indicate that TCM microcosmization can function not only as a modernization strategy, but also as a practical pedagogical resource for cultivating scientifically literate and culturally grounded TCM talent.

In conclusion, microcosmization should be understood as a bridging concept: it neither abandons the holistic spirit of TCM nor simply reduces TCM to laboratory biology. Rather, it provides a structured interpretive pathway

through which TCM can be rearticulated, communicated, and taught in the contemporary era.

Statements and Declarations

Acknowledgments and Notes: The authors thank the reviewers and editors for their constructive comments on the conceptual framing and educational relevance of this manuscript. This work was supported by the Project of Ginseng Fermentation Technology Scientist Studio, Tiansanqi Group, Jilin Province (No. RD2026-1) and the Project of Tiansanqi Institute of Microcosmization in Traditional Chinese Medicine, Jilin (Research on Qi Theory in TCM Microcosmization).

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Kasarello, K., Cudnoch-Jędrzejewska, A., & Czarzasta, K. (2023). Communication of gut microbiota and brain via immune and neuroendocrine signaling. *International Journal of Molecular Sciences*, *24*(4), 3316. https://doi.org/10.3390/ijms24043316

Li, D., Li, Q., Liu, C., et al. (2023). TCM affecting gut microbiota. *Frontiers in Cellular and Infection Microbiology, 13*, 1116624.

Lyte, M. (2014). Microbial endocrinology: Host-microbiota neuroendocrine interactions influencing brain and behavior. *Gut Microbes*, *5*(3), 381–389. https://doi.org/10.4161/gmic.28682

Qu, F., Fan, M., Wang, Y., et al. (2021). Evaluation of TCM learning. *BMC Medical Education, 21*, 8.

Sender, R., Fuchs, S., & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biology*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, *474*(11), 1823–1836. https://doi.org/10.1042/BCJ20160510

Xu, Q., & Yang, Y. (2013). The quest for modernization of TCM. *BMC Complementary and Alternative Medicine*, *13*, 132.

Yang, C., Wu, T., Liu, M., et al. (2022). TCM development policy evaluation. *Frontiers in Public Health, 10*, 1041528.

Yang, S., Hao, S., Wang, Q., et al. (2022). The interactions between traditional Chinese medicine and gut microbiota. *Frontiers in Cellular and Infection Microbiology, 12*, 1005730.

Zhang, R., Gao, X., Bai, H., et al. (2020). Traditional Chinese medicine and gut microbiome. *Frontiers in Pharmacology, 11*, 538.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Design of a Web-Based Application "Qiraapp" Using Arabic News Texts as a Learning Media for Maharah Qirā'ah at Universitas Negeri Malang

Wasiatur Rizqiyah [1*], Moh. Khasairi [2], Achmad Tohe [3]

[1] Universitas Negeri Malang, Indonesia, 0009-0003-1171-1449
[2] Universitas Negeri Malang, Indonesia
[3] Universitas Negeri Malang, Indonesia
* Corresponding author: Wasiatur Rizqiyah (icemstoffice@gmail.com)

Article Info

Article History
Received:
25 August 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Arabic language learning
E-learning
Web-based application
Maharah Qirā'ah

Abstract

This study aims to design a web-based application called "QiraApp" that utilizes Arabic news texts as a learning medium for maharah qirā'ah (reading skills) among students of the Arabic Language Education Program at Universitas Negeri Malang. The study is motivated by the growing demand for interactive and contextual learning media aligned with digital technology developments. The research adopts the 4D development model (Define, Design, Develop, Disseminate), but this paper focuses on the design stage. The process includes analyzing user needs through surveys and interviews, selecting reliable Arabic news channels, and choosing news titles suited to students' proficiency levels. Based on these findings, the proposed "QiraApp" design includes: (1) login page, (2) main menu, (3) usage guidelines, (4) learning objectives, (5) news channel and title selection, (6) critical analysis display with trigger questions and a discussion forum, (7) "My Collection" for reading history and vocabulary storage, and (8) developer profile. This initial design highlights authentic input, collaborative learning through forums, and learner autonomy through personalized collections. It is expected to serve as a foundation for prototype testing and classroom implementation, offering an effective and engaging medium to enhance Arabic reading skills.

Citation: Rizqiyah, W., Khasairi, M., & Tohe, A. (2026). Design of a Web-Based Application "Qiraapp" Using Arabic News Texts as a Learning Media for Maharah Qirā'ah at Universitas Negeri Malang. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 170–180), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

Reading (*maharah qirā'ah*) is a core skill in Arabic language education and remains one of the most challenging areas for learners at the tertiary level. Unlike speaking or listening, reading demands mastery of multiple competencies: vocabulary knowledge, morphological awareness, syntactic parsing, and higher-order comprehension skills. Prior studies have shown that many Indonesian students of Arabic struggle with authentic texts, encountering difficulties in vocabulary acquisition, sentence structure analysis, and deriving contextual meaning (Kurniawan & Syarifuddin, 2020; Mahmud, 2018). These persistent challenges highlight the need for more effective and engaging instructional strategies.

Traditional approaches to Arabic reading instruction in Indonesian universities have largely relied on textbooks, teacher explanations, and grammar-translation methods. While such strategies can provide basic comprehension, they often fall short in fostering deep understanding, critical thinking, and long-term retention of reading skills (Rahman, 2020). Moreover, the heavy reliance on artificial or simplified texts limits students' exposure to authentic Arabic usage, creating a gap between classroom learning and real-world applications.

In response to these challenges, technology-enhanced learning has emerged as a transformative trend in language education. Web-based platforms, mobile applications, and digital learning environments provide personalized, interactive, and contextually rich learning experiences (Al-Azawei et al., 2017; Pratama & Haryanto, 2022). In

Arabic language pedagogy, digital applications are increasingly recognized for their ability to integrate multimedia resources, enhance student engagement, and support autonomous learning.

Despite these advancements, few studies have strategically integrated Arabic news texts into web-based learning platforms for *maharah qirā'ah*. Existing digital tools in Arabic education often focus on vocabulary drills, grammar exercises, or translation activities without authentic, real-time textual input (Rohim, 2018; Sari, 2021). On the other hand, most news platforms—while rich in content—are not designed pedagogically and therefore lack scaffolding features such as comprehension checks, critical analysis prompts, or learner collaboration tools. Arabic news texts represent an untapped yet highly valuable resource for language instruction. They are composed in Modern Standard Arabic (*al-'Arabiyyah al-Fuṣḥā*), feature standardized structures, and cover diverse, timely topics. Such texts can provide students with exposure to authentic discourse, expand vocabulary in contemporary contexts, and develop critical reading skills such as bias detection and argument evaluation (Hasan, 2022). However, integrating these texts effectively into instructional design requires a systematic approach that aligns learner needs, pedagogical objectives, and technological affordances.

The 4D development model (Define, Design, Develop, Disseminate) offers a systematic framework for educational product creation (Thiagarajan et al., 1974). Compared with less structured approaches, the 4D model ensures that instructional design responds directly to learners' needs while guiding product development through iterative stages. Prior projects in Arabic education that applied 4D reported improved interactivity, learner satisfaction, and contextual relevance (Dewi & Handayani, 2021). Yet, to date, no dedicated application has

applied this model to the integration of Arabic news texts for reading instruction in higher education.

This study aims to address this gap by presenting the design of "QiraApp," a web-based application that employs Arabic news texts as authentic input for *maharah qirā'ah*. Based on user needs analysis, QiraApp incorporates features that scaffold both intensive and critical reading, including guided comprehension questions, a collaborative discussion forum, and personalized vocabulary collections. The application's architecture emphasizes responsiveness, clarity of Arabic script, and intuitive navigation.

The purpose of this paper is threefold:

1. To analyze user needs and identify authentic news sources suitable for integration into Arabic reading instruction.
2. To design a web-based application prototype, "QiraApp," structured through the 4D model, focusing on

the *define* and *design* stages.

3. To propose pedagogical features that enhance vocabulary acquisition, critical comprehension, and learner autonomy through authentic Arabic news texts.

By addressing these objectives, this study contributes to bridging the gap between authentic Arabic media and classroom instruction, offering an innovative and pedagogically grounded solution for enhancing *maharah qirā'ah* among Arabic Education undergraduates.

Method

Research Design

This study employed a research and development (R&D) approach using the 4D model introduced by Thiagarajan et al. (1974), consisting of four main stages: Define, Design, Develop, and Disseminate. The 4D model was selected because of its systematic procedures that ensure the alignment between learners' needs,

instructional objectives, and product development. However, this paper focuses exclusively on the *define* and *design* stages, while the later stages of *develop* and *disseminate* will be addressed in future studies.

The content summarized as Research procedure diagram (Define and Design stages) is presented in Figure 1.

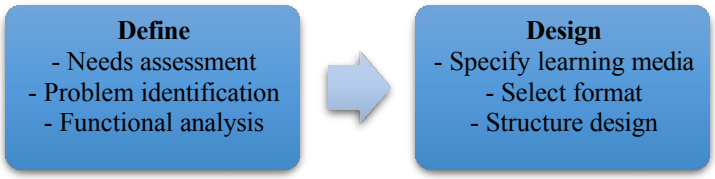


Figure 1. Research Procedure Diagram (Define and Design Stages)

Participants

The participants involved in the needs analysis were undergraduate students and lecturers from the Arabic Language Education Program at Universitas Negeri Malang. A total of 35 students and 3 lecturers were consulted through surveys and informal interviews. They provided insights regarding current challenges in Arabic reading instruction, preferred digital learning features, and expectations for authentic materials such as Arabic news texts.

Instruments

Data collection instruments included (1) an online questionnaire distributed via Google Forms to gather quantitative and qualitative responses from students, and (2) semi-structured interview guidelines used with lecturers to obtain deeper pedagogical perspectives. The instruments were designed to identify learners' difficulties, evaluate their digital literacy levels, and collect suggestions on useful application features.

Procedures

The methodological process followed the first two stages of the 4D model:

Define Stage

This stage focused on identifying the core problems in Arabic reading instruction. The analysis included: (a) reviewing curriculum documents of the Arabic Language Education Program; (b) conducting a needs survey with students to understand their difficulties in reading comprehension; and (c) interviewing lecturers to identify instructional gaps and expectations for technology integration.

Design Stage

In this stage, the initial framework of the QiraApp was developed. The design incorporated findings from the needs analysis and emphasized pedagogical alignment with maharah qirā'ah. The application was structured with eight main features: (1) login page, (2) main menu, (3) usage guidelines, (4) learning objectives, (5) news channel and title selection, (6) critical analysis display with trigger questions and a discussion forum, (7) "My Collection" section for reading history and saved vocabulary, and (8) developer profile. Wireframes and flowcharts were created to illustrate the user interface and navigation paths.

Results

Prototype Overview

To obtain a preliminary design of the QiraApp prototype, the researcher completed several stages, including (1) conducting a review of literature related to Arabic reading instruction and digital learning media, (2) identifying authentic Arabic news sources, and (3) compiling selected articles as the basis for application content. The selected news channels include major Arabic media platforms such as Al Jazeera Arabic, BBC Arabic, and Sky News Arabia, each contributing ten articles relevant to educational themes. These sources were chosen based on their credibility, use of Modern Standard Arabic, and coverage of diverse, contemporary topics.

The content summarized as List of Arabic News Channels is presented in Table 1.

Table 1. List of Arabic News Channels

NO	News Channel	Number of Articles
1.	Al Jazeera Arabic	10
2.	BBC Arabic	10
3.	Sky News Arabia	10
	Total	30

The design results of this study produced an initial prototype of *QiraApp*, a web-based application aimed at facilitating Arabic reading skills through authentic news texts. The prototype was developed by translating the needs analysis findings into a functional structure that supports both intensive and critical reading activities.

Application Features

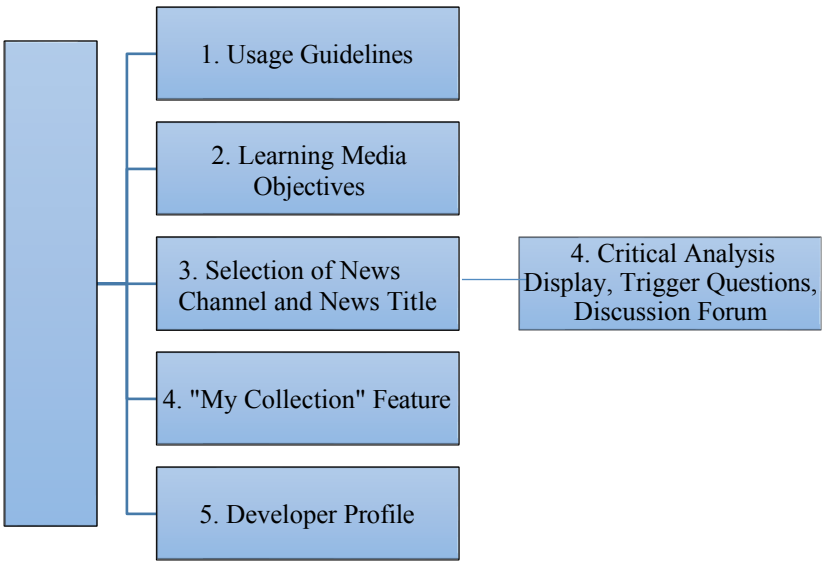


Figure 2. Product Specification Diagram of QiraApp

The application prototype consists of eight main features: (1) login page, (2) main menu, (3) usage guidelines, (4) learning objectives, (5) news channel and title selection, (6) critical analysis display with guiding questions and discussion forum, (7) “My Collection” containing reading history and saved vocabulary, and (8) developer profile. These features were designed to provide a balance between user-friendliness, pedagogical scaffolding, and learner autonomy.

