THE INTERNATIONAL REVIEW
OF RESEARCH IN OPEN AND
DISTRIBUTED LEARNING

Vol 27
No 1

February 2026

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Editorial – Volume 27, Issue 1

Constance Blomgren
Associate Editor, Athabasca University

Welcome to 2026 and to another year of research publications in the world of open and distributed learning. The IRRODL editors continue to receive many submissions, and we are currently developing a special 2026 issue on educational change driven by GenAI and AIED, to be released in the coming months. For this first issue of the new year, IRRODL offers nine research articles, a book review, a policy note, and three literature reviews.

Research Articles

Evaluating the Effectiveness of Online, In-Person, and Hybrid Learning: A Case Study of Engineering Disciplines at a Chinese Technical University offers research findings based on modes of delivery. The research of **Guo, Lagutkina, and Mamedova** indicates a statistically significant improvement in technical learning with a blended approach.

Chen, Chen, Zou, Xie, and **Wang** authored *Exploring Cognitive Presence in Online Collaborative Knowledge-Building: Structural, Temporal, and Social Perspectives*. This study examines three complementary analytic approaches: epistemic network analysis, sequential pattern mining, and social network analysis. Through the analysis of idea connection, temporal analysis, and group interaction patterns, useful insights were gained for designing and guiding collaborative and online knowledge building.

Innovating Interprofessional Continuing Professional Development: Applying the Community of Inquiry Framework to Digital Learning Platforms examines interprofessional continuing professional development for those working in healthcare. **Manganello and Aleo's** analysis of CoI and digital interprofessional education studies revealed four distinct adaptation patterns: communication tool convergence, evolving divergence in collaborative technologies, implementation gaps in reflective tools, and healthcare-specific adaptations. The authors use the findings to theorize CoI application in the design of online healthcare professional development.

Chang and Windeatt looked to *Usability Testing for an Open Educational Resource to Teach Language and Culture*. This study extends the knowledge of creating, testing, and developing open educational resources for migrants and refugees learning English as an additional language.

Tobondo examines *The Role of Open and Distance Education in Reducing the Educational Gap in Indonesia*, with an emphasis on differences between urban and rural areas. The study's findings apply to the country itself and contribute to a broader understanding of the educational significance of the urban-rural divide.

Analyzing Middle School Students' Distance Education Experiences in COVID-19 via Sentiment Analysis and Topic Modeling contributes to scholarship regarding emergency remote education for children and youth. **Bahçekapılı, Kandemir, and Kablan** applied artificial intelligence text analysis to reveal that students valued the flexibility of remote learning but faced significant physical, technological and pedagogical challenges.

Liu, Jiang, Xiong, and Zhao conducted research in China to examine *How Task and Individual Characteristics Affect Students' Cognitive Load: The Moderating Role of AI-Generated Content*. Applying a structural model, the authors studied task characteristics, cognitive load, and individual characteristics of undergraduate participants encountering AI-generated content. The results reveal both positive and negative correlations, including differences found between the two genders studied.

Multimodal Engagement and Sentiment Analytics in Health Science Education: A Learning Analytics Framework Integrating AI and Pedagogical Theory contributes to the growing body of knowledge on integrating AI into online learning. **Fang, Mu, Xing, Chen, and Wong** studied two image-based data sets and textual data to produce a multimodal learning analytics framework that infers learner emotions and engagement.

In this Turkish study, **Ayar** used a mixed-methods approach in *Microphones on Unmute: Perceived Online English-Speaking Anxiety of Non-Native EFL Educators*. The findings showed a lack of perceived competence, troubles with online technologies, and learners' English proficiency contributing to the speaking anxiety when teaching online for these language educators.

This issue includes **Belawati's** book review of the *Handbook of Open Universities Around the World*, edited by Mishra and Panda. A timely print book with over 100 contributions by scholars and practitioners from open universities worldwide, the reviewer recommends diving into the chapters and furthering one's understanding of higher education openness, past and present. With its theoretical initial chapter alongside entries with empirical robustness and comparative themes, the volume ends with chapters examining educational transformations wrought by digital and artificial intelligence.

The technical note, *Regulation of Distance Learning Courses in Brazilian Higher Education*, offers insights for practitioners and those involved in policy development. **Reis, Monteiro, Bertoia, Klement, Prado, Dahmer, Rodrigues, and Passone** examine educational quality, accessibility, and institutional accountability in Brazil as part of recent regulatory change concerning distance education and higher education.

Three literature reviews contribute to understanding digital literacy and collaborative teaching, MOOCS for health education, and the influence of GenAI upon learning achievement. **Wider, Saad, Mahmood, Ishak, Aziz, Wu, Wu, and Tanucan** executed a systematic review exploring *Digital Literacy in Enhancing Collaborative Teaching*. Findings indicate that digital literacy is part of pedagogical capacity that best aligns with institutional culture, curriculum design, and ongoing professional learning. **Arslan,**

Ata, and Kucuk used a systematic review to examine *MOOCs Reshaping Undergraduate Health Education*. Given MOOCs' potential to support health education, the comprehensive literature review identified the need for innovative, practical implementation strategies alongside further research. **Doo and Park** provide *A Meta-Analysis of ChatGPT's Influence on Learning Achievement*. Examining studies between 2022 and 2024, the PRISMA systematic review offers implications and directions for future research regarding the educational purposes of GenAI, including ChatGPT.

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Evaluating the Effectiveness of Online, In-Person, and Hybrid Learning: A Case Study of Engineering Disciplines at a Chinese Technical University

Lingyan Guo[1]*, Margarita Lagutkina[2], and Larisa Mamedova[3]
*[1]Jiangxi Science and Technology Normal University, Nanchang, China; [2] Peoples Friendship University of Russia, Russia; [3]Technical State Autonomous Educational Institute of Higher Professional Education, Neryungri, Russia; *Corresponding Author*

Abstract

The effectiveness of technical education may vary depending on the delivery method. This study compared the effects of online, face-to-face (F2F), and hybrid learning on engineering students' academic performance. The study involved 450 second-year students pursuing an engineering degree at a technical university in China. The pre-test and post-test scores for the five core academic subjects (i.e., computer programming, further mathematics, physics, electrical engineering, and analytical mechanics) revealed a statistically significant improvement in academic performance across all subjects after use of hybrid learning ($p < 0.000$). The average gains were 3.46 points in computer programming, 4.07 points in further mathematics, 3.24 points in physics, 2.5 points in electrical engineering, and 3.06 points in analytical mechanics. The online and F2F delivery groups exhibited a statistically significant improvement with respect to scores for electrical engineering ($p < 0.000$) and physics ($p < 0.002$) only. The one-way ANOVA and Scheffe's test results revealed that the hybrid model had the strongest learning effects compared to online and F2F. A SWOT analysis helped to further explore students' perceptions of the three delivery formats. The present findings, which highlighted the effectiveness of hybrid learning, can be helpful in creating adaptive learning programs for engineering students.

Keywords: education, face-to-face learning, hybrid learning, online learning, students, technical specialties, university

Introduction

Advances in technology have brought new threats and opportunities to education (Lightner & Lightner-Laws, 2024), and technical education is no exception (Iatsyshyn et al., 2020; Li et al., 2024; Sayfullayeva et al., 2021). Recognizing the importance of effective teaching in the 21st century, scholars have looked for the best way to deliver education in universities (Monge et al., 2023). Currently, there is a paucity of empirical evidence on how online, face-to-face (F2F), and hybrid teaching strategies affect the academic outcomes of students in the field of engineering, whereas for other fields, the knowledge base is larger (Singh et al., 2022). The COVID-19 pandemic, which paralyzed the education system in 2020, revealed that educational institutions around the world were able to adapt to the new reality and even reap the benefits of the associated transition (Potra et al., 2021). The active incorporation of digital technologies into teaching sparked interest in different learning formats (Singh et al., 2022; Wahas & Syed, 2024). Researchers have endeavoured to understand how online, F2F, and hybrid instruction can influence academic performance, in order to integrate those that are most effective and beneficial to learners (Meshko et al., 2021; Sayfullayeva et al., 2021). Online learning, despite its advantages of flexibility and virtual presence (Shen et al., 2023), makes live interactions impossible and challenges the ability of students to motivate themselves (Shadiev et al., 2024). The F2F option is centuries old and needs to be updated, even though it has long been considered the gold standard (Valieiev et al., 2021). Hybrid models, which have gained popularity over the past few years, attempt to combine the advantages of these two approaches (Kruchkova & Grigoriev, 2022).

The current study sought to narrow the knowledge gap by comparing the effectiveness of online, F2F, and hybrid delivery within the context of engineering education. The results may be useful to educational institutions and firms in the tech industry that want to optimize learning. The study may help further research on academic performance.

Literature Review

Universities that produce engineering professionals typically offer courses in automation, electronics, computer science, engineering, and other technical fields (Theobald et al., 2020). Technical professionals are expected to possess the skills and knowledge needed to improve existing technologies and develop new ones (Khamidjanovna et al., 2022). In China, due to global and local developments in the technology sector (Wu et al., 2020; Zhang et al., 2022), the most popular technical fields have been electrical engineering, mechanical engineering, and electronics (Lin et al., 2021). To remain competitive in the international market, China has acknowledged the need to enhance the quality of technical education (Ren & Ji, 2021). It is impossible to boost production and develop advanced technologies without properly educated specialists (Kashiramka et al., 2021; Mohammad Shafi et al., 2021). Hence, in countries with high population densities, the quality of technical education has been closely linked to economic growth and quality of life (Lewis, 2020).

From a theoretical standpoint, this study was grounded in the constructivist framework, which posits that knowledge is constructed through learners' active engagement in the educational process (Efgivia et al., 2021). This theory emphasizes the importance of interaction between students and instructors, as well as the necessity of applying acquired knowledge in real-world contexts (Nurhasnah et al., 2024). The hybrid learning format, which integrates both online and face-to-face components, aligns with the core principles of constructivist theory, as it fosters students' active and conscious participation, self-regulation, and the contextual application of knowledge (Niyomves et al., 2024; Wang & Bhagat, 2025). Furthermore, the theory of blended learning—an extension of constructivism—highlights the adaptability of the educational process and the strategic use of digital technologies to enhance learning opportunities (Liu et al., 2024; Singh et al., 2021).

Academic outcomes (e.g., grades and performance on various tests, exams, thesis presentations) are qualitative and quantitative indicators of academic success (Kaya & Erdem, 2021; Miller et al., 2021). They serve as markers that help judge the quality of education, the effectiveness of the teaching strategies employed, and the extent to which learners have engaged with the course (Demir et al., 2021). Higher academic results indicate more effective teaching (Rafiola et al., 2020), and more effective teaching raises students' chances of future success in their profession (Hayat et al., 2020). From a practical standpoint, academic results can be helpful in adapting the teaching strategy to the needs of students and teachers (Reis et al., 2021).

Seminars, lectures, and practical sessions can be delivered in different ways. The online approach implies that the learning activities occur within a virtual space (Nambiar, 2020). Online platforms have allowed teachers and students to communicate with each other via the Internet, thereby enabling learning without physical presence in the classroom (Farrell & Brunton, 2020; Kadhim et al., 2023). Online education has offered access to a wide range of educational resources, which have become abundant since the onset of the COVID-19 pandemic, and enhanced flexibility (Cramarenco et al., 2023). Traditional F2F delivery requires the presence of both the teacher and student in a physical location (Stevens et al., 2021) with knowledge exchanged through live interaction, which helps to build and maintain social connections (Gherheș et al., 2021; Louis-Jean & Cenat, 2020). The hybrid format, as the name suggests, combines the best of both online and F2F instructional methods (Singh et al., 2021). Hybrid learning can be implemented in different ways. For example, laboratory and practical sessions can take place in an online environment, while discussions can be traditionally styled. In this case, the learning materials may be posted online (Haningsih & Rohmi, 2022). There is a need to identify which of these three formats is most effective in improving students' grades in engineering education (DeChenne-Peters et al., 2022; Owston et al., 2020). It is also important to understand their strengths and weaknesses and the opportunities and threats they bring.

Problem Statement

This study compared the effectiveness of different delivery formats used in university courses (i.e., face-to-face, online, hybrid) to determine which worked best with engineering students. Hypothetically, hybrid learning will be the most effective method as it combines the best of both face-to-face and online instructional methods. The insights provided here may be useful in education reform efforts and help boost the quality of technical education. The current study explored the pros and cons of each delivery format through the lens of a strengths, weaknesses, opportunities, and threats (SWOT) framework using data derived from student interviews. The objectives of the study were to (a) compare the pre-test and post-test scores for academic performance across three study groups; (b) determine whether the results of these groups were different (ANOVA analysis); (c) determine which group was significantly different compared to others (Scheffe's post-hoc analysis); and (d) identify strengths, weaknesses, opportunities, and threats associated with each delivery format under consideration (SWOT analysis).

Materials and Methods

In this study, student performance was assessed across five academic subjects (i.e., computer programming, further mathematics, physics, electrical engineering, and analytical mechanics) using final course grades ranging from 0 to 100 as measures of academic performance. The five academic disciplines in question are extremely important as a source of in-demand technical professionals. The final course grades were derived from the anonymized academic performance report provided by the university.

Participants

The study involved 450 second-year students pursuing an engineering degree at a technical university in China. The potential participants entered the degree course a year prior to the intervention; hence, they were familiar with the five academic subjects under consideration and possessed some amount of knowledge in the field. All students were randomly divided into three equal groups: online, traditional F2F, and hybrid (for more details, see Table 1).

Table 1

Data on Student Participants

Group	Number	Male	Female	Mean age	SD
Online delivery	150	109	41	19.27	0.87
F2F delivery	150	110	40	19.58	0.71
Hybrid delivery	150	114	36	19.17	0.83
Total	450	333	117	-	-
Overall mean	-	-	-	19.34	0.80

As shown in the table above, male students were overrepresented in each group, most likely because in China, males dominate in technical specialties. The students were all roughly the same age, which meant the groups were compatible with each other in terms of age. The homogeneity of each group was also tested with respect to pre-test subject scores. For this, a t-test was employed.

Study Design

The intervention was conducted from September 2022 to July 2023. During the preparatory phase, participants were selected, and informed consent was obtained. Students were randomly assigned to one of three groups (online, face-to-face, and hybrid learning). The homogeneity of the groups was verified based on initial academic performance, using final grades from the first year in the same five engineering courses. These academic results served as a reliable baseline or pre-test for the study.

During the main phase of the research, which began in late September 2022, each group received instruction in one of the three designated formats. In the online group, all lectures and practical sessions were delivered remotely via DingTalk and MOOC platforms. The F2F group attended all instructional activities in traditional classroom settings. The hybrid group participated in a combination of online and face-to-face lectures and practical classes. Additionally, both the online and hybrid groups engaged in in-person sessions for consultations, group work, and demonstration-based instruction that required physical presence. These face-to-face components were integrated to enhance the educational impact of the primary delivery format.

The total instructional load (in hours) was equal across all three groups. Final academic results (i.e., post-tests) were collected at the end of the academic year and served to assess students' learning outcomes following instruction in their respective formats. Furthermore, between June and July 2023, interviews with students were conducted to inform a SWOT analysis based on their experiences and feedback.

Online Delivery Group

Students in this group studied the learning material online using the DingTalk app and the MOOC platform. DingTalk, a video conferencing platform, has been widely used in China (DingTalk, 2023). In this study, DingTalk served as a tool to distribute course materials, submit homework, and receive feedback from teachers.

The MOOC platform has hosted various courses, including those in technical disciplines (MOOC China, 2023). It enables learners to participate in discussion forums, communicate with their classmates and teachers, complete online assignments and tests, and study independently.

F2F Delivery Group

Students in this group received F2F instructions only. Lectures, practical lessons, seminars, and discussions were traditionally styled, and no computer programs or digital platforms were used. One exception was the professional software that students were expected to cover as part of their training program.

Hybrid Delivery Group

Students in this group took both online and F2F classes. The digital resources used to support online interactions were the MOOC platform and the DingTalk app.

Data Analysis

All data regarding student performance were transferred to an SPSS v26 file for analysis. The results were presented as descriptive statistics (i.e., means, standard errors of the mean, standard deviations, skewness, and kurtosis). The means were then compared using the following statistical methods: student's t-test, one-way analysis of variance (ANOVA), and Scheffé's method. Data for the SWOT analysis were derived from student interviews.

Ethical Issues

The ethics committee of the participating university approved the study procedures. All students who participated in the project did so voluntarily. The subjects were informed about the nature of the study and gave written consent to participate. Confidentiality and anonymity were guaranteed.

Results

Table 2 presents the descriptive statistics for student performance across five different subject areas before (pre-test) and after (post-test) the intervention. Students in the hybrid delivery group improved significantly from the pre-test, with average gains of 3.46 points in computer programming, 4.07 points in further mathematics, 3.24 points in physics, 2.5 points in electrical engineering, and 3.06 points in analytical mechanics. Improvements in the online and F2F delivery groups were less significant; the average gains there ranged from 0.25 to 1.69 points.

Table 2

Descriptive Statistics of Pre-Test and Post-Test Subject Scores: Online, F2F, and Hybrid Delivery

Group	Measure	Computer programming (pre-test)	Computer programming (post-test)	Further mathematics (pre-test)	Further mathematics (post-test)	Physics (pre-test)	Physics (post-test)	Electrical engineering (pre-test)	Electrical engineering (post-test)	Analytical mechanics (pre-test)	Analytical mechanics (post-test)
Online	*M*	75.70	76.54	71.18	72.17	76.56	77.17	75.28	76.48	70.32	70.57
	SD	4.084	4.129	4.183	3.879	4.411	4.279	3.307	2.991	4.455	4.052
	SEM	.332	.336	.340	.316	.359	.348	.269	.243	.363	.330
	Skewness	.159	-.052	.142	-.093	-.186	-.101	.157	-.209	-.166	.060
	Kurtosis	-1.044	-1.260	-1.252	-1.244	-1.011	-1.383	-1.090	-1.030	-1.197	-1.219
F2F	*M*	75.83	76.68	71.88	72.17	76.19	77.88	74.95	75.97	70.12	70.70
	SD	4.341	4.061	4.049	3.823	4.549	3.980	3.475	3.299	4.047	4.130
	SEM	.356	.333	.332	.313	.373	.326	.285	.270	.332	.338
	Skewness	-.053	-.006	-.088	-.006	.025	-.182	.268	.023	-.130	-.094
	Kurtosis	-1.352	-1.241	-1.220	-1.206	-1.257	-1.099	-1.156	-1.336	-1.122	-1.279
Hybrid	*M*	76.01	79.47	70.99	75.06	76.27	79.51	74.87	77.37	70.35	73.41
	SD	4.343	2.388	3.865	2.017	4.711	3.036	3.492	2.914	4.334	2.335
	SEM	.355	.195	.316	.165	.385	.248	.285	.238	.354	.191
	Skewness	-.021	-.012	.125	-.028	.135	.013	.169	-.033	.021	-.059
	Kurtosis	-1.238	-1.355	-1.105	-1.304	-1.242	-1.298	-1.182	-1.298	-1.249	-1.244

The student's t-test showed that differences between the pre-test and post-test subject scores in the hybrid delivery group were statistically significant ($p < 0.000$). This finding pointed to the effectiveness of the hybrid instructional format. In contrast, the only statistically significant improvements observed in the online and F2F delivery groups after the intervention were found with respect to the final scores for electrical engineering (p <0.000) and physics (p <0.002). For more details, see Table 3. Even though using online and F2F teaching strategies ensured the academic growth of students in just one subject, no downward trends were detected. Therefore, both instructional approaches were equally compatible.