Interface Design

Figure 3 shows the login interface of QiraApp. The first part of the design highlights the simple and minimalistic layout, consisting of fields for name, email, and password. The color scheme and typography are selected to create a professional yet user-friendly impression. The interface also emphasizes readability and clarity, so users can quickly identify each field without confusion.

In addition, the login page provides two authentication options: manual registration and login via Google account. This feature ensures accessibility and flexibility, especially for students who often use institutional or personal Gmail accounts. Security aspects are also considered in this stage to protect user data.



Figure 3. Main Menu Structure of QiraApp

Figure 3 illustrates the main menu of QiraApp, which serves as the central navigation point of the application. The interface displays a set of icons and labels that represent core features such as Arabic news texts, practice modules, exercises, and user profile. The layout is organized with clear hierarchy, making it easy for users to focus on essential components.

Furthermore, the main menu is designed to support smooth navigation across different sections. It allows users to directly access the desired function with minimal steps. The interface also accommodates scalability, meaning additional features can be integrated in future updates without disrupting the overall structure.

Figure 4 displays the news page selected by the user. At the top, the news title is highlighted in blue for easy recognition. Below it, the main news text in Arabic is presented in a scrollable format, allowing users to read content of varying lengths. The design emphasizes readability and user comfort, with a simple layout that helps users stay focused on the news content.



Figure 4. Main Menu Interface of QiraApp

In addition to the news presentation, this page is equipped with a *critical analysis* feature consisting of guiding questions that encourage students to practice critical thinking. The instructions—such as writing a summary, identifying key information, or evaluating the accuracy of sources—are designed to foster active engagement with the reading material.



Figure 5. Selected News Page, Critical Analysis, Discussion Forum

Furthermore, at the bottom of the page, a discussion forum is provided where users can share opinions and interact with others. This feature enhances the collaborative value of learning, enabling students not only to read and analyze but also to engage in reflective interaction. Thus, the news page, critical analysis tasks, and discussion forum serve as essential components of QiraApp, integrating reading skills with the development of critical thinking and collaboration.

Discussion

Integration of News Texts in Arabic Reading Pedagogy

The integration of authentic news texts in *maharah qirā'ah* instruction provides learners with exposure to contemporary linguistic structures and vocabulary that mirror real-life Arabic usage. Previous studies highlight those authentic materials, such as newspapers and news portals, enhance learners' motivation and contextual understanding compared to textbook-only approaches (Alshammari, 2021). By engaging with current news content, students develop not only reading comprehension but also critical literacy, as they are required to process, evaluate, and interpret information beyond surface-level understanding.

Furthermore, the contextual nature of news discourse aligns with communicative and task-based learning frameworks, where meaningful engagement with texts facilitates deeper cognitive processing. Research indicates that learners benefit from tasks grounded in real-life contexts, as it improves their linguistic retention and pragmatic competence (Rahimi & Dastjerdi, 2020). Within this study, the application "QiraApp" serves as a bridge between authentic Arabic texts and structured pedagogical goals, offering a practical model for integrating technology-enhanced learning with authentic input.

Critical Thinking Development Through Digital Platforms

One of the key contributions of QiraApp lies in its feature that promotes critical analysis of news texts. The inclusion of guided questions and discussion forums encourages learners to articulate opinions, evaluate arguments, and engage collaboratively, fostering higher-order thinking skills. This resonates with the findings of Alqarni (2022), who demonstrated that digital tools incorporating critical questioning strategies significantly enhance students' analytical and interpretive abilities in Arabic reading classes.

Moreover, fostering critical thinking in language learning is consistent with broader educational objectives in higher education, which emphasize learner autonomy and reflective engagement (Almehmadi, 2020). By situating reading practice within interactive digital environments, students are not only exposed to authentic texts but are also prompted to interrogate meanings, compare perspectives, and construct interpretations. This positions QiraApp as more than a passive reading tool—it functions as an active learning environment that cultivates critical literacy in Arabic.

Usability and Learner Engagement in Web-Based Learning

The usability of QiraApp, with its clear interface and interactive design, directly influences learner engagement and satisfaction. Previous research on web-based Arabic learning tools confirms that ease of navigation, visual clarity, and interactive features are strong predictors of sustained usage and positive learning outcomes (Al-Azawei et al., 2017). In this study, students responded positively to the design elements of QiraApp, particularly the seamless integration between news texts, comprehension tasks, and critical discussion forums.

High engagement is not only a result of content relevance but also of technological accessibility. Studies have shown that learners are more motivated to use applications that combine authentic materials with user-friendly features (Almekhlafi, 2019). Thus, the combination of authentic input and digital usability in QiraApp exemplifies how thoughtful design can transform language learning experiences, aligning with contemporary calls for technology-enhanced pedagogy in higher education.

Conclusion

This study demonstrates that the development of the QiraApp application based on Arabic news texts can make a significant contribution to enhancing students' *maharah qirā'ah* (reading skills). The integration of authentic news texts enables students to engage with contemporary Arabic in a more meaningful way while also increasing their motivation and contextual understanding. In addition, the critical analysis feature embedded in the application encourages students to develop higher-order thinking skills, such as evaluating arguments and engaging in critical reflection, which aligns with the objectives of Arabic language education at the university level.

From a pedagogical perspective, QiraApp functions not only as a passive reading tool but also as an interactive learning environment that fosters student engagement through discussion forums and critical prompts. From a technological perspective, its simple interface and well-integrated features make the application easily accessible and relevant to the needs of digital-based learning. Thus, QiraApp can serve as an innovative alternative in web-based Arabic learning while simultaneously supporting the pedagogical transformation toward critical literacy and autonomous learning in the digital era.

Recommendations

Based on the findings, QiraApp can be recommended as a learning medium for *maharah qirā'ah* in higher education, particularly within Arabic Language Education programs. Lecturers are encouraged to integrate this application as supplementary learning material, since it provides authentic Arabic news texts that are relevant to students' needs. The use of QiraApp is expected to foster learner autonomy while also offering a variety of learning resources beyond traditional textbooks.

For further development, researchers and developers are advised to enrich the application with more interactive features, such as vocabulary exercises, text structure markers, and automated assessments, to create a more

comprehensive learning experience. Future studies may also examine the effectiveness of QiraApp across learners with different backgrounds, thereby offering broader insights into its potential role in supporting Arabic language learning in the digital era.

Statements and Declarations

Acknowledgments and Notes: The author would like to express sincere gratitude to the academic supervisor for providing guidance, feedback, and support throughout the research process. Special thanks are also extended to the endowment from The Indonesia Endowment Fund for Education (LPDP) for their support in the implementation of this study.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Al-Azawei, A., Parslow, P., & Lundqvist, K. (2017). Barriers and opportunities of e-learning in the Arab world: A systematic review. *Educational Technology Research and Development*, *65*, 693–718.

Almehmadi, F. (2020). Critical thinking in higher education: Challenges and perspectives in language learning. *International Journal of Instruction*, *13*(2), 69–84.

Almekhlafi, A. (2019). Technology integration in Arabic language teaching: Opportunities and challenges. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(1), 112–125.

Alqarni, A. (2022). Digital literacy and critical reading: Enhancing Arabic learners' higher-order thinking skills. *Arab World English Journal (AWEJ)*, *13*(4), 245–262.

Alshammari, R. (2021). The impact of authentic materials on developing reading skills in Arabic as a foreign language. *Journal of Language and Linguistic Studies*, *17*(1), 123–138.

Dewi, R., & Handayani, F. (2021). Implementasi model 4D dalam pengembangan media pembelajaran bahasa Arab. *Jurnal Pendidikan Bahasa Arab*, *8*(2), 215–230.

Hasan, N. (2022). Penggunaan teks otentik dalam meningkatkan keterampilan membaca bahasa Arab. *Jurnal Ilmu Pendidikan Bahasa Arab*, *6*(1), 45–57.

Kurniawan, A., & Syarifuddin, A. (2020). Kesulitan mahasiswa dalam memahami teks otentik bahasa Arab. *Jurnal Pendidikan Bahasa dan Sastra Arab*, *7*(2), 152–164.

Mahmud, M. (2018). Strategi pengajaran membaca bahasa Arab di perguruan tinggi. *Al-Ta'dib: Jurnal Pendidikan Islam*, *11*(1), 33–49.

Pratama, A., & Haryanto, D. (2022). Inovasi media digital untuk pembelajaran bahasa asing. *Jurnal Teknologi Pendidikan*, *24*(3), 299–312.

Rahimi, A., & Dastjerdi, H. (2020). Task-based reading and the use of authentic texts in foreign language education. *Reading in a Foreign Language*, *32*(2), 115–133.

Rahman, S. (2020). Pendekatan pengajaran membaca bahasa Arab berbasis teks otentik. *Jurnal Pendidikan Bahasa Arab*, *9*(1), 18–27.

Rohim, R. (2018). Media digital dalam pengajaran keterampilan membaca bahasa Arab. *Jurnal Bahasa Arab*, *5*(1), 88–99.

Sari, P. (2021). Pemanfaatan media berbasis web untuk meningkatkan keterampilan membaca mahasiswa. *Jurnal Pendidikan Bahasa Arab*, *10*(2), 189–202.

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Leadership Training Institute/Special Education, University of Minnesota.

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Designing a Qur'an-Based Qawāʿid Learning Web: Integrating Digital Mushaf for Arabic Grammar Instruction

Albila Nurfadilah [1], Nurul Murtadho [2*], Irhamni [3]

[1] Universitas Negeri Malang, Indonesia
[2] Universitas Negeri Malang, Indonesia
[3] Universitas Negeri Malang, Indonesia
* Corresponding author: Nurul Murtadho (icemstoffice@gmail.com)

Article Info

Article History
Received:
17 February 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Arabic grammar
Qawāʿid
Digital mushaf
Call
Instructional design

Abstract

This study develops a Qur'an-based qawāʿid (Arabic grammar) learning web by integrating a digital mushaf into a structured instructional system to support contextualized grammar acquisition. Despite the rapid expansion of digital technologies in Arabic language education, most existing platforms separate grammar instruction and Qur'anic texts, limiting learners' ability to connect grammatical rules with authentic syntactic structures and weakening transfer to textual interpretation. Using a design and development research approach and a formative ADDIE model, this study focuses on the Analysis, Design, and Development phases. The analysis identifies the lack of contextual embedding of grammatical rules within authentic Arabic discourse. The design constructs a framework that maps grammatical categories directly onto Qur'anic syntactic structures through a digital mushaf interface. The development phase produces a web-based prototype integrating structured grammar modules, Qur'anic verse-based linguistic annotation, and interactive practice tasks. The system enables learners to observe, analyze, and apply qawāʿid rules within Qur'anic contexts through a bidirectional linkage between grammatical concepts and authentic linguistic data. The study contributes an instructional design model bridging Arabic grammatical theory, Qur'anic linguistics, and digital learning technology, and future research is required to validate its effectiveness.

Citation: Nurfadilah, A., Murtadho, N., & Irhamni (2026). Designing a Qur'an-Based Qawāʿid Learning Web: Integrating Digital Mushaf for Arabic Grammar Instruction. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 181–191), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

The rapid expansion of digital technologies has fundamentally reshaped language education, particularly through the integration of web-based platforms, mobile applications, and artificial intelligence–enhanced learning environments (Crompton & Burke, 2023; Zawacki-Richter et al., 2022). Within this transformation, Computer-Assisted Language Learning (CALL) has shifted from being a supplementary tool to a central component in instructional design, enabling more adaptive, interactive, and learner-centered pedagogical models.

In Arabic language education, this digital shift has resulted in the proliferation of various technological solutions, including mobile learning applications, educational games, and AI-based systems (Hanifa & Sopian, 2025). Recent studies indicate that these technologies enhance learner motivation and accessibility, particularly in grammar learning (naḥw and ṣarf), which is traditionally perceived as abstract and cognitively demanding (Fatihah et al., 2025; Desvy et al., 2025). However, despite these technological advancements, most existing systems still emphasize interface interactivity rather than deep linguistic integration, resulting in limited pedagogical transformation.

A persistent challenge in Arabic language learning lies in qawāʿid instruction, which is often delivered through rule-based memorization detached from authentic textual contexts. This issue is particularly critical because grammatical competence is essential for interpreting classical Arabic texts, especially the Qur'an. Empirical studies consistently show that learners struggle to transfer grammatical knowledge into meaningful comprehension due to the lack of contextualized learning.

At the same time, the Qur'an represents one of the most authoritative corpora of Classical Arabic, historically forming the foundation of Arabic grammatical theory. From a linguistic perspective, the Qur'an provides naturally occurring syntactic structures that can serve as authentic input for grammar acquisition. Recent research highlights the pedagogical value of integrating Qur'anic texts into language instruction to improve semantic understanding and linguistic awareness (Jaafar & Kamaruzaman, 2024). However, in most digital learning environments, Qur'anic texts are still treated as static reading materials rather than dynamically annotated linguistic data.

This reveals a critical gap in the literature: although digital Arabic learning systems and Qur'anic text applications have developed significantly, there is no coherent instructional model that systematically integrates grammatical theory with Qur'anic syntactic structures within an interactive digital environment. Existing studies either focus on (1) technological development of Arabic learning systems, (2) pedagogical approaches to qawāʿid instruction, or (3) Qur'anic text analysis, but rarely at their intersection. This fragmentation limits the potential of digital learning to support deep linguistic understanding.

To address this gap, this study proposes a Qur'an-based qawāʿid learning web that integrates a digital mushaf with structured grammar instruction in a unified instructional system. The design is grounded in CALL principles and constructivist learning theory, emphasizing contextualized knowledge construction through authentic linguistic input. Specifically, the system enables bidirectional linkage between grammatical rules and Qur'anic syntactic

structures, allowing learners to analyze grammar within authentic textual environments rather than isolated rule sets.