Table 3

Student's t-Test Results: Performance Levels of the Online, F2F, and Hybrid Delivery

Group	Subject	Paired differences					*t*	*Dfdf*	Sig. (2-tailed)
		M	*Std. DeviationSD*	*SEM*	95% CI				
					Lower	Upper			
Online	Computer programming	-.813	5.934	.485	-1.771	.144	-1.679	149	.095
	Further mathematics	-.953	5.632	.460	-1.862	-.045	-2.073	149	.040
	Physics	-.680	5.704	.466	-1.600	.240	-1.460	149	.146
	Electrical engineering	-1.227	4.065	.332	-1.883	-.571	-3.696	149	.000
	Analytical mechanics	-.220	5.546	.453	-1.115	.675	-.486	149	.628
F2F	Computer programming	-.873	5.701	.465	-1.793	.046	-1.876	149	.063
	Further mathematics	-.340	5.623	.459	-1.247	.567	-.741	149	.460
	Physics	-1.607	6.132	.501	-2.596	-.617	-3.209	149	.002
	Electrical engineering	-1.000	5.194	.424	-1.838	-.162	-2.358	149	.020
	Analytical mechanics	-.607	5.956	.486	-1.568	.354	-1.247	149	.214
Hybrid	Computer programming	-3.460	4.832	.395	-4.240	-2.680	-8.769	149	.000
	Further mathematics	-4.073	4.448	.363	-4.791	-3.356	-11.215	149	.000
	Physics	-3.247	5.703	.466	-4.167	-2.327	-6.973	149	.000
	Electrical engineering	-2.507	4.517	.369	-3.235	-1.778	-6.797	149	.000
	Analytical mechanics	-3.060	4.912	.401	-3.852	-2.268	-7.630	149	.000

There were no statistically significant differences in pre-test subject scores between groups as determined by one-way ANOVA (Table 4). At the same time, differences between the post-test scores were statistically significant across all subjects (p < 0.000), indicating that not all delivery formats are highly effective delivery formats.

Table 4

Performance Variance in the Three Comparison Groups (One-Way ANOVA Findings)

Subjects	Comparison Groups	*Sum of SquaresSS*	*df*	*Mean SquareMS*	*F*	Sig.
Computer programming (pre-test)	Between groups	7.032	2	3.516	.194	.824
	Within groups	8101.388	447	18.124		
	Total	8108.420	449			
Computer programming (post-test)	Between groups	817.417	2	408.709	31.244	.000
	Within groups	5847.340	447	13.081		
	Total	6664.758	449			
Further mathematics (pre-test)	Between groups	65.949	2	32.974	2.026	.133
	Within groups	7275.971	447	16.277		
	Total	7341.920	449			
Further mathematics (post-test)	Between groups	833.285	2	416.642	37.044	.000
	Within groups	5027.446	447	11.247		
	Total	5860.731	449			
Physics (pre-test)	Between groups	11.756	2	5.878	.283	.754
	Within groups	9289.224	447	20.781		
	Total	9300.980	449			
Physics (post-test)	Between groups	435.531	2	217.766	15.059	.000
	Within groups	6464.160	447	14.461		
	Total	6899.691	449			
Electrical engineering (pre-test)	Between groups	14.343	2	7.171	.611	.543
	Within groups	5245.222	447	11.734		
	Total	5259.564	449			
Electrical engineering (post-test)	Between groups	151.530	2	75.765	8.030	.000
	Within groups	4217.634	447	9.435		
	Total	4369.164	449			
Analytical mechanics (pre-test)	Between groups	4.802	2	2.401	.131	.877
	Within groups	8199.198	447	18.343		
	Total	8204.000	449			
Analytical mechanics (post-test)	Between groups	772.358	2	386.179	29.760	.000
	Within groups	5800.400	447	12.976		
	Total	6572.758	449			

The results of Scheffe's post-hoc test analysis revealed significant differences between hybrid delivery and the other two formats across all subjects ($p < 0.001$). This finding suggested that the hybrid instructional format was the most effective option among the three formats. For more details, see Table 5. The significant differences between online and F2F formats were not present for every subject, indicating that these two delivery strategies had similar levels of impact on learning.

Table 5

Scheffe's Post-Hoc Test Findings

Subject	(I)	(J)	Mean difference (I - J)	*Std. ErrorSE*	Sig.	99.9% CI Lower	99.9% CI Upper
Computer programming	Online	F2F	-.135	.418	.949	-1.70	1.43
		Hybrid	-2.924*	.417	.000	-4.49	-1.36
	F2F	Online	.135	.418	.949	-1.43	1.70
		Hybrid	-2.789*	.418	.000	-4.36	-1.22
	Hybrid	Online	2.924*	.417	.000	1.36	4.49
		F2F	2.789*	.418	.000	1.22	4.36
Further mathematics	Online	F2F	-.002	.387	1.000	-1.45	1.45
		Hybrid	-2.888*	.387	.000	-4.34	-1.44
	F2F	Online	.002	.387	1.000	-1.45	1.45
		Hybrid	-2.886*	.388	.000	-4.34	-1.43
	Hybrid	Online	2.888*	.387	.000	1.44	4.34
		F2F	2.886*	.388	.000	1.43	4.34
Physics	Online	F2F	-.714	.439	.268	-2.36	.93
		Hybrid	-2.348*	.438	.000	-3.99	-.71
	F2F	Online	.714	.439	.268	-.93	2.36
		Hybrid	-1.634	.440	.001	-3.28	.01
	Hybrid	Online	2.348*	.438	.000	.71	3.99
		F2F	1.634	.440	.001	-.01	3.28
Electrical engineering	Online	F2F	.517	.355	.347	-.81	1.85
		Hybrid	-.890	.354	.043	-2.22	.44
	F2F	Online	-.517	.355	.347	-1.85	.81
		Hybrid	-1.407*	.355	.000	-2.74	-.08
	Hybrid	Online	.890	.354	.043	-.44	2.22
		F2F	1.407*	.355	.000	.08	2.74
Analytical mechanics	Online	F2F	-.135	.416	.949	-1.69	1.42
		Hybrid	-2.844*	.415	.000	-4.40	-1.29
	F2F	Online	.135	.416	.949	-1.42	1.69
		Hybrid	-2.709*	.417	.000	-4.27	-1.15
	Hybrid	Online	2.844*	.415	.000	1.29	4.40
		F2F	2.709*	.417	.000	1.15	4.27

Note. *Correlation is significant at the 0.001 level.

In terms of advantages (Table 6), delivering courses online means that (a) more materials can be made available to students, (b) students get more autonomy in deciding their own schedule, and (c) these courses can be completed anywhere. At the same time, learning online leads to a lack of interpersonal communication and creates a need for self-regulation. The F2F delivery format fosters direct interaction and allows teachers to structure classes in a way that makes it easy to follow the curriculum. The downsides of this strategy are that you can only study in a physical classroom and that learning feels monotonous and outdated. The hybrid format combines the advantages of both methods (i.e., interactivity and hands-on experience), but technological compatibility may be a challenge.

Table 6

SWOT Analysis of Online, F2F, and Hybrid Modalities

Category	Online	F2F	Hybrid
Strength	Wide access to educational resources, no geographical constrains, customized learning schedule	Clarity, structured learning and management, in-person contact with teacher and peers	Combines the advantages of online and F2F methods (i.e., interactivity + practical and live experience)
Weakness	Lack of in-person interaction, difficulty focusing without teacher's supervision, poor learning effectiveness of some digital platforms	Inability to learn outside of campus, monotony, outdated training methods	Difficulty adjusting, incompatibility of digital technologies, forced alternation between online and F2F contexts
Opportunity	Adaptation of existing courses, freedom of distribution, technological progress	Deep student-teacher interaction, socialization, living the life of a university student	Higher academic performance, geographical flexibility, access to learning in time of crisis (e.g., during the COVID-19 pandemic)
Threat	Isolation, slow Internet, technical problems with audio and video materials, technological incompetence	Less time saved, dependence on physical presence, no access to learning in time of crisis (e.g., during the COVID-19 pandemic)	High optimization effort required, technological incompetence

Discussion

Integrating online learning models with hands-on activities was reported to be beneficial to STEM education (Shen et al., 2023), which resonates with the current study. Yet, this process has presented challenges for teachers and students unfamiliar with the new technology (Wahas & Syed, 2024). The present study revealed similar obstacles. Our results highlighted the need to create effective teaching strategies and prepare educators to implement new technology. This study was not the first attempt at using distance technologies to teach engineering students. Previous research supported the present findings regarding the importance of innovative teaching methods (Meshko et al., 2021; Sayfullayeva et al., 2021). Blended learning has been associated with higher levels of student satisfaction when compared to F2F and distance modes (Valieiev et al., 2021). In this study, hybrid learning had a superior boosting effect on academic performance. Similarly, other researchers have argued that a hybrid delivery format can significantly improve the quality of education (Kruchkova & Grigoriev, 2022).

Some scholars have pointed out the growing need for skilled technical professionals and the current education system's shortcomings in producing such specialists through traditional routes (Lewis, 2020). Furthermore, other researchers have acknowledged the need to adapt curricula to the current demands of the labor market and facilitate collaboration between educational institutions and employers (Mohammad Shafi et al., 2021). Some researchers offered an automated assessment system as a possible solution (Kadhim et al., 2023). Their system was designed using Internet of Things technology, which supported the idea that innovation makes learning more effective. Previous research has shown that online training can be as effective as F2F (Stevens et al., 2021), which aligned with the results of the current study. Developing education programs that combine the best aspects of traditional and online learning can offer effective solutions to challenges associated with online transition (Singh et al., 2021).

Online learning has proven effective during the COVID-19 pandemic, highlighting the need to develop hybrid educational models (Haningsih & Rohmi, 2022). This finding, consistent with the current study, showed that hybrid models not only enhanced learning but also ensured greater access to learning while meeting the diverse needs of students. Some researchers provided an interesting perspective on STEM students' preferences and perceptions towards blended learning (Owston et al., 2020). Significant academic gains were reported, despite a relatively low level of satisfaction with blended courses (Owston et al., 2020). This finding pointed out the need to further adapt and individualize learning approaches to satisfy students' needs. There is evidence that course-based research experiences can be adapted to online or hybrid modality with minimal impact on student performance (DeChenne-Peters et al., 2022), which is especially important in light of online transition. Hybrid learning can be more effective than online and F2F formats as demonstrated by both previous research (Felder, 2021) and our current study. This is especially true in the field of engineering. Thanks to a combination of interactive and flexible learning experiences, hybrid courses can meet student needs better than can other modalities.