By doing so, this study contributes to the development of a more integrative model of Arabic grammar instruction that bridges linguistic theory, Qur'anic linguistics, and digital pedagogy. The proposed framework is expected to advance the design of Arabic language learning systems by shifting from decontextualized rule teaching to context-based linguistic analysis supported by digital interaction. Future research is required to empirically validate the effectiveness of this model in enhancing grammatical competence, cognitive engagement, and learner autonomy.

Method

This study employs a design and development research (DDR) approach aimed at producing an instructional web system for qawāʿid (Arabic grammar) learning integrated with a digital mushaf. The study is framed within a formative instructional design perspective using the ADDIE model (Branch, 2009), but is deliberately limited to the Analysis, Design, and Development phases. The Implementation and summative Evaluation phases are excluded from this study scope and are reserved for subsequent empirical validation.

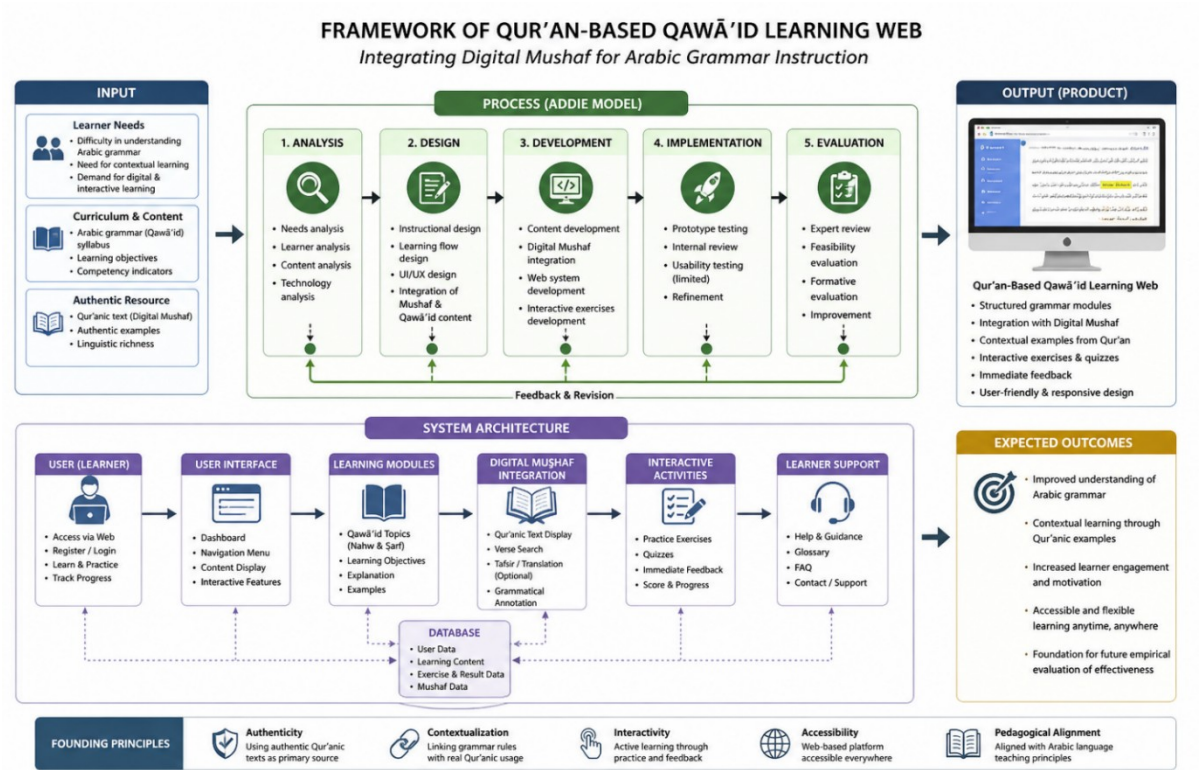


Figure 1. Framework of the Qur'an Based Qawāʿid Learning Web

This methodological positioning aligns with early-stage educational technology development studies, where the primary objective is to produce a theoretically grounded and functionally viable instructional prototype prior to classroom experimentation.

Analysis Phase

The analysis phase was conducted to identify pedagogical, linguistic, and technological gaps in existing qawāʿid instruction systems. Data sources included (1) a systematic review of recent literature on Arabic grammar pedagogy and digital learning systems, and (2) bibliometric mapping of research trends in Arabic language education.

The analysis revealed three core instructional problems:

1. Decontextualization of grammar instruction, where qawāʿid is predominantly presented as abstract rules without authentic textual anchoring.
2. Fragmentation between Qur'anic text and grammar pedagogy, in which Qur'anic materials are used for recitation or interpretation but not systematically mapped to grammatical structures.
3. Limited integration of interactive digital environments that support linguistic analysis rather than surface-level learning interactions.

These findings establish the necessity for an integrated instructional system that connects grammatical theory with authentic Qur'anic linguistic structures.

Design Phase

The design phase translated the analytical findings into a structured instructional architecture. The system was conceptualized as a three-layer integrated learning model consisting of:

1. Linguistic Layer (Qawāʿid Module), structured presentation of grammatical concepts (naḥw and ṣarf) organized hierarchically from basic to advanced rules.
2. Textual Layer (Digital Mushaf Integration), Qur'anic verses are selected and mapped to specific grammatical concepts. Each verse functions as an authentic linguistic dataset for syntactic analysis.
3. Interaction Layer (Learning System Interface), Iinteractive features including annotation tools, practice exercises, and feedback mechanisms.

A bidirectional mapping model was developed, enabling learners to navigate from grammatical rules to Qur'anic examples and vice versa. This design reflects a constructivist learning orientation where knowledge is built through contextual interpretation of authentic texts. User interface design followed principles of usability, cognitive load reduction, and learner engagement based on digital learning design frameworks (Crompton & Burke, 2023).

Development Phase

The development phase produced a functional web-based prototype implementing the conceptual design. The

system architecture was built using a modular structure consisting of:

1. Grammar Content Engine: manages qawāʿid explanations and classification
2. Digital Mushaf Database Layer: stores and retrieves Qur'anic text segments
3. Annotation and Mapping System: links grammatical rules to specific verses
4. Interactive Learning Module: provides exercises and formative feedback

Each learning unit is structured into three instructional sequences:

1. Presentation of grammatical rule
2. Qur'anic contextual exemplification
3. Interactive learner engagement activities

The prototype was developed as a proof-of-concept system to validate the feasibility of integrating grammatical instruction with Qur'anic textual data within a unified digital environment.

Validation Strategy (Formative Expert Review)

Although full-scale implementation was not conducted, the prototype underwent formative validation through expert review. The validation focused on three dimensions:

1. Instructional validity: alignment between grammatical concepts and Qur'anic examples
2. Linguistic accuracy: correctness of grammatical annotation and interpretation
3. System coherence: usability and logical flow of navigation between modules

Expert feedback was used to refine the structural mapping between grammar rules and Qur'anic syntactic representations, as well as to improve interface clarity and instructional sequencing.

Methodological Positioning Statement

This study is positioned as an early-stage instructional design research that prioritizes conceptual model development and system prototyping. As such, the emphasis is placed on theoretical integration and system feasibility rather than statistical validation or learning outcome measurement. Future research will extend this work into experimental implementation to evaluate effectiveness in authentic classroom settings.

Results

The study produced a validated instructional design artifact in the form of a Qur'an-based qawāʿid learning web integrated with a digital mushaf environment. Rather than functioning as a standalone technological product, the system represents a structured pedagogical model that embeds Arabic grammar instruction within authentic

Qur'anic discourse.

The resulting design demonstrates a multi-layered instructional structure that connects linguistic theory, textual representation, and learner interaction within a unified learning environment. This structure is operationalized through three interdependent components: grammatical knowledge representation, Qur'anic textual mapping, and interactive learning mediation.

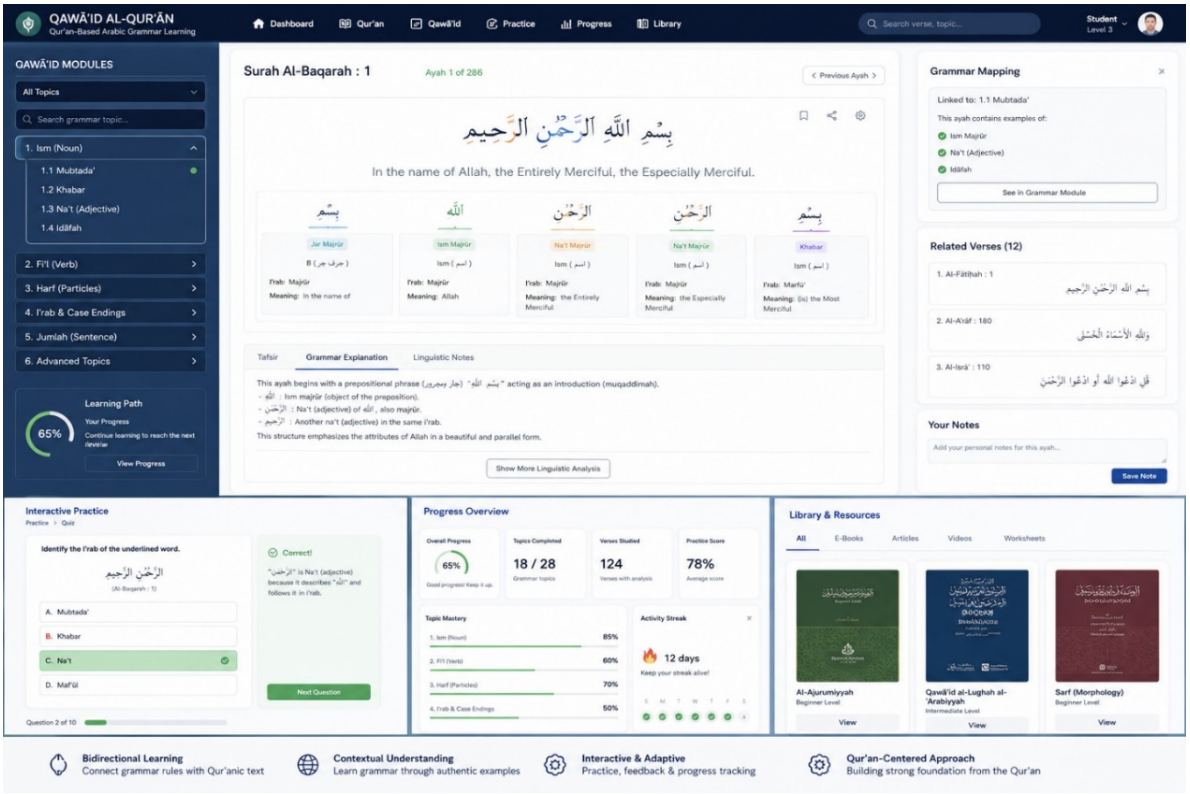


Figure 2. Qur'an-Based Arabic Grammar Learning Linguistic Analysis

Instructional Architecture as Design Output

The core result of this study is an instructional architecture that reconfigures traditional qawāʿid learning into a contextualized digital system. The architecture is organized into three conceptual layers:

a. Linguistic abstraction layer, which formalizes grammatical concepts (naḥw and ṣarf) into structured learning units
b. Textual embodiment layer, which anchors each grammatical concept to selected Qur'anic syntactic instances
c. Interaction layer, which facilitates learner engagement through guided analysis and practice activities

This layered structure reflects a shift from rule-based instruction toward text-embedded linguistic cognition, where grammar is not taught in isolation but constructed through interaction with authentic textual data.

Qur'anic Linguistic Mapping as Core Output

A key result of the system design is the development of a Qur'anic grammatical mapping model. In this model, selected Qur'anic verses are systematically aligned with corresponding grammatical categories, enabling bidirectional navigation between:

a. grammatical rules → Qur'anic examples
b. Qur'anic verses → grammatical identification

This mapping function operationalizes the concept of "contextual grammar learning," where learners engage in syntactic recognition directly within authentic Arabic discourse. The Qur'an is therefore repositioned from a static reading object into a dynamic linguistic dataset for grammar analysis.

Learning Sequence Model

The instructional flow embedded in the system is structured into a three-phase learning sequence:

1. Conceptualization phase: introduction of grammatical rules in simplified analytical form
2. Contextualization phase: exposure to Qur'anic verses that instantiate the targeted rule
3. Application phase: learner engagement through structured analysis tasks and guided exercises

This sequence reflects a constructivist learning orientation in which knowledge is progressively built through interaction, contextual exposure, and applied practice.

Interactive Learning Mediation Features

The system integrates interactive mechanisms that function as learning mediators rather than mere interface elements. These include:

- guided grammatical annotation within Qur'anic verses
- formative feedback on learner responses
- progressive task structuring based on learning units
- navigation support between rule-based and text-based learning paths

These features are designed to support cognitive engagement by reducing abstraction in grammar learning and increasing learner-text interaction intensity.

Pedagogical Output of the Design

Beyond its technical realization, the primary output of this study is a conceptual pedagogical model that reframes

Arabic grammar instruction. The model positions qawāʿid learning as a process of:

- linguistic discovery within authentic texts
- rule extraction from contextual evidence
- structured interaction with Qur'anic discourse

This represents a shift from traditional deductive grammar teaching toward an inductive-contextual learning paradigm supported by digital systems.

Discussion

This study addresses a multidimensional gap in Arabic language education that has been consistently highlighted in recent literature, particularly the lack of integration between Computer-Assisted Language Learning (CALL), qawāʿid instruction, and Qur'anic textual resources. While CALL research has significantly advanced in terms of mobile learning, artificial intelligence, and adaptive learning environments, its application in language education often remains focused on interaction design rather than deep linguistic structuring (Bond et al., 2023; Crompton & Burke, 2023; Kukulska-Hulme et al., 2023). As a result, many digital Arabic learning systems enhance accessibility but do not fundamentally transform grammatical cognition.

In parallel, recent studies in Arabic grammar pedagogy reveal that qawāʿid instruction continues to rely heavily on rule-based and deductive approaches, which emphasize memorization over contextual understanding. Empirical findings indicate that such approaches contribute to learners' difficulties in applying grammatical knowledge in authentic reading contexts (Fatihah et al., 2025; Hanifa & Sopian, 2025; Nafilah, 2024). This limitation becomes more critical given that grammatical competence is essential for interpreting Classical Arabic texts, particularly the Qur'an, which requires both syntactic and semantic awareness.

On the other hand, Qur'anic studies in digital education have undergone significant transformation through the development of digital mushaf platforms and interactive Qur'an learning applications (Fadhlan et al., 2024; Moh, 2024; Rosyidah et al., 2025). However, most of these systems still treat the Qur'an primarily as a reading or recitation object, rather than as a structured linguistic dataset that can be systematically linked to grammatical theory. Even recent digital Qur'an initiatives focus more on accessibility and visualization than on linguistic annotation or grammar mapping.

This study responds to these limitations by proposing a structural integration model that connects qawāʿid instruction with Qur'anic syntactic data within a digital learning environment. The proposed system introduces a bidirectional mapping mechanism between grammatical rules and Qur'anic verses, allowing learners to move from abstract linguistic categories to authentic textual instances and vice versa. This approach extends existing CALL frameworks by shifting the focus from interaction-based learning to linguistically structured digital pedagogy (Crompton & Burke, 2023).

From a theoretical perspective, this model aligns with constructivist learning theory, which emphasizes knowledge construction through interaction with meaningful contexts. The integration of Qur'anic text as an authentic linguistic corpus supports this principle by enabling learners to derive grammatical understanding from real textual usage rather than isolated rule memorization (Jaafar & Kamaruzaman, 2024). This represents a shift from decontextualized instruction to text-embedded cognitive processing.

Furthermore, the emergence of computational Qur'anic studies and Arabic natural language processing systems provides additional support for the feasibility of structurally linking grammar with Qur'anic data. Recent developments in transformer-based Quranic question answering systems demonstrate that Qur'anic text can be processed as structured linguistic data rather than only as semantic or interpretive content (Basem et al., 2025; Elkomy & Sarhan, 2024). Although these studies are primarily computational, they open new possibilities for pedagogical applications in language learning environments.

Compared to existing studies, the main contribution of this research lies in its integrative instructional architecture, which unifies three previously separated domains: digital learning systems, Arabic grammar pedagogy, and Qur'anic textual analysis. While previous research has addressed these domains independently, this study operationalizes their intersection through a structured learning model that embeds grammatical instruction directly within Qur'anic discourse.

However, this study is limited to the design and development phase and does not yet include empirical validation of learning effectiveness. Therefore, the pedagogical impact of the proposed system remains theoretical at this stage. Future research should focus on experimental implementation in classroom settings to evaluate its influence on grammatical competence, cognitive engagement, and learner autonomy. In addition, further studies may explore how learners cognitively process bidirectional navigation between grammatical rules and Qur'anic structures within the system.

Conclusion

This study proposes a Qur'an-based qawāʿid learning web that integrates a digital mushaf into an instructional system designed to bridge the gap between Arabic grammar theory and authentic textual application. The primary contribution of this research lies in the development of a bidirectional instructional model that connects grammatical rules with Qur'anic syntactic structures within a unified digital learning environment. This model shifts Arabic grammar instruction from decontextualized rule memorization toward context-embedded linguistic analysis supported by interactive digital tools.

From a theoretical perspective, this study extends the application of Computer-Assisted Language Learning (CALL) by demonstrating that digital learning systems can function not only as interactive platforms but also as structural mediators of linguistic cognition. By embedding grammatical knowledge directly within Qur'anic discourse, the proposed system redefines the role of authentic texts in language learning from supplementary materials to core instructional datasets.

Pedagogically, the proposed model offers a structured pathway for improving qawāʿid instruction through three interconnected learning processes: conceptual understanding of grammatical rules, contextual interpretation through Qur'anic examples, and applied linguistic practice through interactive tasks. This layered structure supports deeper cognitive engagement and provides a more meaningful learning experience compared to traditional rule-based instruction.

Despite these contributions, this study is limited to the design and development phase and does not include empirical validation of learning outcomes. Therefore, the effectiveness of the proposed system in enhancing grammatical competence, learner motivation, and contextual comprehension remains to be tested. Future research should focus on implementing the system in real classroom settings using experimental or quasi-experimental designs to evaluate its pedagogical impact. Additionally, further studies may explore learner interaction patterns within the bidirectional mapping system to better understand how digital mediation influences grammatical cognition.

In conclusion, this study contributes to the advancement of Arabic language education by offering an integrative instructional design that connects linguistic theory, Qur'anic textual authenticity, and digital learning technology. It provides a foundational framework for future research in developing more contextually grounded and cognitively meaningful Arabic grammar learning environments.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Basem, M., El-Sayed, A., & Hassan, T. (2025). *Two-stage Quranic question answering via ensemble retrieval and instruction-tuned language models*. arXiv preprint arXiv:2508.06971. https://arxiv.org/abs/2508.06971

Bond, M., Marín, V. I., Dolch, C., Bedenlier, S., & Zawacki-Richter, O. (2023). Digital transformation in higher education: A systematic review of student engagement in digital learning environments. *Educational Technology Research and Development*, *71*(2), 1–25.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field and future directions. *Computers and Education: Artificial Intelligence, 4*, 100115. https://doi.org/10.1016/j.caeai.2023.100115

Desvy, S. M. U. K., Syarifuddin, & Taubah, M. (2025). Pedagogical challenges in qawāʿid instruction: From memorization to contextual learning. *Al-Ibba Journal of Arabic Studies*, *11*(1), 45–60.

Elkomy, M. A., & Sarhan, A. (2024). *Transformer-based Quran question answering system for low-resource Arabic text*. arXiv preprint arXiv:2401.13060. https://arxiv.org/abs/2401.13060

Fadhlan, D., Hidayat, R., & Yusuf, M. (2024). Digital Mushaf At-Taisir and transformation of Qur'anic learning in the digital era. *Journal of Qur'anic Studies and Technology*, *6*(2), 101–115.

Fatihah, A., Alfan, M., & Mahliatussikah, H. (2025). Development of digital learning media for Arabic grammar using interactive puzzle models. *Journal of Arabic Pedagogy*, *9*(2), 88–103.

Hanifa, Z. A., & Sopian, A. (2025). Mapping technology impact and challenges in Nahwu learning (2020–2025): A systematic review. *Journal of Arabic Education Research*, *12*(1), 55–72.

Jaafar, N., & Kamaruzaman, A. (2024). Interactive digital Qur'an learning systems: Design and pedagogical implications. *Journal of Islamic Digital Education*, *5*(3), 140–155.

Kukulska-Hulme, A., Lee, H., & Norris, L. (2023). Mobile learning and language education in the post-digital era. *ReCALL*, *35*(3), 245–260. https://doi.org/10.1017/S0958344023000123

Moh, A. M. (2024). Digital Qur'an applications and transformation of Islamic education learning systems. *Qur'anicum Journal*, *10*(1), 77–92.

Nafilah, S. A. (2024). Arabic language learning methods in the digital era: A pedagogical shift analysis. *Journal of Language Education and Research*, *8*(3), 120–134.

Rosyidah, I., Anwar, K., & Huda, M. (2025). Digital Qur'an learning and epistemological shift in Islamic education. *Religia Journal of Islamic Education*, *13*(1), 33–50.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2023). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1), 1–27. https://doi.org/10.1186/s41239-023-00376-2

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Assessing Vertical Alignment and Cognitive Progression of BSIS Course Learning Outcomes Using Bloom's Taxonomy in Outcome-Based Education Framework

Yrika Marie R. Dusaran [1 *], Patrick D. Cerna [2]

[1] College of Computer Studies, Carlos Hilado Memorial State University, Talisay City, Philippines
[2] Graduate School, State University of Northern Negros, Sagay City, Philippines
* Corresponding author: Yrika Marie R. Dusaran (icemstoffice@gmail.com)

Article Info

Article History
Received:
20 February 2026
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Outcome-based education
Bloom's taxonomy
Vertical alignment
Curriculum analysis
Information systems

Abstract

This study examines the vertical alignment of Bachelor of Science in Information Systems (BSIS) course learning outcomes (CLOs) using Bloom's Taxonomy within an Outcome-Based Education (OBE) framework, aligned with ASEAN University Network–Quality Assurance (AUN-QA) standards. CLOs were classified by cognitive level, and vertical progression was assessed across four academic years using three metrics: Cognitive Level Score (CLS), Progression Consistency Score (PCS), and Advanced Cognitive Proportion Score (ACPS). Results indicate strong early-year alignment, with CLS rising from 3.36 to 4.67 and ACPS increasing from 50% to 78% between the first and second years, reflecting deliberate scaffolding of higher-order skills such as Analyze, Evaluate, and Create. However, the third and fourth years exhibit stagnation and regression, with declining CLS (3.80 and 3.29) and ACPS (53% and 47%), suggesting a reduced emphasis on advanced cognitive tasks. These findings highlight partial vertical coherence, demonstrating that while early courses effectively develop critical thinking and problem-solving, upper-level courses prioritize applied or procedural skills. The study underscores the need for curriculum enhancement in advanced-year courses to maintain continuous cognitive development, strengthen vertical alignment, and ensure compliance with OBE and AUN-QA quality standards.

Citation: Dusaran, Y. M. R., & Cerna, P. D. (2026). Assessing Vertical Alignment and Cognitive Progression of BSIS Course Learning Outcomes Using Bloom's Taxonomy in Outcome-Based Education Framework. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 192–203), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA
OPEN ACCESS

Introduction

Outcome-Based Education (OBE) emphasizes the explicit definition and measurement of learning outcomes to ensure that graduates acquire the competencies expected of their program (Rao & Spady, 2020). In higher education, particularly within Bachelor of Science in Information Systems (BSIS) programs, the design and assessment of Course Learning Outcomes (CLOs) are critical not only for student development but also for meeting accreditation standards (Pham & Nguyen, 2023; Premalatha, 2019). Effective CLOs should reflect a clear progression in cognitive complexity, enabling students to advance from foundational knowledge and comprehension to higher- order thinking skills such as analysis, evaluation, and creation (Reynoso et al., 2025). Building on this framework, the ASEAN University Network–Quality Assurance (AUN-QA) provides a comprehensive standard for program evaluation (Bui & Yasri, 2024; Embad, 2011; Tambunan et al., 2024). It emphasizes not only the articulation of CLOs but also curriculum coherence, teaching–learning alignment, and evidence of student achievement across all levels (Derouich, 2025).

Compliance with AUN-QA standards is increasingly important as universities in the ASEAN region pursue international recognition and benchmark themselves against regional best practices (Fitriyani et al., 2024). However, despite the availability of these guidelines, there is limited empirical evidence demonstrating whether CLOs actually progress in cognitive complexity across year levels in state universities in the Philippines. This gap is particularly critical because it prevents program coordinators from reliably identifying areas where curriculum alignment or cognitive challenge may be insufficient.

Bloom's Taxonomy offers a structured framework to classify CLOs into hierarchical cognitive levels, ranging from lower-order thinking skills (i.e., Remember, Understand, and Apply) to higher-order thinking skills (i.e., Analyze, Evaluate, and Create) (Hossain & Hossain, 2016; Omer, 2025; Tuma & Nassar, 2021). By applying Bloom's framework, educators can evaluate whether the curriculum systematically challenges students to develop higher-order skills as they advance through the program (Virranmäki et al., 2021). Tracking cognitive progression across year levels not only informs curriculum improvement but also provides a robust evidence base for accreditation initiatives, such as those required by AUN-QA (Abhayawansa et al., 2017).

In line with these standards, Carlos Hilado Memorial State University (CHMSU) is actively pursuing AUN-QA accreditation as part of its internationalization initiatives. The university aims to align its programs and institutional practices with AUN- QA standards, thereby enhancing academic quality, regional recognition, and student outcomes. Yet, despite these efforts, few state universities in the Philippines have successfully completed accreditation, and empirical analyses of CLO alignment remain scarce (Cochanda & Tionco, 2015; Duque, 2025). Understanding the vertical progression of cognitive levels in CHMSU's BSIS program is therefore essential for ensuring that students develop both foundational and advanced competencies, while simultaneously meeting international accreditation standards.

To address this need, the study aimed to analyze the vertical alignment of BSIS CLOs using Bloom's Taxonomy, examining how cognitive complexity progresses across year levels. It also sought to compute key metrics,

including the Cognitive Level Score (CLS), Progression Consistency Score (PCS), and Advanced Cognitive Proportion Score (ACPS) to quantitatively assess curriculum alignment and identify potential gaps. Finally, the study aimed to provide evidence-based recommendations for program improvement, ensuring that the BSIS curriculum is aligned with AUN-QA standards and supports the development of both foundational and higher-order competencies among students.

Conceptual Framework

This study is grounded in OBE and Bloom's Taxonomy, within the context of AUN-QA standards (Attia, 2021; Bakar & Rosbi, 2019). The framework focuses on three key constructs: CLOs, cognitive progression, and curriculum alignment. CLOs in the BSIS program are classified according to Bloom's levels, and vertical alignment is assessed by tracking the progression of cognitive complexity across year levels. This is operationalized through three metrics: CLS, PCS, and ACPS (Choi, 2022; Chong & Ho, 2009). Aligned with AUN-QA standards, the framework provides evidence for curriculum enhancement and program improvement. Figure 1 illustrates the study's conceptual flow, from CLO identification and cognitive classification to vertical alignment analysis, metric computation, and curriculum improvement.