Conclusions

Our results showed a statistically significant effect of hybrid learning on academic performance across all five subjects ($p < 0.000$). The average gains were 3.46 points in computer programming, 4.07 points in further mathematics, 3.24 points in physics, 2.5 points in electrical engineering, and 3.06 points in analytical mechanics. Online and F2F learning resulted in a statistically significant improvement of final grades in electrical engineering (p <0.000) and physics (p <0.002). The one-way ANOVA and Scheffe's test findings revealed that hybrid format had the strongest learning effects compared to online and F2F. Based on the results of the SWOT analysis, the strengths of the online format were accessibility and flexibility, while the strengths of the F2F format were live interaction between teacher and learner and structured learning. The hybrid format combines these advantages.

The present study contributes to the growing body of empirical research on the effectiveness of hybrid learning, particularly within the context of technical education. However, its findings may also be relevant to other academic disciplines where the integration of theoretical knowledge and practical skills is essential. Further exploration of the adaptability of hybrid learning formats across diverse educational contexts may support the enhancement of university curricula and vocational training systems. The present findings can be used by other researchers in the field and by educational institutions that seek to integrate hybrid courses into their programs. Future research may also examine the role of student motivation and preferred learning styles within the context of online, in-person, and hybrid learning formats. Such investigations would offer deeper insights into the underlying mechanisms that influence the effectiveness of each instructional approach.

Limitations

This study focused on just five technical disciplines, which may have limited the generalizability of the findings. There were just two learning platforms used throughout the courses, so more studies may be needed to explore the effects of other commonly as well as less frequently used digital tools. Despite the large sample size, all participants were recruited from the same university, making extrapolation of the data difficult. In addition, the study did not examine the longitudinal effects, which may be an interesting topic for future research.

Funding

Department of Education of Jiangxi Province. No.:23YB150

References

Cramarenco, R. E., Burcă-Voicu, M. I., & Dabija, D. C. (2023). Student perceptions of online education and digital technologies during the COVID-19 pandemic: A systematic review. *Electronics, 12*(2), 319. https://doi.org/10.3390/electronics12020319

DeChenne-Peters, S. E., Sargent, E., Mateer, S. C., Machingura, M., Zettler, J., Ness, T., DeMars, G., Cannon, S., & BrofttBailey, J. (2022). Comparison of student outcomes in a course-based undergraduate research experience: Face-to-face, hybrid, and online delivery of a biology laboratory. *International Journal for the Scholarship of Teaching and Learning, 16*(1), 5. https://doi.org/10.20429/ijsotl.2022.160105

Demir, A., Maroof, L., Sabbah Khan, N. U., & Ali, B. J. (2021). The role of e-service quality in shaping online meeting platforms: A case study from higher education sector. *Journal of Applied Research in Higher Education, 13*(5), 1436–1463. https://doi.org/10.1108/JARHE-08-2020-0253

DingTalk. (2023). *DingTalk: Smart Work Platform*. https://www.dingtalk.com/

Efgivia, M. G., Rinanda, R. A., Hidayat, A., Maulana, I., & Budiarjo, A. (2021). Analysis of constructivism learning theory. In *1st UMGESHIC international seminar on health, social science and humanities* (pp. 208–212). Atlantis Press. https://doi.org/10.2991/assehr.k.211020.032

Farrell, O., & Brunton, J. (2020). A balancing act: A window into online student engagement experiences. *International Journal of Educational Technology in Higher Education, 17*(1), 25. https://doi.org/10.1186/s41239-020-00199-x

Felder, R. M. (2021). STEM education: A tale of two paradigms. *Journal of Food Science Education, 20*(1), 8–15. https://doi.org/10.1111/1541-4329.12219

Gherheș, V., Stoian, C. E., Fărcașiu, M. A., & Stanici, M. (2021). E-learning vs. face-to-face learning: Analyzing students' preferences and behaviors. *Sustainability, 13*(8), 4381. https://doi.org/10.3390/su13084381

Haningsih, S., & Rohmi, P. (2022). The pattern of hybrid learning to maintain learning effectiveness at the higher education level post-COVID-19 pandemic. *European Journal of Educational Research, 11*(1), 243–257. https://doi.org/10.12973/eu-jer.11.1.243

Hayat, A. A., Shateri, K., Amini, M., & Shokrpour, N. (2020). Relationships between academic self-efficacy, learning-related emotions, and metacognitive learning strategies with academic performance in medical students: A structural equation model. *BMC Medical Education, 20*(1), 76. https://doi.org/10.1186/s12909-020-01995-9

Iatsyshyn, A. V., Kovach, V. O., Romanenko, Y. O., Deinega, I. I., Iatsyshyn, A. V., Popov, O. O., Kutsan, Y. G., Artemchuk, V. O., Burov, O. Y., & Lytvynova, S. H. (2020). Application of augmented reality technologies for preparation of specialists of new technological era. In A. E. Kiv & M. P. Shyshkina (Eds.), *Augmented reality in education: Proceedings of the 2nd international workshop* (pp. 181–200), Kryvyi Rih, Ukraine. http://ceur-ws.org/Vol-2547/paper14.pdf

Kadhim, J. Q., Aljazaery, I. A., & ALRikabi, H. T. S. (2023). Enhancement of online education in engineering college based on mobile wireless communication networks and IoT. *International Journal of Emerging Technologies in Learning, 18*(1), 176–200. https://doi.org/10.3991/ijet.v18i01.35987

Kashiramka, S., Sagar, M., Dubey, A. K., & Mehndiratta, A. (2021). Critical success factors for next generation technical education institutions. *Benchmarking: An International Journal, 28*(5), 1605–1621. https://doi.org/10.1108/BIJ-06-2018-0176

Kaya, M., & Erdem, C. (2021). Students' well-being and academic achievement: A meta-analysis study. *Child Indicators Research, 14*(5), 1743–1767. https://doi.org/10.1007/s12187-021-09821-4

Khamidjanovna, K. V., Nazarovna, S. S., Sattikhanovna, T. D., & Sabirzhanovna, A. G. (2022). Formation and improvement of word usage skills in the technical field. *Journal of Positive School Psychology, 6*(10), 4318–4322. https://mail.journalppw.com/index.php/jpsp/article/view/14495

Kruchkova, O., & Grigoriev, M. (2022). The introduction of innovative and active-based teaching methods into the complex process of acquiring professional competencies by students of technical specialties. In L. G. Chova, A. L. Martínez, & I. C. Torres (Eds.), *Proceedings of 16th international technology, education and development conference* (pp. 7958–7965). https://doi.org/10.21125/inted.2022.2005

Lewis, P. (2020). Developing technician skills for innovative industries: Theory, evidence from the UK life sciences industry, and policy implications. *British Journal of Industrial Relations, 58*(3), 617–643. https://doi.org/10.1111/bjir.12532

Li, Y.-D., Ding, G.-H., & Zhang, C.-Y. (2024). Effects of learner-centred education on academic achievement: A meta-analysis. *Educational Studies, 50*(3), 285–298. https://doi.org/10.1080/03055698.2021.1940874

Lightner, C. A., & Lightner-Laws, C. A. (2024). A new day in higher ed: HyFlex universities. *Interactive Learning Environments 32*(10), 7364–7381. https://doi.org/10.1080/10494820.2024.2312932

Lin, S., Lin, R., Sun, J., Wang, F., & Wu, W. (2021). Dynamically evaluating technological innovation efficiency of high-tech industry in China: Provincial, regional and industrial perspective. *Socio-Economic Planning Sciences, 74,* 100939. https://doi.org/10.1016/j.seps.2020.100939

Liu, M., Zhao, G., Zhong, Z., Ma, J., & Wang, W. (2024). Theoretical foundations for blended learning. In M. Li, X. Han, & J. Cheng (Eds.), *Handbook of educational reform through blended learning* (pp. 1–44). Springer. https://doi.org/10.1007/978-981-99-6269-3_1

Louis-Jean, J., & Cenat, K. (2020). Beyond the face-to-face learning: A contextual analysis. *Pedagogical Research, 5*(4), em0077. https://doi.org/10.29333/pr/8466

Meshko, H. M., Habrusieva, N. V., & Kryskov, A. A. (2021). Research of professional responsibility of students of technical specialities by means of information and communication technologies. *Journal of Physics: Conference Series, 1840*(1), 012058. https://doi.org/10.1088/1742-6596/1840/1/012058

Miller, A. L., Fassett, K. T., & Palmer, D. L. (2021). Achievement goal orientation: A predictor of student engagement in higher education. *Motivation and Emotion, 45,* 327–344. https://doi.org/10.1007/s11031-021-09881-7

Mohammad Shafi, M., Neyestani, M. R., Jafari, S. E. M., & Taghvaei, V. (2021). The quality improvement indicators of the curriculum at the technical and vocational higher education. *International Journal of Instruction, 14*(1), 65–84. https://doi.org/10.29333/iji.2021.1415a

Monge, C., Montalvo, D., Santos, F., & Torrego, J. C. (2023). Teaching innovation factors: Analysis from three Ibero-American countries. *Educational Studies, 51*(4), 482–502. https://doi.org/10.1080/03055698.2023.2256920

MOOC China. (2023). *MOOC China - Online Learning Platform*. https://www.icourse163.org/

Nambiar, D. (2020). The impact of online learning during COVID-19: Students' and teachers' perspective. *The International Journal of Indian Psychology, 8*(2), 783–793. https://doi.org/10.25215/0802.094

Niyomves, B., Kunacheva, N., & Sutadarat, S. (2024). Hybrid learning: A combination of face-to-face and online learning. *Journal of Education and Learning Reviews*, *1*(3), 11–20. https://doi.org/10.60027/jelr.2024.673

Nurhasnah, N., Sepriyanti, N., & Kustati, M. (2024). Learning theories according to constructivism theory. *Journal International Inspire Education Technology*, *3*(1), 19–30. https://doi.org/10.55849/jiiet.v3i1.577