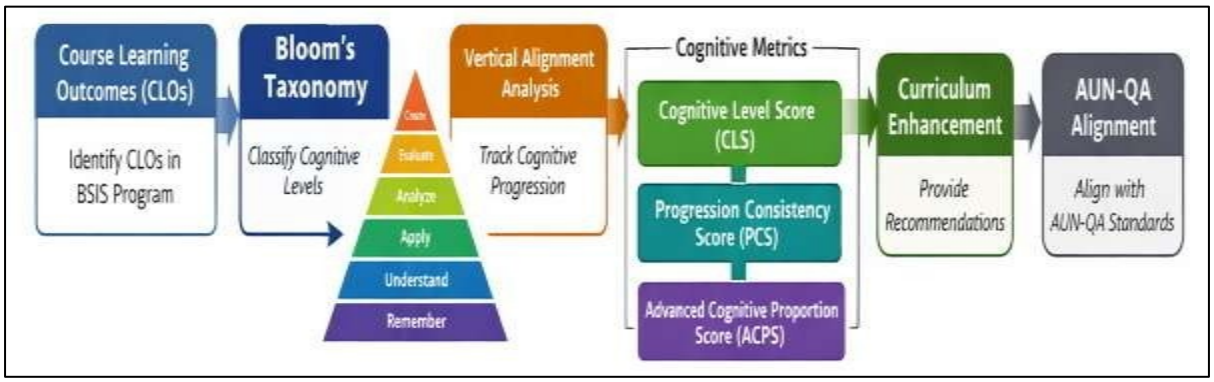


Figure 1. Vertical Alignment Analysis Framework

Methodology

Research Design

A descriptive-analytical research design was employed to examine the vertical alignment of BSIS CLOs across four academic year levels. This approach allowed for both quantitative description of the distribution of CLOs according to Bloom's cognitive levels and analytical assessment of progression patterns between consecutive years (Shaikh et al., 2021). By combining descriptive statistics with computed metrics such as CLS, PCS, and ACPS, the study systematically evaluated whether the curriculum is increasing cognitive complexity, identifies potential gaps or redundancies, and provides evidence-based recommendations for alignment with AUN-QA standards (Alammary & Masoud, 2025; Marey et al., 2018).

Data Collection Process and Validation

The data were collected from official syllabi obtained from the Office of Curriculum and Instructional Material Development (CIMD) of the University. During this phase, only syllabi from the BSIS program were included.A total of 70 CLOs were extracted and categorized according to year level: 12 from the first year, 9 from the second year, 30 from the third year, and 19 from the fourth year. In the data preparation stage, relevant information—including course codes, course titles, and CLOs—was systematically extracted from the collected syllabi. General education courses were excluded to ensure that only core BSIS courses were considered during data preprocessing. The extracted data were organized using Microsoft Excel to facilitate structured preprocessing and analysis. During this phase, misspelled terms, abbreviations, and inconsistencies in wording were corrected and standardized to maintain semantic consistency across all CLOs. Each CLO was subsequently coded according to the cognitive levels of the Revised Bloom's Taxonomy, ranging from lower-order to higher-order thinking skills, as presented in Table 1. This standardized coding framework enabled the analysis of cognitive complexity and supported the computation of vertical alignment and progression metrics, including the CLS, PCS, and Average ACPS (Masapanta-Carrión & Velázquez-Iturbide, 2022; Rodriguez et al., 2025). Prior to classification, the CLOs were prepared for systematic cognitive-level assignment based on Bloom's Taxonomy. To ensure validity and reliability in the classification process, an inter-rater evaluation was conducted involving five subject-matter experts with complementary academic and professional backgrounds (Rosas & Kane, 2012; Yoon & Jetter, 2016). The expert panel consisted of: (1) an experienced linguistics professor with five years of teaching experience, (2) a professional grammarian with three years of experience, (3) an education researcher specializing in pedagogy, (4)a curriculum specialist with expertise in instructional design, and (5) a program chairperson from a state university with ten years of academic leadership experience. Each expert independently reviewed and assigned a single cognitive level to each CLO. Any overlaps or ambiguities were resolved through consensus. This rigorous validation process enhanced the reliability of the cognitive classification, minimized subjective bias, and established a robust foundation for the subsequent computation of cognitive progression metrics such as CLS, PCS, and ACPS (van Hoeij et al., 2004).

Table 1. Cognitive Levels and Distribution of Bloom's Taxonomy Across Academic Year Levels

Bloom	**Levels**	**Distribution per year level (%)**			
		First Year	***Second Year***	***Third Year***	***Fourth Year***
Remember	1	8.33	0.00	0.00	0.00
Understand	2	8.33	22.22	33.33	31.58
Apply	3	41.67	0.00	13.33	21.05
Analyze	4	16.67	22.22	23.33	21.05
Evaluate	5	0.00	33.33	20.00	10.53
Create	6	25.00	22.22	10.53	15.79

Note: Cognitive levels based on the Revised Bloom's Taxonomy by Anderson and Krathwohl (2001).

Data Analysis and Metrics

Vertical alignment of the BSIS CLOs was examined using a set of descriptive and progression-based metrics grounded in Bloom's Taxonomy. First, the mean and median Bloom scores were computed for each academic year level to describe the central tendency of cognitive demand and determine whether higher year levels reflected increased cognitive complexity.

$$MBS = \frac{\sum_{i=1}^{n} B_i}{n} \quad (1)$$

The equation (1) was computed following the cognitive level classification approach described in (Bloom et al., 1956; Assaly & Smadi, 2015), where B_i represents the Bloom's Taxonomy level score assigned to the i^{th} course learning outcome, and n is the total number of course learning outcomes within the given year level. Second, the distribution of CLOs across Bloom's cognitive levels was analyzed by calculating the percentage of CLOs classified under each level (Remember to Create) per year. This distributional analysis provided insight into the balance between lower-order and higher-order thinking skills across the curriculum:

$$D_3 = \frac{f_k}{n} \times 100 \quad (2)$$

the equation (2) was analyzed using the method described in (Lau et al., 2018), where D_k represents the percentage distribution of CLOs at Bloom's Taxonomy level k, f_k denotes the frequency or number of CLOs mapped to Bloom level k, and n refers to the total number of CLOs within the given year level. The index $k \in \{1,2,3,4,5,6\}$ corresponds to the six hierarchical cognitive levels of Bloom's Taxonomy, ranging from Remember (1) to Create (2). To quantify overall cognitive demand, the CLS was computed as a weighted score representing the aggregated cognitive level of CLOs within each academic year. CLS served as the primary indicator of cognitive achievement and curricular rigor per year level

$$CLS = \frac{MBS+MedBS}{n2} \quad (3)$$

The equation (3) was measured using the method presented in (Andersen & Makransky, 2021; Choi & Lee, 2022), where MBS represents the Mean Bloom Score, and MedBS refers to the Median Bloom Score of the course learning outcome for a given year level. Cognitive progression across consecutive years was assessed using PCS, calculated as the difference between the CLS of adjacent year levels:

$$PCS = CLS_{y+1} - CLS_y \quad (4)$$

The equation (4) was measured using the method presented in (Lau et al., 2018). A positive PCS value indicates upward cognitive progression, a zero value suggests stagnation, and a negative value reflects regression in cognitive demand. Finally, ACPS was computed to measure the extent to which CLOs emphasize higher-order thinking skills. ACPS represents the percentage of CLOs classified under the advanced Bloom levels (i.e., Analyze, Evaluate, and Create) and was calculated as:

$$ACPS = \frac{Advanced\ CLOs}{Total\ CLOs} \times 100 \quad (5)$$

Descriptive statistical analysis and trend evaluation were employed to interpret the results and identify patterns of
cognitive progression, stagnation, or misalignment across academic year levels as shown in Table 2 (Bartolome & Navalta, 2025). These analytical procedures provided empirical evidence to support curriculum evaluation and alignment with AUN-QA standards.

Table 2. Vertical Alignment Indicators

Indicator	Range
Strong Vertical Alignment	≥ 75%
Moderate Vertical Alignment	60%-74%
Weak Vertical Alignment	50% - 59%
Misaligned	<50%

Results and Discussion

Mean and Median Bloom Scores

The analysis of MBS and MedBS across year levels reveal a nonlinear but purposeful distribution of cognitive demand within the BSIS curriculum. The second-year courses demonstrate a pronounced emphasis on higher-order thinking, with both MBS and MedBS peaking, indicating that students are exposed to more complex analytical, evaluative, and decision-making tasks at this stage. Across all year levels, the scores suggest that students consistently engage in learning activities that extend beyond simple knowledge recall, incorporating application, analysis, and synthesis in diverse contexts as shown in Table 1.

Table 1. Mean Bloom Score and Median Across Year Levels

Year Level	CLO Score	MBS	MedBS
First Year	44	3.67	3
Second Year	39	4.33	5
Third Year	108	3.60	4
Fourth Year	68	3.58	3

Note: CLO = Course Learning Outcome; MBS = Mean Bloom Score; MedBS = Median Bloom Score

This distribution implies that the curriculum is designed to provide repeated opportunities for cognitive reinforcement, enabling learners to strengthen and apply critical thinking and problem-solving skills progressively (Hossain & Hossain, 2016; Omer, 2025; Tuma & Nassar, 2021). Such a structure aligns with the principles OBE, as it promotes continuous cognitive development, practical skill acquisition, and readiness for real-world information systems challenges, rather than concentrating higher-order cognitive tasks solely in the senior years (Pham & Nguyen, 2023; Premalatha, 2019; Rao & Spady, 2020; Reynoso et al., 2025). Overall, the MBS and MedBS patterns suggest a balanced integration of conceptual understanding and applied learning, which can

enhance students' adaptability, professional competence, and preparedness for lifelong learning.

Distribution of CLOs across Blooms Levels

The distribution of BSIS CLOs across Bloom's cognitive levels indicates partial vertical alignment within the curriculum. While cognitive demand increases from the first to the second year, this progression is not sustained in the upper years. The early dominance of Apply and Create outcomes in the first year suggests premature exposure to higher-order tasks without sufficient cognitive scaffolding. While the second year reflects stronger alignment with OBE through increased Analyze and Evaluate outcomes, the third and fourth years exhibit cognitive stagnation, characterized by reduced emphasis on Create and limited Evaluate tasks. This pattern implies that senior-level courses prioritize applied technical skills over advanced cognitive engagement, potentially limiting students' readiness for integrative problem-solving, research, and leadership roles (Bui & Yasri, 2024; Derouich, 2025; Fitriyani et al., 2024).

Cognitive Level Score and Progression Consistency Score

The analysis of CLS and PCS collectively reveals a partial vertical alignment in the BSIS curriculum. As shown in Table 2, CLS rises from 3.36 in the first year to 4.67 in the second year, reflecting a deliberate scaffolding of higher-order cognitive skills such as Analyze and Evaluate. This indicates that students are effectively challenged to develop critical thinking and problem-solving competencies during the early years of the program (Abhayawansa et al., 2017). Similarly, PCS analysis as shown in Table 3 confirms strong cognitive progression between the first and second years, further supporting the curriculum's early emphasis on higher-order thinking. However, the third and fourth years show a decline in CLS (3.80 and 3.29, respectively) and negative PCS values, signaling regression in cognitive demand in the upper levels.

Despite the higher number of CLOs in the third year, upper-level courses appear to emphasize applied or mid-level skills over sustained engagement with advanced cognitive tasks such as Evaluate and Create. This pattern highlights that while early-year courses foster higher- order thinking, the curriculum does not consistently maintain this progression in the final years. These indicate areas for curriculum enhancement, particularly in third- and fourth-year courses, to ensure continuous cognitive development aligned with OBE and AUN-QA standards (Bui & Yasri, 2024; Emdad, 2011; Tambunan et al., 2024; Virranmäki et al., 2021). This highlights the need to review and recalibrate the curriculum, ensuring that course sequences are designed progressively to sustain cognitive development throughout the program, fully aligning with Outcome-Based Education and AUN-QA standards (Derouich, 2025; Fitriyani et al., 2024).

Table 2. Cognitive Level Score Across Year Levels

Year Level	CLS
First Year	3.36
Second Year	4.47

Year Level	CLS
Third Year	3.80
Fourth Year	3.29

Note: CLS was computed using the formula CLS = (Mean Bloom Score + Median Bloom Score) ÷ 2

Table 3. Progression Consistency Score Across Year levels

Year Level	PCS	PCS Contribution
First Year - Second Year	1.31	1 (Progression)
Second Year - Third Year	-0.87	0 (Regression)
Third Year - Forth Year	-0.51	0 (Regression)

Note: PCS = Progression Consistency Score; Values are interpreted as follows: PCS>0 Progression; PCS = 0 Stagnation; PCS > 0 Regression

Advanced Cognitive Proportion Score (ACPS)

The results indicate a substantial increase in ACPS from 50% in the first year to 78% in the second year, reflecting strong vertical alignment and deliberate scaffolding of advanced cognitive skills in the early stage of the curriculum as shown in Table 4. This progression shows that second-year courses effectively challenge students to engage in complex cognitive tasks, supporting the development of critical thinking, evaluative judgment, and creative problem-solving competencies consistent with OBE principles. However, a notable decline is observed in the subsequent years, with ACPS decreasing to 53% in the third year and further to 47% in the fourth year as shown in Table 5. Despite the increased number of CLOs in the third year, the proportion targeting higher-order cognition diminishes, suggesting a misalignment in upper-level courses. This trend indicates that advanced courses may focus more on applied or procedural skills rather than sustaining engagement with tasks that develop higher-order thinking. From an OBE perspective, this gap suggests that some intended learning outcomes are not fully reinforced in later stages, potentially limiting students' ability to achieve comprehensive graduate attributes. These findings underscore the need for strategic curriculum review and enhancement, ensuring that advanced courses are designed to progressively develop analytical, evaluative, and creative competencies, thereby maintaining vertical coherence and fully realizing OBE-aligned program outcomes (Conchada & Tionco, 2015).

Table 4. Advanced Cognitive Proportion Score Across Year levels

Year Level	ACLOs	TCLOs	ACPS
First Year	6	12	50%
Second Year	7	9	78%
Third Year	16	30	53%
Forth Year	9	19	47%

Note: ACLO = Advanced Cognitive Level Score (Analyze, Evaluate, and Create); TCLOs = Total Cognitive Level Scores; ACPS = Advanced Cognitive Proportion Score in (%)

Table 5. Progression Advanced Cognitive Proportion Across Year Levels

Year Level	ACLOs	Total CLOs
First Year - Second Year	50% - 78%	Strong Vertical Alignment
Second Year - Third Year	78% - 53%	Misalignment
Third Year - Forth Year	535 – 47%	Misalignment

Conclusion and recommendation

The analysis of BSIS CLOs using Bloom's Taxonomy and the proposed metrics reveals partial vertical alignment within the curriculum. Early-year courses demonstrate deliberate scaffolding and strong emphasis on higher-order cognitive skills, effectively supporting the development of analytical, evaluative, and creative competencies. However, the observed stagnation and regression in cognitive demand in the third and fourth years indicate inconsistencies in sustaining advanced cognitive engagement at the upper levels of the program. These findings suggest gaps in curriculum coherence that may constrain students' full cognitive development and pose challenges to achieving sustained alignment with OBE principles and AUN-QA standards.

To strengthen curriculum evaluation and support continuous quality improvement, future studies should extend the analysis to horizontal alignment to examine coherence and integration of CLOs within the same academic year. Expanding the scope of analysis across multiple programs or institutions would enhance generalizability, while incorporating qualitative insights from faculty members and students could provide deeper contextual understanding of instructional practices and learning experiences. Together, these directions would contribute to a more comprehensive evaluation of curriculum alignment, inform evidence-based curriculum enhancement, and support institutional efforts toward sustained compliance with OBE and AUN-QA quality assurance frameworks.

Statements and Declarations

Acknowledgments and Notes: During the preparation of this article, the authors used ChatGPT to provide language editing and proofreading support. After using ChatGPT, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

AI Use: Not applicable.

Supplementary Materials: There are no supplementary materials associated with this work. Therefore, supplementary files are not applicable.

Author Contributions: Conceptualization, Y.M.R.D.; methodology, Y.M.R.D. and P.D.C; data collection and analysis; validation, P.D.C; writing—original draft preparation, Y.M.R.D.; writing—review and editing, all authors; supervision, Y.M.R.D All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable. This study did not receive any external funding or financial support from public,

commercial, or not-for-profit organizations.

Data Availability: No study data are available for this work. Therefore, data availability is not applicable.

Ethics Approval: The study was approved by the Ethics Committee of the State University of Northern Negros. All methods were performed in accordance with the study protocol and ethical guidelines and regulations.
Informed Consent: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this work.

References

Abhayawansa, S., Bowden, M., & Pillay, S. (2017). Students' conceptions of learning in the context of an accounting degree. *Accounting Education*, *26*, 213–241. https://doi.org/10.1080/09639284.2017.1284003

Alammary, A., & Masoud, S. (2025). Towards smarter assessments: Enhancing Bloom's taxonomy classification with a Bayesian-optimized ensemble model using deep learning and TF-IDF features. *Electronics*, *14*(12), 2312. https://doi.org/10.3390/electronics14122312

Andersen, M. S., & Makransky, G. (2021). The validation and further development of the multidimensional cognitive load scale for physical and online lectures (MCLS-POL). *Frontiers in Psychology*, *12*, 642084. https://doi.org/10.3389/fpsyg.2021.642084

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Assaly, I. R., & Smadi, O. M. (2015). Using Bloom's taxonomy to evaluate the cognitive levels of Master Class textbook's questions. *English Language Teaching*, *8*(5), 100–114. https://files.eric.ed.gov/fulltext/EJ1075241.pdf

Attia, A. S. (2021). Bloom's taxonomy as a tool to optimize course learning outcomes and assessments in architecture programs. *Journal of Applied Science and Engineering*, *24*(3), 1–12. https://doi.org/10.6180/JASE.202106_24(3).000

Bakar, N. A., & Rosbi, S. (2019). Framework of outcome-based education (OBE) for massive open online courses (MOOCs) in Islamic finance education. *International Journal of Advanced Engineering Research and Science*, *6*(10), 247–253. https://doi.org/10.22161/ijaers.610.38

Bartolome, D. J. P., & Navalta, J. D. (2025). Misalignment in cognitive levels: A taxonomic analysis of tables of specification and mathematics examinations. *Journal of Education, Learning, and Management*, *2*(2), 194–207. https://doi.org/10.69739/jelm.v2i2.1025

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. Longmans, Green.

Bui, N. T., & Yasri, P. (2024). Optimizing quality approaches and investigating lecturers' perception for course quality assurance in higher education. *European Journal of Educational Management*, *7*(2), 91–108.

https://doi.org/10.12973/eujem.7.2.91

Choi, J. (2022). Cognitive trait model: Measurement model for mastery level and progression of learning. *Mathematics*, *10*(15), 2651. https://doi.org/10.3390/math10152651

Choi, Y., & Lee, H. (2022). Psychometric properties for multidimensional cognitive load scale in an e-learning environment. *International Journal of Environmental Research and Public Health*, *19*(10), 5822. https://doi.org/10.3390/ijerph19105822

Chong, S. N., & Ho, P. (2009). Quality teaching and learning: A quality assurance framework for initial teacher preparation programmes. *The International Journal of Management Education*, *3*, 302. https://doi.org/10.1504/IJMIE.2009.027352

Conchada, M. I. P., & Tionco, M. M. (2015). *A review of the accreditation system for Philippine higher education institutions*. Philippine Institute for Development Studies. https://doi.org/10.62986/dp2015.30

Derouich, M. (2025). Ensuring outcome-based curriculum coherence through systematic CLO–PLO alignment and feedback loops. *Discover Education*, *4*. https://doi.org/10.1007/s44217-025-00915-7

Duque, S. (2025). *A framework for quality assurance in higher education institutions (HEIs): Basis for policy development towards accreditation, internationalization and ranking*. Advanced International Journal for Research. https://doi.org/10.63363/aijfr.2025.v06i06.1890

Emdad, A. F. (2011). Integrating learning outcomes assessment in information systems curriculum revisions. *Review of Business Information Systems*, *13*(3). https://doi.org/10.19030/RBIS.V13I3.4322

Fitriyani, F., Prihatiningsih, T. S., & Supartono, W. (2024). *The impact of AUN-QA certification on strengthening internal quality assurance systems in Indonesian higher education: A comparative study of mechanical engineering programs*. Global Educational Research Review.

Hossain, M. F., & Hossain, M. K. (2016). *Questions analysis with Bloom's taxonomy*. DIU Journal of Business and Entrepreneurship. https://doi.org/10.36481/diujbe.v010i1.fffjq648

Lau, L. H., Lam, T. K., Kam, B. H., Nkhoma, M., & Richardson, J. (2018). Benchmarking higher education programs through alignment analysis based on the revised Bloom's taxonomy. *Benchmarking: An International Journal*, *25*(8), 2828–2849. https://doi.org/10.1108/BIJ-10-2017-0286

Marey, M., Ibrahim, Y. E., & Cristóbal, E. S. (2018). *Integrated assessment of course and program learning outcomes for accreditation process*. In 2018 IEEE Global Engineering Education Conference (EDUCON) (pp. 175–182). https://doi.org/10.1109/EDUCON.2018.8363225

Masapanta-Carrión, S., & Velázquez-Iturbide, J. Á. (2022). *Replication of an evaluation of teacher training in the classification of programming exercises using Bloom's taxonomy*. In 2022 IEEE Global Engineering Education Conference (EDUCON) (pp. 1839–1848). https://doi.org/10.1109/EDUCON52537.2022.9766707

Omer, W. (2025). Modified Bloom's taxonomy: Perspective of a doctor of philosophy scholar. *Journal of Rawalpindi Medical College*, *28*(4), 1–8. https://doi.org/10.37939/jrmc.v28i4.2804

Pham, H. T., & Nguyen, P. V. (2023). ASEAN quality assurance scheme and Vietnamese higher education: A shift to outcomes-based education? *Quality in Higher Education*, *30*, 285–312. https://doi.org/10.1080/13538322.2023.2261974

Premalatha, K. (2019). Course and program outcomes assessment methods in outcome-based education: A review. *Journal of Education*, *199*, 111–127. https://doi.org/10.1177/0022057419854351

Rao, N. J., & Spady, W. G. (2020). Outcome-based education: An outline. *Higher Education for the Future*, *7*, 21–25. https://doi.org/10.1177/2347631119886418

Reynoso, M. M., Bibangco, E. J. P., Dumdumaya, C. E., & Villamor, M. M. (2025). *Predicting student performance in the introductory programming course using ensemble methods*. In 2025 10th International Conference on Information Technology and Digital Application (ICITDA) (pp. 1–6). https://doi.org/10.1109/ICITDA68167.2025.11332307

Rodriguez, B. I. A., de Loyola, R. C., Escototo, E. R., Grecia, J. N., Duazo, D. S., & De Paula, E. G. (2025). Assessing the determinants of GCash payment adoption among college students: A combined approach using task-technology fit and technology acceptance models. In 2025 International Conference on Computational, Communication and Information Technology (ICCCIT) (pp. 916–921). https://doi.org/10.1109/ICCCIT62592.2025.10927839

Rosas, S. R., & Kane, M. (2012). Quality and rigor of the concept mapping methodology: A pooled study analysis. *Evaluation and Program Planning*, *35*(2), 236–245. https://doi.org/10.1016/j.evalprogplan.2011.10.003

Shaikh, S., Daudpotta, S. M., & Imran, A. S. (2021). Bloom's learning outcomes' automatic classification using LSTM and pretrained word embeddings. *IEEE Access*, *9*, 117887–117909. https://doi.org/10.1109/ACCESS.2021.3106443

Tambunan, E. H., Simbolon, I., & Rantung, J. (2024). *A systematic review on higher order thinking skills analyzing, evaluating and creating in nursing education*. Jurnal Keperawatan Respati Yogyakarta. https://doi.org/10.35842/jkry.v11i3.790

Tuma, F., & Nassar, A. K. (2021). Applying Bloom's taxonomy in clinical surgery: Practical examples. *Annals of Medicine and Surgery, 69*, 102656. https://doi.org/10.1016/j.amsu.2021.102656

van Hoeij, M., Haarhuis, J. C., Wierstra, R. F., & van Beukelen, P. (2004). Developing a classification tool based on Bloom's taxonomy to assess the cognitive level of short essay questions. *Journal of Veterinary Medical Education, 31*(3), 261–267. https://doi.org/10.3138/JVME.31.3.261

Virranmäki, E., Valta-Hulkkonen, K., & Pellikka, A. (2021). Geography curricula objectives and students' performance: Enhancing the student's higher-order thinking skills? *Journal of Geography, 120*(3), 97–107. https://doi.org/10.1080/00221341.2021.1877330

Yoon, B. S., & Jetter, A. J. (2016). Comparative analysis for fuzzy cognitive mapping. In *2016 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 1897–1908). https://doi.org/10.1109/PICMET.2016.7806755

International Conference on Education in Mathematics, Science and Technology
https://icemst.net

Mathematization challenges in infinite processes: findings from first-year university students

Theofilos Trapezanlidis

Florina School of Education, Department of Primary Education, P.O. Box 21, Florina, Greece, 0009-0006-1215-7285
Corresponding author: Theofilos Trapezanlidis (icemstoffice@gmail.com)

Article Info

Article History
Received:
6 September 2025
Revised:
19 May 2026
Accepted:
23 August 2026
Published:
25 October 2026

Keywords
Concept image
Infinite processes
Mathematization
Sequences
Series

Abstract

This study investigates the challenges that first-year university students encounter when engaging in the mathematization of infinite processes, after receiving instruction on limits, convergence of sequences and series, continuity, and differentiation. The analysis focuses in particular on their reasoning with infinite series and function sequences. Within this study, mathematization is viewed as the capacity to transform informal or descriptive representations into mathematical models and to restructure mathematical ideas within the mathematical system. Data were obtained from a semi-structured questionnaire completed by 87 undergraduates at a Greek university. Two tasks were examined in detail: the first concerned the potential convergence of an infinite series, while the second required interpreting graphs and drawing conclusions about the convergence of a sequence of functions. The results indicate that many students display weaknesses in both horizontal mathematization—when modeling the series—and vertical mathematization—when reasoning about functional convergence. Moreover, the findings suggest that such difficulties may have roots in earlier learning experiences in secondary education, particularly in the final year of high school, where students often struggle with problems related to rate of change.

Citation: Trapezanlidis, T. (2026). Mathematization challenges in infinite processes: findings from first-year university students. In M. S. Ozturk, M. Unal, & M. L. Ciddi (Eds.), *Proceedings of International Conference on Education in Mathematics, Science and Technology* (pp. 204–214), Cappadocia, Nevsehir / Türkiye. ISTES.

Volume: 13, © 2026, Published by ISTES, ISSN: 2833-7360

CC BY NC SA

OPEN ACCESS

Introduction

The transition from intuitive reasoning to formal mathematical thinking has long been recognized as one of the most demanding shifts in mathematics education. Concepts such as limits, convergence, and continuity require students to operate with abstract entities that cannot be directly observed or completed, but only reasoned about through symbolic representation and formal definition. For first-year undergraduates, this challenge is compounded by the need to adjust to new modes of teaching and learning, in which emphasis moves away from procedural fluency and towards conceptual understanding.