Owston, R., York, D. N., Malhotra, T., & Sitthiworachart, J. (2020). Blended learning in STEM and non-STEM courses: How do student performance and perceptions compare? *Online Learning, 24*(3), 203–221. https://doi.org/10.24059/olj.v24i3.2151

Potra, S., Pugna, A., Pop, M. D., Negrea, R., & Dungan, L. (2021). Facing COVID-19 challenges: 1st-year students' experience with the Romanian hybrid higher educational system. *International Journal of Environmental Research and Public Health, 18*(6), 3058. https://doi.org/10.3390/ijerph18063058

Rafiola, R., Setyosari, P., Radjah, C., & Ramli, M. (2020). The effect of learning motivation, self-efficacy, and blended learning on students' achievement in the Industrial Revolution 4.0. *International*

Journal of Emerging Technologies in Learning, 15(8), 71–82. https://doi.org/10.3991/ijet.v15i08.12525

Reis, S. M., Westberg, K. L., Kulikowich, J. M., & Purcell, J. H. (2021). Curriculum compacting and achievement test scores: What does the research say? In J. Renzulli & S. M. Reis (Eds.), *Reflections on gifted education* (pp. 271–284). Routledge. https://doi.org/10.4324/9781003237693-15

Ren, W., & Ji, J. (2021). How do environmental regulation and technological innovation affect the sustainable development of marine economy: New evidence from China's coastal provinces and cities. *Marine Policy, 128,* 104468. https://doi.org/10.1016/j.marpol.2021.104468

Sayfullayeva, D. A., Tosheva, N. M., Nematova, L. H., Zokirova, D. N., & Inoyatov, I. S. (2021). Methodology of using innovative technologies in technical institutions. *Annals of the Romanian Society for Cell Biology, 25*(4), 7505–7522. https://annalsofrscb.ro/index.php/journal/article/view/3397

Shadiev, R., Yi, S., & Altinay, F. (2024). Cultivating self-directed learning abilities in K–12 students through immersive online virtual tours. *Interactive Learning Environments, 32*(10), 7338–7363. https://doi.org/10.1080/10494820.2024.2312923

Shen, F., Roccosalvo, J., Zhang, J., Tian, Y., & Yi, Y. (2023). Online technological STEM education project management. *Education and Information Technologies, 28*(10), 12715–12735. https://doi.org/10.1007/s10639-022-11521-7

Singh, J., Evans, E., Reed, A., Karch, L., Qualey, K., Singh, L., & Wiersma, H. (2022). Online, hybrid, and face-to-face learning through the eyes of faculty, students, administrators, and instructional designers: Lessons learned and directions for the post-vaccine and post-pandemic/COVID-19 world. *Journal of Educational Technology Systems, 50*(3), 301–326. https://doi.org/10.1177/00472395211063754

Singh, J., Steele, K., & Singh, L. (2021). Combining the best of online and face-to-face learning: Hybrid and blended learning approach for COVID-19, post vaccine, & post-pandemic world. *Journal of Educational Technology Systems, 50*(2), 140–171. https://doi.org/10.1177/00472395211047865

Stevens, G. J., Bienz, T., Wali, N., Condie, J., & Schismenos, S. (2021). Online university education is the new normal: But is face-to-face better? *Interactive Technology and Smart Education, 18*(3), 278–297. https://doi.org/10.1108/ITSE-08-2020-0181

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones II, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, . . . Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences, 117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117

Valieiev, R., Pokaichuk, Y., Zhbanchyk, A., Polyvaniuk, V., Nykyforova, O., & Nedria, K. (2021). In the search for the golden mean: Students satisfaction with face-to-face, blended and distance learning. *Revista Romaneasca Pentru Educatie Multidimensionala, 13*(1), 20–40. https://doi.org/10.18662/rrem/13.1/357

Wahas, Y. M. A., & Syed, A. J. A. (2024). E-assessment challenges during e-learning in higher education: A case study. *Education and Information Technologies, 29,* 14431–14450. https://doi.org/10.1007/s10639-023-12421-0

Wang, T., & Bhagat, K. K. (2025). Theoretical foundations, models, and frameworks of blended learning. In *Case studies on blended learning in higher education: Design, development, and delivery* (pp. 19–40). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-0722-8_2

Wu, T., Yang, S., & Tan, J. (2020). Impacts of government R&D subsidies on venture capital and renewable energy investment: An empirical study in China. *Resources Policy, 68,* 101715. https://doi.org/10.1016/j.resourpol.2020.101715

Zhang, L., Ma, X., Ock, Y. S., & Qing, L. (2022). Research on regional differences and influencing factors of Chinese industrial green technology innovation efficiency based on dagum gini coefficient decomposition. *Land, 11*(1), 122. https://doi.org/10.3390/land11010122

International Review of Research in Open and Distributed Learning
Volume 27, Number 1

February – 2026

Exploring Cognitive Presence in Online Collaborative Knowledge-Building: Structural, Temporal, and Social Perspectives

Xieling Chen[1], Huimei Chen[1], Di Zou[2*], Haoran Xie[3], and Fu Lee Wang[4]
[1]School of Education, Guangzhou University, Guangzhou, China; [2]Department of English and Communication, The Hong Kong Polytechnic University, Hong Kong SAR; [3]School of Data Science, Lingnan University, Hong Kong SAR; [4]School of Science and Technology, Hong Kong Metropolitan University, Hong Kong SAR; [*]Corresponding Author

Abstract

Collaborative knowledge-building is an important mode of learning in which students' cognitive presence has a significant impact on learning outcomes. To better understand how cognitive presence influences collaborative learning, this study applied three complementary analytic approaches: epistemic network analysis, which maps how ideas are connected in discussions; sequential pattern mining, which identifies temporal sequences; and social network analysis, which examines the interaction patterns and roles among group members. Using data from 37 students divided into 8 groups in a university course on academic reading and writing, we compared high-performing groups (HPGs) and low-performing groups (LPGs). The results showed that HPGs demonstrated stronger exploratory, integrative, and problem-solving abilities in their cognitive networks, with members actively exchanging ideas, questioning, and summarizing. In contrast, LPGs relied more on encouragement and reminders to sustain discussions. Furthermore, HPGs displayed more complex and varied behavioral sequences and clearer leadership and facilitation roles within their social networks, whereas LPGs showed simpler and less developed interaction patterns and lacked core members in their networks. These findings provide insights for instructors on how to better design and guide group knowledge-building to enhance online collaborative learning outcomes.

Keywords: online collaborative knowledge-building, cognitive presence, structural pattern, temporal sequence, social connection

Introduction

Collaborative knowledge-building, as a learning mode that emphasizes knowledge co-construction through sustained dialogue, mutual support, and shared responsibility among learners, aligns well with the pedagogical shift from teacher-centered instruction to learner-centered and interaction-driven approaches. This process fosters not only the development of domain-specific knowledge but also higher-order skills and competencies such as critical thinking, communication, and teamwork, which are essential in the 21st-century learning paradigm (Jiang et al., 2023). For example, in engaging students in working together in teams to participate in tasks, reach consensus through discussion, share resources and ideas, and correct misconceptions, collaborative knowledge-building contributes to students' development of a deeper understanding of new knowledge.

Central to the effectiveness of collaborative knowledge-building is cognitive presence, defined as the extent to which learners can construct and confirm meaning through sustained reflection and discourse (Baanqud et al., 2020). Effective cognitive engagement is critical for supporting higher-order thinking in collaborative processes (Chen et al., 2025). However, variations in the structure, dynamics, and social patterns of cognitive presence can affect learning outcomes (Moon et al., 2024). Students who struggle to engage cognitively or interact effectively within groups may experience lower academic performance; thus, it is important to understand these patterns (Liu et al., 2022).

While previous studies have underscored the importance of cognitive presence as a critical indicator of deep learning in collaborative settings (Morueta et al., 2016), several issues remain. Many prior studies have relied on static or linear analyses (Bhutoria & Aljabri, 2022), failing to capture the complex, dynamic, and socially situated nature of learners' cognitive presence. There is also limited research on comparing cognitive presence structures, temporal patterns, and social interaction dynamics across learners with different academic performance levels. These gaps suggest an incomplete understanding of how cognitive presence unfolds and difficulty in identifying strategies to support at-risk learners due to less effective cognitive and social engagement.

To address these challenges, this study proposed a multidimensional analytical framework that integrated epistemic network analysis (ENA), sequential pattern mining (SPM), and social network analysis (SNA) to investigate cognitive presence in online collaborative knowledge-building. Specifically, this study identified and quantified the connections among coded cognitive presence elements by visualizing structural patterns across high-performing groups (HPGs) and low-performing groups (LPGs). Second, temporal sequences of cognitive activities were visualized through sequential analysis to understand how learners' engagement evolved during collaboration. In addition, interaction patterns and members' positions were examined through social network visualization to reveal relational dynamics that influence cognitive development. Accordingly, this study aimed to address three research questions:

RQ1: What are the structural differences in cognitive presence between HPGs and LPGs, and how can these be identified through ENA and validated statistically?

RQ2: How do sequential patterns of cognitive presence differ between HPGs and LPGs?

RQ3: What are the differences in social network characteristics between HPGs and LPGs, and how can these be identified through SNA and validated statistically?

This study aimed to provide a fine-grained, multi-layered analysis of cognitive presence, capturing structural, temporal, and social dimensions. Findings are expected to inform instructional design to optimize cognitive engagement and academic performance for diverse learners.

Literature Review

Collaborative Knowledge-Building and Cognitive Presence

Collaborative knowledge-building, as a process in which individuals work together to form wisdom products such as opinions, ideas, and methods, represents a shift from traditional didactic instruction toward learner-centered, inquiry-based pedagogies that emphasize knowledge co-construction through social interaction, discourse, and reflective thinking (Shea et al., 2022). Within this context, cognitive presence, rooted in the community of inquiry (CoI) framework (Wilson & Berge, 2023), refers to learners' capacity to construct and confirm meaning through sustained reflection and dialogue (Maranna et al., 2022). In the CoI framework, there are four progressive phases of cognitive presence: triggering event, exploration, integration, and resolution, which capture how learners move from encountering a problem to proposing actionable solutions, highlighting a pathway for deep, meaningful learning.

In collaborative learning contexts, where learners co-construct meaning through processes such as idea-sharing, negotiation, conflict resolution, and joint problem-solving, cognitive presence is closely intertwined with social interaction and metacognitive regulation (Moon et al., 2024). As conceptualized by Sharma et al. (2024), collaborative knowledge-building, which involves iterative cycles of sharing, arguing, negotiating, creating, and reflecting, underscores the complexity of cognitive engagement within social systems. In analyzing the cognitive presence, existing research has often relied on linear or static models; thus, the challenge remains in fully capturing the dynamics of collaborative knowledge-building in real-world learning environments from temporal and networked perspectives.

ENA and Its Application in Education

ENA, as a methodological innovation in learning analytics, has attained attention for its ability to model and visualize complex cognitive structures over time by analyzing coded discourse or behavioral data to construct network models that quantify the strength and structure of connections among epistemic elements. Its visual representation ability allows the visualization of the development process of thinking and cognitive change rules, thus offering insights into how learners' ideas evolve, cohere, and relate to one another during collaborative learning (Elmoazen et al., 2024).

ENA has been applied in varied contexts such as collaborative problem-solving, multi-task learning, and peer assessment. For instance, by analyzing students' problem-solving process, Gao et al. (2022) found that the communication between excellent group members was more active, showing stronger social cognition and a sense of responsibility. Fougt et al. (2018) analyzed students' essays to understand the differences in cognitive structure between high-quality and low-quality essays. However, ENA has been underused in studies that explicitly compare cognitive presence structures between students of varying academic performance levels, particularly within naturally occurring student discussions. This study addresses this gap by applying ENA to contrast the cognitive structures of HPGs and LPGs in online collaborative discussions.

SPM and Its Application in Education

SPM, including methods such as lag sequential analysis, which are particularly effective for uncovering hidden behavioral and cognitive patterns that unfold over time, allows researchers to examine temporal and logical dependencies between events or behaviors in learning processes in the field of education. For instance, Cheng et al. (2022) tracked online learners' activity flows for predicting learning achievement based on behavioral sequences. He et al. (2021) used sequence-mining of process data from a large-scale assessment to identify test-takers' problem-solving strategies, finding that optimal strategies were linked to higher performance, with older adults and women more likely to use sub-optimal approaches. Despite its effectiveness, SPM has seldom been applied to examine differences in cognitive presence between HPGs and LPGs in the context of collaborative group discussions. Thus, this study explored how learners' cognitive trajectories diverged during online collaboration to allow a more nuanced understanding of the temporal aspects of cognitive engagement.

SNA and Its Application in Education

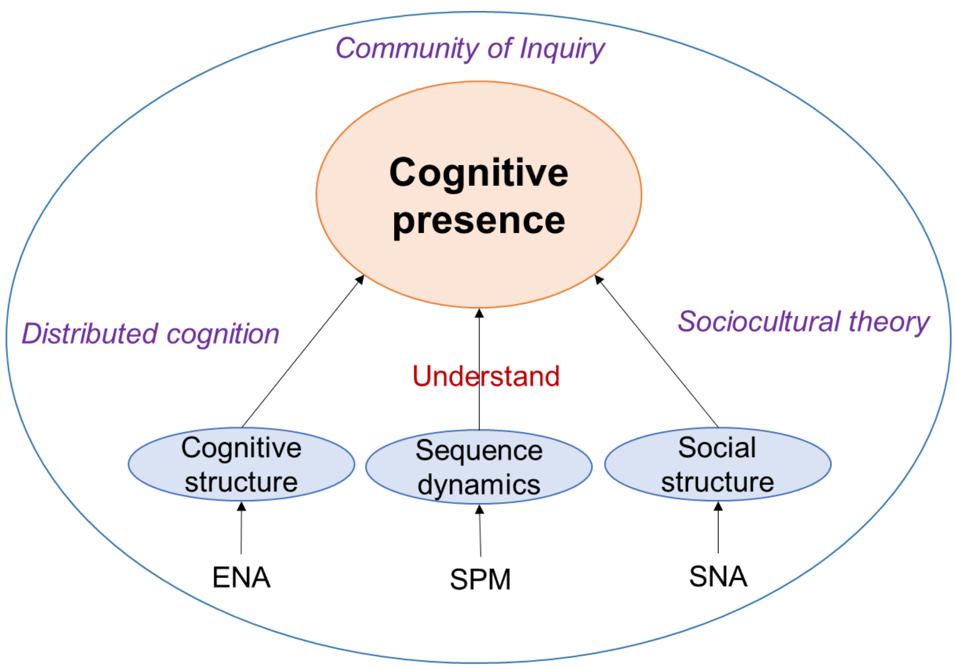
SNA, by mapping and analyzing the relationships among individuals in a learning community, is able to capture interaction patterns, information flow, and positional dynamics through metrics such as degree centrality, betweenness centrality, and network density to provide a structural perspective on learning. For instance, Zou et al. (2021) used machine learning and network analytics to show that positive social presence indicators in MOOC forum posts—such as asking questions or expressing gratitude—are linked to higher learner prestige, while negative expressions reduce it, highlighting the effect of social behavior on online learning dynamics. Norz et al. (2023) analyzed social presence in online learning and found significant relationships that can help track social presence in real time. However, there remains a lack of research that compares social interaction patterns in relation to cognitive presence and performance levels. Also, while studies have explored general interaction frequency or network centrality, few have linked these structural features to specific phases of cognitive presence or learning outcomes. This study sought to bridge these gaps by examining how social network characteristics varied between HPG and LPG to understand how social positioning relates to cognitive engagement.

Theoretical Perspectives

Cognitive presence encompasses inquiry-related discourse (e.g., triggering events, exploration, integration, resolution), is increasingly shaped by the temporal sequencing of inquiry processes (sequence dynamics), and is influenced by social affordances such as interaction patterns, frequency, and participant roles (Chen et al., 2025). Our study adopted a framework that linked theoretical perspectives to analytic lenses (see Figure 1) by integrating CoI, sociocultural theory, and distributed cognition. In our framework, cognitive presence was understood as sustained inquiry, learning was viewed as social and culturally situated, and knowledge was seen as embedded in interactions and social relations. Specifically, ENA captures the structure of cognitive elements and idea integration, SPM represents the temporal sequencing of inquiry moves, and SNA depicts the social structure and pathways of scaffolding and information flow. These lenses enable multilevel mapping from interactions, as shown in micro-level transcripts, through processes in meso-level sequences and networks, to outcomes reflected in group performance and sustained cognitive presence.

Figure 1

Integrative Theoretical and Analytic Framework Used in This Study



Note. ENA = epistemic network analysis; SPM = sequential pattern mining; SNA = social network analysis

Research Design

This study employed a case study of a higher-education course to examine how 37 third-year Educational Technology majors, grouped by academic performance, engaged cognitively and socially in online collaborative knowledge-building using a multi-method approach combining ENA, SPM, and SNA. To examine whether observed patterns differed statistically between HPGs and LPGs, we conducted multivariate analyses of variance (MANOVAs), complementing our study's qualitative, exploratory design with quantitative analyses that provided statistical support for observed differences.

Context and Data Collection

The participants were enrolled in the course Academic Reading and Writing. Students were assigned to eight groups (four to six members each), with each group working in its own dedicated WeChat group chat. The instructor posted one open-ended discussion prompt every 2 weeks to all groups simultaneously to stimulate inquiry, argumentation, and collaborative meaning-making. Students responded asynchronously within their group chat, contributing ideas, asking questions, and engaging in collaborative reasoning throughout the week. The five prompts were: (a) How can information literacy be improved? (b) Why is a literature review important in academic research? (c) What techniques are effective for reading English-language scholarly articles? (d) Does critical thinking necessarily imply rejecting or dismissing others' ideas? (e) Is it necessary to include every article in a literature review? Chat transcripts that captured all learner contributions, reflecting both cognitive and social dimensions of participation in a naturally occurring, text-based collaborative environment, were collected.

At the end of the semester, groups submitted a joint assignment that required them to analyze and critique a scholarly article and present their insights to the class. Course grades, based on the quality of this assignment, were used to categorize the top four performing groups as HPGs and the bottom four as LPGs for comparative analysis.

Data Coding

To systematically examine cognitive presence in online group discussions, a qualitative content analysis was conducted using a structured coding scheme adapted from the CoI framework and prior research (e.g., Ba et al., 2023; Zhang et al., 2022). The scheme encompassed five primary cognitive themes: triggering event, exploration, integration, resolution, and encouragement, and comprised 17 specific behavior codes (see Table 1).

A total of 521 comments were coded by two trained coders with backgrounds in educational technology and experience in discourse analysis, following a three-step approach. First, the two coders participated in a joint training session where they thoroughly reviewed the coding scheme and conducted a pilot coding round on a randomly selected 10% of the dataset; subsequently, the coders independently applied the codes and discussed discrepancies to align their interpretations and refine category boundaries as needed. After achieving initial consensus, the coders independently coded the full dataset, considering both linguistic content and contextual cues. To assess the consistency of coding between the two coders, Cohen's kappa coefficient was computed, which was 0.94, indicating a high level of agreement and suggesting that the coding scheme was reliably applied. Discrepancies were resolved through discussion and consensus, and the coded data were used for data analysis.

According to the frequency distribution presented in Table 1, the analysis shows a clear emphasis on exploration (33%) and integration (39%), with resolution also well-represented (26%), whereas triggering (3.64%) behaviors were comparatively rare. This reflects that the students actively shared and built ideas in online collaborative learning, although more support may be needed to foster deeper inquiry and thinking.