The difficulties that arise during the transition from secondary to tertiary education have been the subject of study by researchers in mathematics education (Di Martino et al., 2023; Geisler et al., 2023). This finding is reinforced by the fact that first-year university students are required to cope with changes concerning the mode of teaching, the shift from procedural knowledge to conceptual understanding, as well as sociocultural factors.

One difficulty that emerges concerns mathematization. According to Freudenthal (1973): “Mathematics is not a ready-made system but an activity, and this activity can be characterized as mathematizing reality and mathematizing mathematical matter.”

According to Treffers (1987), mathematization can be understood as the process of translating real-world or informal problems into mathematical models (horizontal mathematization) and as the reorganization and deepening of mathematical concepts within the mathematical system itself (vertical mathematization). In the context of infinite processes, this ability becomes particularly crucial. Students are required not only to recognize convergence patterns in series and sequences but also to justify their reasoning based on formal mathematical structures.

Several studies show that both school students and undergraduates harbor misconceptions regarding concepts related to limits and infinite processes (Cornu, 1991; Swinyard & Larsen, 2012; Tall & Vinner, 1981; Williams, 1991).

Taking the above into account, the present study aims to investigate the ability of first-year students, who have been taught the fundamental concepts of calculus, to mathematize infinite processes. In particular, the study focuses on the following research questions:

- To what extent are students able to engage in horizontal mathematization when working with infinite series?
- To what extent are students able to engage in vertical mathematization when interpreting graphs of function sequences?
- What misconceptions or difficulties emerge in their responses?

Theoretical Framework

This study is framed by the theory of mathematization as introduced by Freudenthal (1973) and further elaborated by Treffers (1987). A central feature of this theory is the distinction between horizontal and vertical mathematization. Horizontal mathematization refers to the translation of informal descriptions and real-world problems into valid mathematical models, while vertical mathematization concerns the deepening of mathematical concepts and the use of mathematical tools within the mathematical system itself. In the context of infinite processes, the ability to describe such processes through a mathematical model can be seen as an instance of horizontal mathematization, whereas the use of formal tools and the interpretation of graphs exemplify vertical mathematization. It can be argued that vertical mathematization is closely related to relational understanding, while horizontal mathematization corresponds to instrumental understanding, as described by Skemp (1976).

An additional perspective that complements the theory of mathematization is provided by Realistic Mathematics Education (RME), developed at the Freudenthal Institute. Within RME, the process of learning mathematics is described as a guided reinvention, where students gradually move from informal strategies and intuitive reasoning to more formal mathematical expressions. This transition is strongly connected to the distinction between horizontal and vertical mathematization. In practice, learners are first encouraged to translate a realistic or context-based situation into a mathematical model (horizontal mathematization), and subsequently, to refine, restructure, and deepen their reasoning by means of formal concepts and tools (vertical mathematization) (Treffers, 1987). The RME perspective helps to explain why students often find contextualized problems more accessible: such tasks activate their intuitive resources and provide an entry point for mathematization. However, many learners encounter difficulties in advancing from this initial stage to the formal level, especially when infinite processes are involved.

Sierpinska (1990) have emphasized that certain misconceptions are not simply errors to be corrected but rather robust mental constructions that may be productive in some contexts yet become barriers in others. In the case of limits, sequences, and series, one of the most persistent obstacles is the belief that an infinite process can never be completed and therefore cannot yield a final result. This line of reasoning resonates with the ancient paradoxes of Zeno and continues to shape students' concept images of convergence (Cornu, 1991). Such obstacles reveal why learners frequently accept the idea of "approach" but hesitate to endorse the existence of an actual limit.

Recognizing the role of conceptual obstacles highlights that progress in learning infinite processes requires more than the transfer of formal definitions. It entails a reconstruction of the learner's conceptual framework. Instruction must therefore provide opportunities for students to confront the limits of their intuitive reasoning and to reflect critically on their own mental images. Studies in mathematics education have shown that when learners are explicitly guided to compare their personal concept images with formal concept definitions, deeper understanding of convergence and related ideas is more likely to emerge (Swinyard & Larsen, 2012; Tall & Vinner, 1981). From this perspective, the combined lens of mathematization and conceptual obstacles offers a comprehensive theoretical basis for analyzing students' strategies and the persistence of their misconceptions when reasoning about infinite processes.

In addition, this study draws on the theoretical framework of concept image and concept definition developed by Tall and Vinner (1981). Specifically, the concept image is defined as "the total cognitive structure that is associated with a concept, which includes all the mental pictures, properties and processes associated with the concept, and the connections between them. The concept image may change as the individual meets new stimuli and matures" (Tall & Vinner, 1981, p. 152). On the other hand, the concept definition is "a form of words used to specify that concept. This may be a personal reconstruction of the concept definition, which the individual accepts as giving meaning to the concept name" (Tall & Vinner, 1981, p. 153).

Within the domain of mathematics education, successful teaching involves enriching or even reshaping students' concept images in order to align them with the formal concept definitions. The integration of the mathematization framework with the concept image/definition distinction provides a comprehensive lens for analyzing students' strategies and misconceptions when reasoning about infinite processes.

Given these challenges, it becomes crucial to investigate not only whether students can apply procedures correctly but also how they construct meaning when faced with infinite processes. By focusing on the interplay between students' concept images and formal concept definitions, as well as their ability to move from horizontal to vertical mathematization, this study aims to shed light on the nature of their reasoning and the obstacles that arise in the transition to university mathematics. This perspective provides the foundation for the theoretical framework that guides the analysis presented in the following section.

Method

This study was conducted in the Department of Mathematics at a Greek university. The participants were 87 first-year undergraduate students who had attended the course Calculus I, which covers, among other topics, the concept of limit, the convergence of sequences and series, and infinite processes.

Data Collection

Data were collected through a semi-structured questionnaire consisting of open-ended questions designed to capture students' procedural and conceptual understanding of infinite processes. Participation in the study was voluntary, anonymous, and approved by the department, and the completion time was approximately 45 minutes.

In this paper, we focus on the following two items:

- Q1. Suppose that starting today you begin saving money in an infinite process, following this rule: Today you save 1 euro, tomorrow ½ euro, the day after tomorrow ¼ euro – in other words, each day you save half as much as the previous day. Will there ever be a day when the total amount of money you have saved is greater than or equal to 2 euros? Justify your answer.
- Q2. The sequence of functions $f_n(x) = lnx + n,\ n = 0,1,2,\ldots$ is given: In the following Cartesian

coordinate system, the graphs of $f_n(x)$, $n=0,1,2,3,4$ are represented. Consider whether there such that the point of intersection of the graph of with the x-axis is at the minimum possible distance from the origin of the axes, if it exists. Justify your answer.

The content summarized as Graphical Representations of functions fnx, n=0,1,2,3,4 is presented in Figure 1.

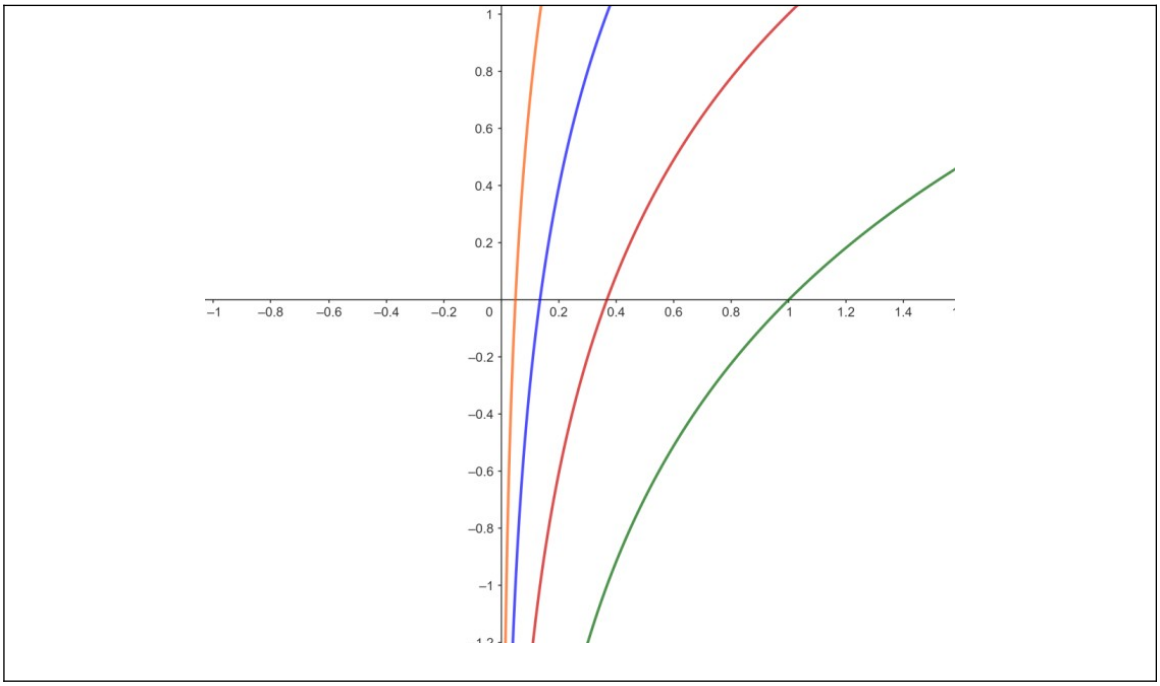


Figure 1. Graphical Representations of Functions fnx, n=0,1,2,3,4

Data Analysis

The analysis of students' responses was guided by two theoretical frameworks:

The theory of mathematization (Freudenthal, 1973; Treffers, 1987), which was used to identify evidence of horizontal and vertical mathematization in students' reasoning.

The theory of concept image and concept definition (Tall & Vinner, 1981), which was employed to explore the relationship between students' personal representations (concept images) and the formal definitions (concept definitions) they utilized.

Responses were categorized through a combination of qualitative content analysis and quantitative recording of the frequency of particular strategies and misconceptions.

Results

The students' responses to the questions were initially evaluated based on whether or not an answer was provided and, in cases where an answer was given, whether it was correct or incorrect. The results according to these

parameters of analysis are presented in the following table:

The content summarized as Student Responses to Q1 and Q2 is presented in Figure 2.

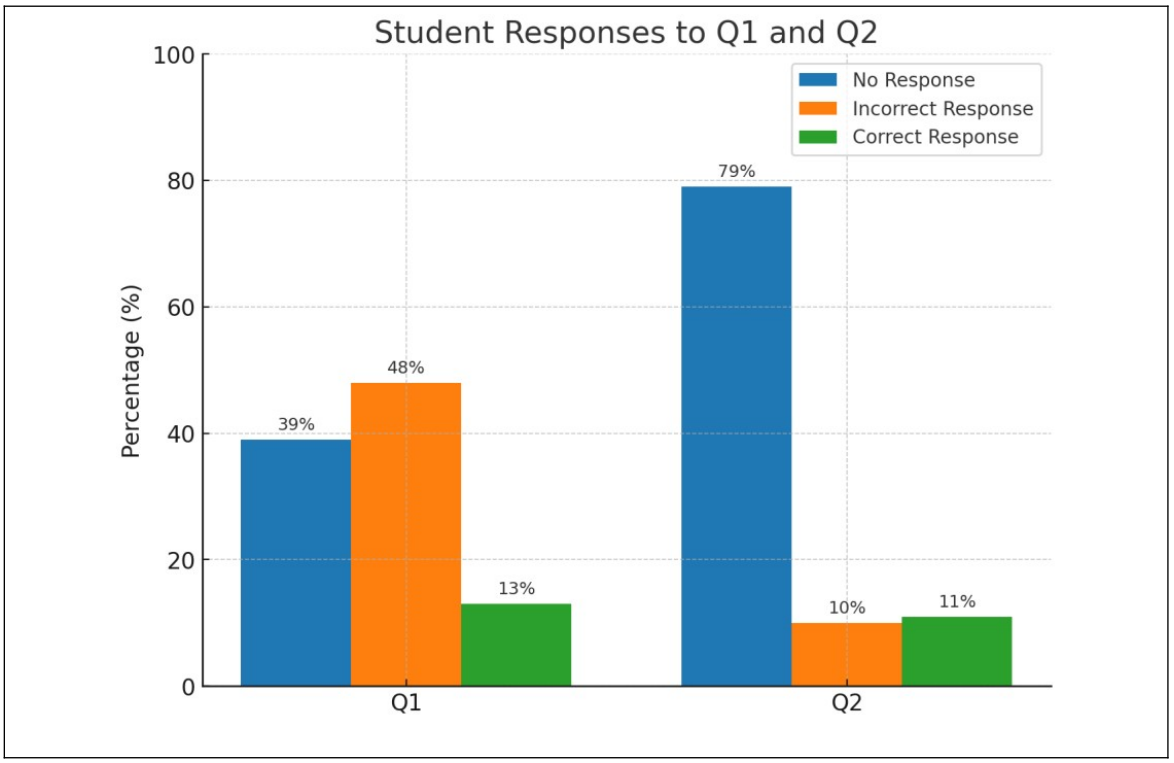


Figure 2. Student Responses to Q1 and Q2

At a second stage, the students' responses (incorrect and correct) were examined with respect to the degree of achievement of vertical and horizontal mathematization, while the students' emerging concept images were also studied.

Q1 (Saving problem):

In question Q1 the incorrect answers amounted to 48%. From this percentage only 24% managed to interpret the problem as a mathematical process (horizontal mathematization). From those who achieved this goal 29% were able to answer the problem correctly (vertical mathematization) using the sum of a geometric series. Finally, it was found that almost universally the students' concept images are characterized by misconceptions, such as the belief that the process always remains incomplete and therefore no final result can be produced, something that shows limited understanding of the notion of convergence.

The content summarized as Q1 (Saving Problems) is presented in Figure 3. In question Q2 the percentage of students who did not give any answer is particularly high (79%). From the remaining 21% the percentage of students who can reduce the problem to a mathematical model reaches 35% and from them 41% manage to complete with an overall mathematically correct answer. Regarding the prevailing concept image of the students who answer this question we can claim that Convergence as attainment of a minimum distance. More specifically, from those who manage to find that the intersection point of the graph of the function has abscissa by looking at

the sequence of graphs they think that there must exist n, so that the minimum distance is attainable.

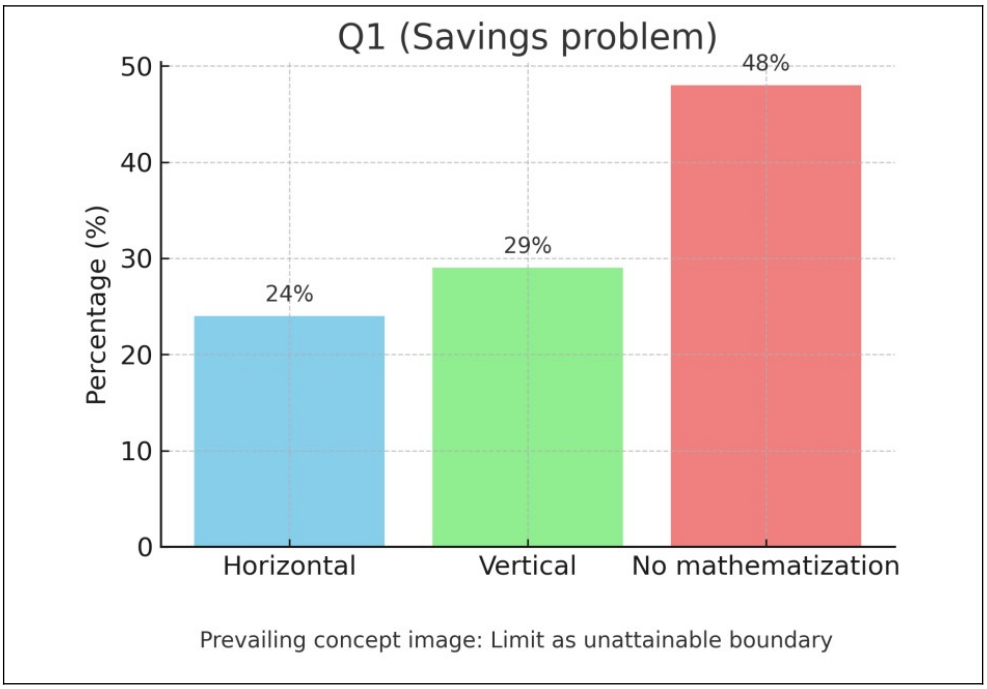


Figure 3. Q1 (Saving Problems)

The content summarized as Q3 (Sequence of functions problem) is presented in Figure 4.

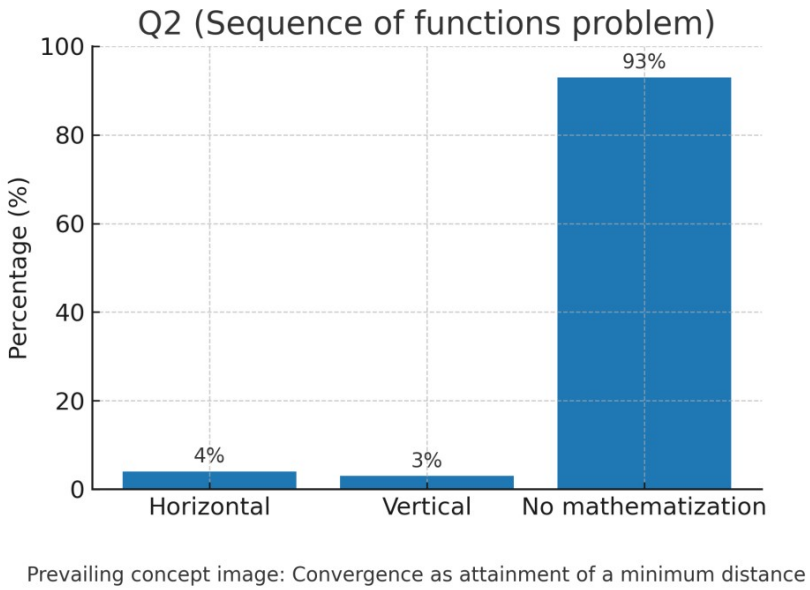


Figure 4. Q3 (Sequence of Functions Problem)

Overall, the findings indicate that students are more likely to engage in horizontal mathematization when tasks are embedded in familiar real-world contexts, but they struggle to advance to vertical mathematization when formal mathematical structures are required. Moreover, the recurring misconceptions observed in both items point to persistent gaps between students' concept images and the formal definitions of limit and convergence.

Discussion

According to the analysis of the results, we observe the difficulties that students face during the transition from secondary to university mathematics. In the Greek high school curriculum, the ε–δ definition of limit is not part of the prescribed content, and therefore students encounter infinite processes mainly through intuitive approaches, often relying on graphical representations of functions. By contrast, emphasis is placed on computational techniques, which leads students to develop stronger procedural skills. It would be expected that, during their studies in a mathematics department, their level of conceptual understanding of infinite processes would evolve and that they would become more capable of engaging with formal definitions. However, the findings indicate that students' concept images remain disconnected from the formal concept definitions, resulting in persistent misconceptions.

These results also highlight the tension between procedural fluency and conceptual understanding. While students may successfully perform routine calculations, their reasoning often lacks the underlying structure required to justify results in a rigorous way. This imbalance suggests that the instructional emphasis on techniques in secondary education has long-lasting effects, shaping the ways in which students approach new mathematical objects at university. Addressing this gap requires explicit pedagogical strategies that not only expose learners to formal definitions but also guide them in reconciling their intuitive approaches with the abstract framework of analysis.

In Q1 (Savings problem), the rate of horizontal mathematization confirms that tasks situated in familiar, real-world contexts are more accessible to students. This finding is consistent with research highlighting the potential of such tasks to bridge intuitive reasoning with formal mathematics (Jupri & Drijvers, 2016). Nevertheless, only a minority of students advanced to vertical mathematization by applying the formula for the sum of an infinite geometric series. The prevailing misconception that the process "never reaches" the target value is in line with earlier studies showing that students maintain a dynamic but incomplete view of convergence (Tall & Vinner, 1981).

In Q2 (Sequence of functions problem), the very high percentage of non-responses (79%) demonstrates that many students were unable to engage in any form of mathematization. Among the small proportion who attempted the task, misconceptions such as interpreting convergence as the attainment of a minimum distance were common. These findings are consistent with recent studies documenting students' persistent difficulties with functional convergence and their tendency to rely on intuitive graphical representations while neglecting the underlying formal mathematical reasoning (Di Martino et al., 2023; Geisler et al., 2023).

In addition to these observations, it is important to recognize that students' responses reveal not only gaps in procedural and conceptual understanding but also broader epistemological tensions. Many participants appeared to rely on an operational view of convergence, focusing on repeated computation or graphical inspection, while neglecting the structural perspective required to justify results formally. This imbalance between operational and

structural reasoning has been documented in earlier studies (Sfard, 1991) and remains a significant challenge in the teaching of analysis.

Another dimension worth highlighting concerns the affective factors that accompany students' reasoning. The high percentage of non-responses in the second task may not only reflect cognitive difficulty but also feelings of insecurity and lack of confidence when dealing with unfamiliar mathematical objects. Research at the transition to university has shown that such affective dimensions—particularly anxiety and low self-efficacy—can exacerbate students' difficulties and even contribute to dropout (Geisler et al., 2023). Therefore, addressing students' misconceptions cannot be separated from creating supportive learning environments that encourage risk-taking and value partial attempts as steps toward deeper understanding.

Furthermore, the findings suggest that the way mathematical concepts are introduced in earlier education plays a decisive role in shaping students' later difficulties. The absence of the formal definition of the limit in secondary school appears to leave learners with an incomplete conceptual image, one that resists integration with the more formal approaches encountered at university. This points to the need for a more coherent curricular progression, where intuitive representations are deliberately connected to their formal counterparts instead of being left to coexist in isolation.

Finally, the study highlights the potential benefits of integrating technological tools and dynamic visualizations into instruction. Interactive graphing software, for instance, could allow students to explore how sequences of functions behave across different domains, making the abstract notion of convergence more tangible. When combined with carefully designed tasks that encourage reflection, such tools may help bridge the persistent gap between students' intuitive images and formal mathematical definitions.

Conclusion

This study explored how first-year university students deal with infinite processes in two different tasks. The results suggest that students are more comfortable with problems in a real-world setting but face major obstacles when they need to rely only on formal structures.

For teaching, this means that more attention should be given to connecting intuitive ideas with formal reasoning. Classroom activities could be designed to help students move gradually from informal approaches to the use of precise mathematical tools. Future work could test such approaches and study how students' understanding of limits and convergence develops during their studies.

This study has examined the ways in which first-year university students approach infinite processes, focusing on their ability to mathematize tasks involving series and sequences of functions. The findings suggest that while students are often able to engage with real-world contexts and initiate horizontal mathematization, they encounter significant barriers when asked to progress to vertical mathematization. This difficulty is not only a matter of technical skill but also reflects the persistence of concept images that remain misaligned with formal definitions.

The results carry important implications for instruction. If learners' initial understanding is largely shaped by intuitive reasoning and procedural habits from secondary education, then university teaching must explicitly create opportunities to reconstruct these mental models. Rather than treating misconceptions as simple mistakes, they can be recognized as conceptual obstacles that require careful intervention. Tasks that invite students to articulate their reasoning, compare different solution paths, and critically reflect on the adequacy of their arguments may support this reconstruction process.

From a curricular perspective, greater continuity between secondary and tertiary education appears essential. The absence of the formal ε–δ definition of limit in high school leaves many students ill-prepared to engage with rigorous definitions at university. Collaborative curriculum planning across educational levels could help to smooth this transition, ensuring that students encounter the central ideas of convergence gradually and in a way that connects intuitive approaches to formal reasoning.

Pedagogically, strategies inspired by Realistic Mathematics Education (RME) may be particularly effective. By embedding infinite processes in realistic or meaningful contexts, instructors can make horizontal mathematization more accessible, while subsequent guided reflection can help students progress toward vertical mathematization. In parallel, the integration of visual tools—such as dynamic graphing environments—can foster deeper engagement with the notion of convergence, allowing students to move beyond static images and explore the dynamic nature of infinite processes.

Finally, the study points to several directions for future research. Longitudinal investigations could trace how students' concept images of limits and convergence evolve during their undergraduate years, while intervention studies could test specific teaching designs aimed at bridging the gap between informal reasoning and formal mathematical structures. Comparative studies across different curricula and cultural contexts may also shed light on how systemic factors shape students' difficulties with infinite processes. Taken together, the findings reaffirm that understanding infinite processes is not simply a matter of mastering symbolic techniques, but rather of engaging students in the complex transition from intuition to formalization. By addressing conceptual obstacles, fostering mathematization in both its horizontal and vertical forms, and ensuring smoother curricular continuity, mathematics educators can better support learners in developing robust and coherent understandings of fundamental concepts in analysis.

Statements and Declarations

Acknowledgments and Notes: Not applicable.

AI Use: Not applicable.

Supplementary Materials: Not applicable.

Author Contributions: Not applicable.

Funding: Not applicable.

Data Availability: Not applicable.

Ethics Approval: Not applicable.

Informed Consent: Not applicable.

Conflicts of Interest: Not applicable.

References

Cornu, B. (1991). *Limits.* In D. Tall (Ed.), Advanced mathematical thinking (pp. 153–166). Dordrecht: Kluwer Academic Publishers.

Di Martino, P., Gregorio, F., & Iannone, P. (2023). Transition from school into university mathematics: Experiences across educational contexts. *Educational Studies in Mathematics*, *113*(1), 1–5.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Reidel Publishing Company.

Geisler, S., Rolka, K., & Rach, S. (2023). Development of affect at the transition to university mathematics and its relation to dropout—Identifying related learning situations and deriving possible support measures. *Educational Studies in Mathematics*, *113*(1), 35–56.

Jupri, A., & Drijvers, P. H. M. (2016). Student Difficulties in Mathematizing Word Problems in Algebra. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(9), 2481–2502. https://doi.org/10.12973/eurasia.2016.1299a

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1–36. https://doi.org/10.1007/BF00302715

Sierpinska, A. (1990). Some remarks on understanding in mathematics. For the Learning of Mathematics, 10(3), 24–36, 41.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, *77*, 20–26.

Swinyard, C. A., & Larsen, S. (2012). Coming to understand the formal definition of limit. *Journal for Research in Mathematics Education*, *43*(4), 465–493.

Tall, D., & Vinner, S. (1981). Concept image and concept definition mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169. https://doi.org/10.1007/BF00305619

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction—the Wiskobas project*. Dordrecht: Reidel.

Williams, S. R. (1991). Models of limit held by college calculus students. *Journal for Research in Mathematics Education*, *22*(3), 219–236.

BOOKS

SonSES
Studies on Social and Education Sciences

SonEST
Studies on Education, Science, and Technology

JOURNALS

IJEMST
International Journal of Education in Mathematics, Science and Technology

IJRES
International Journal of Research in Education and Science

IJonSES
International Journal on Social and Education Sciences

IJTE
International Journal of Technology in Education

IJTES
International Journal of Technology in Education and Science

IJonSE
International Journal on Studies in Education

IJSES
International Journal of Studies in Education and Science

IJonEST
International Journal on Engineering, Science and Technology

IJofES
International Journal of Education Science

IJofTES
International journal of Technology in Education Science



www.istes.org/event/icemst