Table 1

Coding Scheme for Discourse Analysis to Understand Cognitive Presence in Online Group Discussions

Cognitive theme	Behavior codes	Code	Description	Example	Frequency, *n* (%)
Triggering	Clarification	CL	Clarify	"Clarifying, we are discussing literature searches."	14 (2.69)
	Restating	RE	Restate	"Let me repeat that point."	5 (0.95)
Exploration	Agreement	AWS	Agreement without substantiation	"I think so."	46 (8.83)
	Information sharing	IS	Stating a fact, policy, or rule; citing a source	"Here is some news from the Internet for reference."	60 (11.52)
	Divergence	DIS	Disagreement	"I don't agree."	1 (0.19)
	Personal narration	PN	Story, relating an incident, describing practices at their job	"I usually translate passages whole."	6 (1.15)
	Opinion	OP	Belief, judgment, personal view, or attitude based on grounds insufficient to conclude factual	"I think we should pay closer attention to information and its associations."	56 (10.75)
Integration	Building on	BO	Augmenting a point made by self or by another earlier	"To build on these foundations, strengthen practice, explore new skills, and stay current with emerging technologies."	84 (16.12)
	Creating solutions	CS	Novel conclusion	"So, reading the literature turns ignorance into insight, revealing solutions otherwise overlooked."	29 (5.57)
	Justified hypothesis	JH	A tentative assumption made to draw out and test its logical consequence to prove or show to be reasonable; coming to a conclusion predicted by ongoing discussion but supported with relevant reasons	"If we neglect the paper, conclusions risk vagueness; careful attention should yield clarity and coherence."	32 (6.14)
	Supported divergence	SD	Disagreement with the reason stated	"I disagree, since some papers contain information unrelated to our project."	7 (1.34)
	Supported agreement	SA	Agreement with reason stated	"Yes, a process is needed, involving keyword searches and filtering results."	9 (1.73)
	Wrap-up	WU	Concluding; summarizing	"So there are three things: get the information, understand the information, and apply it."	39 (7.49)

Cognitive theme	Behavior codes	Code	Description	Example	Frequency, *n* (%)
Resolution	Thought experiment	TE	Questioning "What if?" or "What do you think about?"	"What if we search papers in different databases?"	45 (8.64)
	Apply, test, and defend	ATD	Any of the three, but not retrospective narrative; must be an application of new thought initiated by the discussion present	"To keep pace with society, we should track social trends and emerging computer technologies to strengthen our information literacy."	51 (9.79)
Encourage-ment	Encouraging answer	EA	Encouraging answer	"What do you think?"	37 (7.10)

Data Analysis

Data analysis proceeded in three stages, each aligned with a specific research question, employing a multi-method approach that integrated qualitative coding with quantitative modeling and statistical testing to provide detailed process insights alongside comparative group-level validation.

To examine structural patterns of cognitive presence (RQ1), ENA was employed to model the co-occurrence of coded cognitive presence elements within student discussions. ENA constructs networks in which nodes represent specific cognitive elements (e.g., triggering event, exploration, integration, resolution) and edges represent their co-occurrence within the same discourse segment, such that frequently co-occurring elements form stronger connections, visually depicting group thinking structures. To assess whether these structural patterns differed between HPGs and LPGs, a MANOVA was adopted to test differences in the frequency of the five main cognitive presence themes.

Second, SPM was employed to identify recurring sequences of cognitive presence moves (RQ2), capturing the evolution of inquiry processes in online discussions. Learners' activities were analyzed as ordered behavior sequences (e.g., posting a question → providing explanation → making a connection) to highlight patterns occurring more frequently than expected by chance. In the resulting behavioral transition diagram, nodes represent specific behaviors, with size proportional to frequency, while directed arrows indicate significant transitions and line thickness reflects the magnitude of adjusted residuals (Z-scores), allowing typical discourse trajectories and critical inquiry transitions to be observed for each group.

In addition, SNA was used to examine interactional structures and individual centrality within groups (RQ3). In the resulting network diagrams, nodes represent members, lines indicate interactions with thickness proportional to communication frequency, and arrows show the direction of information flow. Centrality indices quantified each member's role and influence, and group-level differences were tested via MANOVA to determine whether HPGs and LPGs differed in social structure.

This multi-method study captures the structural, temporal, and social dimensions of cognitive presence in online collaborative learning, using MANOVA to confirm whether observed structural and social differences are statistically significant, and provides a comprehensive view of its manifestation and evolution.

Results

Structural Patterns of Cognitive Presence (RQ1)

According to Table 2, which summarizes the frequency of cognitive presence themes, the average frequency of all cognitive presence themes is consistently higher in the HPGs compared to the LPGs. For example, regarding exploration, HPGs averaged 60.5 instances, more than double the LPG average of 24.5. Considering integration, HPGs (79.5) demonstrated significantly more behaviors related to synthesizing and building upon ideas than LPGs (41.25), highlighting a more advanced level of meaning-making. In terms of resolution, HPGs showed stronger evidence of applying and testing knowledge, averaging 99 instances versus 33 in LPGs. In addition, regarding triggering, it was absent in LPGs (0), whereas HPGs engaged in some level of initiating inquiry (4.5), reflecting a stronger ability or tendency to pose problems or raise questions during discussion.

Table 2

Mean Frequency of the Five Cognitive Themes

Theme	Group	
	High performing	Low performing
Triggering	4.5	0
Exploration	60.5	24.5
Integration	79.5	41.25
Resolution	99	33
Encouragement	27.5	22.5

A MANOVA analysis was further applied to see if the frequency of the five main themes differed for HPGs and LPGs, with results detailed in Table 3. The Box's M test [χ^2 (15, n = 36) = Inf, p < 2.2e-16] and Pillai's Trace test revealed a significant MANOVA effect [Pillai's Trace = 0.59473, F(1, 36) = 8.8049, p = 3.145e-05] and suggest statistically significant differences between HPGs and LPGs. The univariate tests showed significant differences between groups, as shown in Table 3.

Table 3

Results of the Multivariate Analysis of Variance (MANOVA) on the Frequency of Cognitive Themes Between HPGs and LPGs

Cognitive themes	Group				Univariate test		
	High-performing		Low-performing				
	M	*SD*	*M*	*SD*	*F*	η^2	*Sig.*
Triggering	1	0.268	0	0	6.70*	.165	0.014
Exploration	6.72	0.939	2.72	0.463	13.9***	.29	0.000
Integration	5.89	1.02	3.06	0.392	14.6***	.3	0.000
Resolution	5.5	0.772	1.83	0.232	20.7***	.378	0.000
Encouragement	1.22	0.348	1	0.302	0.232	.007	0.633

Note. ***: p < 0.001; **: p < 0.01; *: p < 0.05.

Table 4 provides a more granular analysis of cognitive behaviors within each group and compares the average percentage of these behaviors between HPGs and LPGs. The percentages were calculated by dividing the mean frequency of each category by the total mean frequency for the respective group (HPGs or LPGs), then multiplying by 100. For example, in the HPGs, the frequency of clarification was 3.5, and the total frequency was 91.5. The percentage is therefore calculated as: 3.5/91.5 × 100 = 3.83%. Results in Table 4 show that agreement without substantiation occurred more in HPGs than in LPGs, indicating a greater tendency among HPGs to reach consensus, even if not fully reasoned. Also, thought experiment behaviors were significantly more frequent in HPGs than in LPGs, suggesting a more active engagement in speculative reasoning and hypothetical thinking in HPGs. Regarding resource sharing and elaboration, LPGs exhibited higher proportions of behaviors of information sharing, opinion, and building on than HPGs, suggesting that while LPGs were highly engaged in idea sharing and elaboration, they might have lacked deeper synthesis and critical evaluation compared to HPGs.

Table 4

Mean Frequency of Behavioral Codes Between HPGs and LPGs

Behavioral Codes	*Abbreviations*	Group			
		High performing		Low performing	
		n	%	*n*	%
Clarification	CL	3.5	3.83	0	0
Restating	RE	1	1.09	0.25	0.65
Agreement	AWS	9.25	10.11	2.25	5.81
Information sharing	IS	10	10.93	5	12.90
Divergence	DIS	0	0.00	0.25	0.65
Leap to conclusion	LTC	0	0.00	0	0.00
Personal narration	PN	1.5	1.64	0	0.00
Opinion	OP	9.5	10.38	4.5	11.61
Building on	BO	13.5	14.75	7.5	19.35
Creating solution	CS	5	5.46	2.25	5.81
Justified hypothesis	JH	4.75	5.19	3.25	8.39
Supported divergence	SD	1.5	1.64	0.25	0.65
Supported agreement	SA	1.75	1.91	0.5	1.29
Wrap-up	WU	5.5	6.01	4.25	10.97
Thought experiment	TE	10.5	11.48	0.75	1.94
Apply, test, and defend	ATD	8.75	9.56	4	10.32
Encouraging answer	EA	5.5	6.01	3.75	9.68
Total		91.5	100	38.75	100

The contrasting cognitive network structures presented in Figure 2 reveal that HPGs were characterized by cognitively rich, conceptually integrated discussions, while LPGs tended to rely more heavily on surface-level behaviors and motivational prompts to maintain dialogue. For example, for the HPGs, the centrality of building on (BO) among their cognitive behaviors suggests that members frequently engaged in elaborating and expanding upon others' ideas, indicating deeper cognitive processing and collaborative meaning-making. The strongest co-occurrence connections reflected by the darkest lines in the network were observed between BO and opinion (OP), BO and wrap-up (WU), thought experiment (TE) and BO, and apply, test, and defend (ATD) and BO, demonstrating that group members not only extended discussions but also effectively integrated personal opinions, summarized conclusions, engaged in hypothetical reasoning, and applied conceptual knowledge. Additionally, justified hypothesis (JH), TE, and

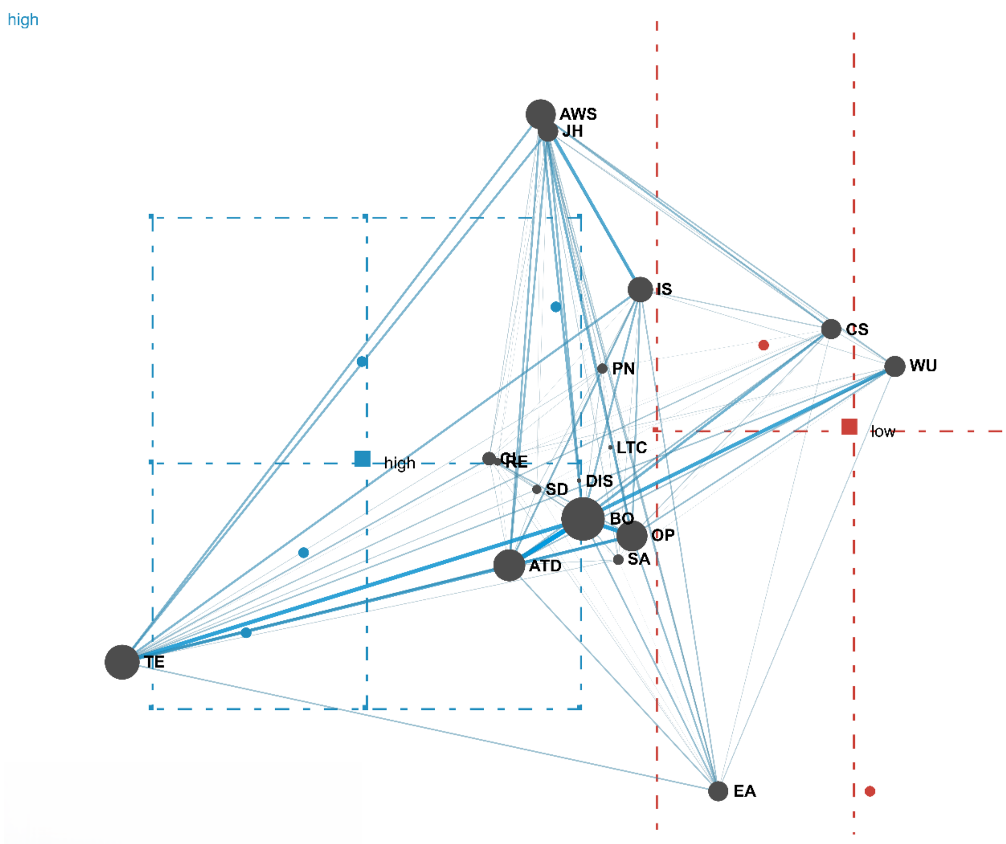
ATD exhibited high connectivity with others, suggesting that introducing conceptual challenges or application-based scenarios facilitated the emergence of higher-order thinking and sustained the cognitive engagement of the group (see Morueta et al., 2016).

In contrast, the cognitive network of the LPGs shows a markedly less complex structure, with a lower density of connections between behavioral codes. For example, the strongest behavioral associations between OP and BO, BO and WU, and BO and encouraging answer (EA) indicate a tendency among LPG members to move from expressing opinions to building on them and summarizing, often with peer encouragement as a mediating factor. Furthermore, the relatively darker lines connecting EA to BO, OP, and WU imply that encouragement behaviors were instrumental in sustaining interaction and prompting further contributions, albeit possibly at the expense of more advanced integrative or exploratory behaviors (see Zou et al., 2021).

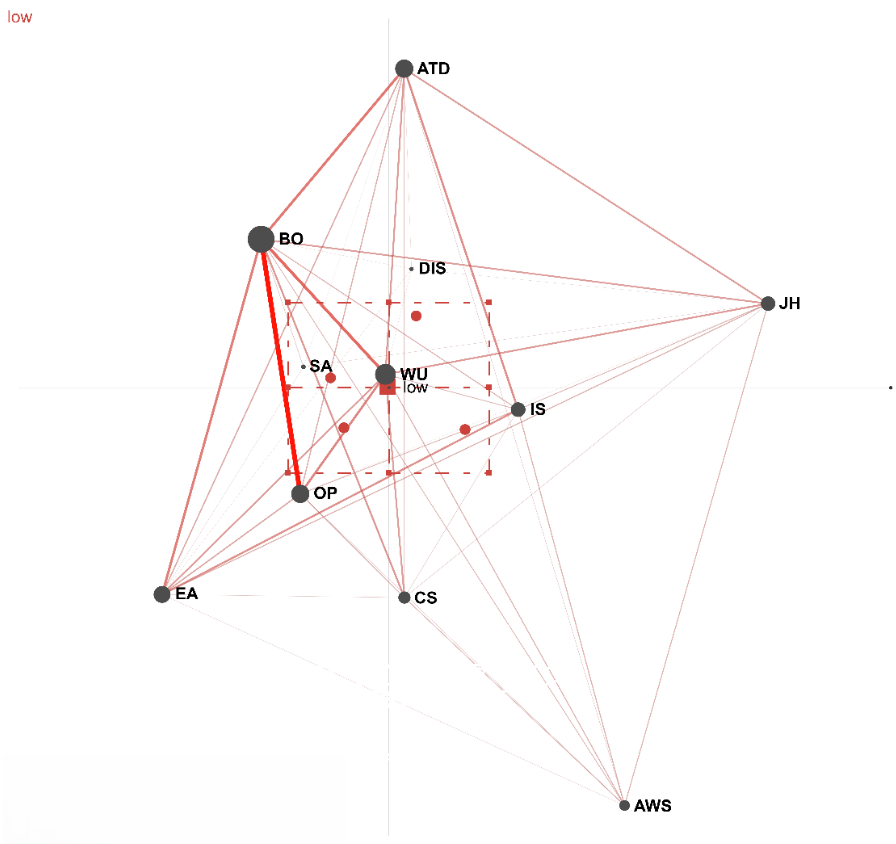
Figure 2

Cognitive Network Structures of HPGs and LPGs

A



B

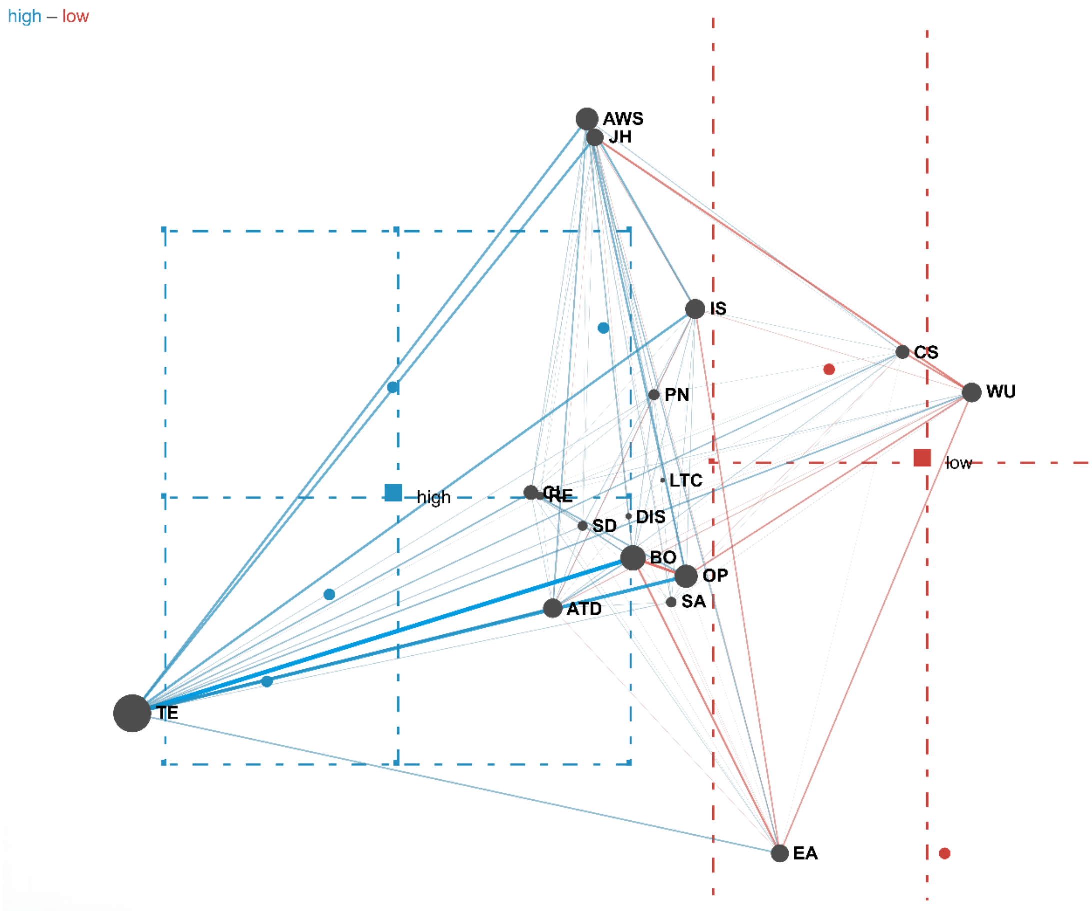


Note. HPG = high-performing group; LPG = low-performing group. Panel A: Interactions between behavior codes in HPGs. Panel B: Interactions between behavior codes in LPGs. ATD = apply, test, and defend; AWS = agreement; BO = building on; CL = clarification; CS = creating solutions; DIS = divergence; EA = encouraging answer; IS = information sharing; JH = justified hypothesis; LTC = leap to conclusion; OP = opinion; PN = personal narration; RE = restating; SA = supported agreement; SD = supported divergence; TE = thought experiment; WU = wrap-up. Blue dots and lines represent the interactions within the HPG, while red dots and lines represent the interactions within the LPGs. The thicker lines indicate stronger interactions between behavior codes.

According to Figure 3, which presents the superimposed subtraction plots of the cognitive networks for HPGs and LPGs, HPGs demonstrated a more integrated and exploratory cognitive discourse structure compared to the more interactionally supportive yet cognitively superficial patterns observed in LPGs. For example, for HPGs, TE, which emerges as a central anchor, exhibited strong connections with ATD, OP, BO, agreement (AWS), and JH. Conversely, for the LPGs, the most prominent behavioral pairings of OP-BO, JH-WU, and BO-EA suggest that members tended to focus more on basic opinion expression and summarization, with encouragement behaviors playing a notable role in sustaining participation rather than fostering conceptual elaboration or cognitive engagement (Chen et al., 2022).

Figure 3

Superimposed Subtraction Plots of the Cognitive Networks for HPGs and LPGs



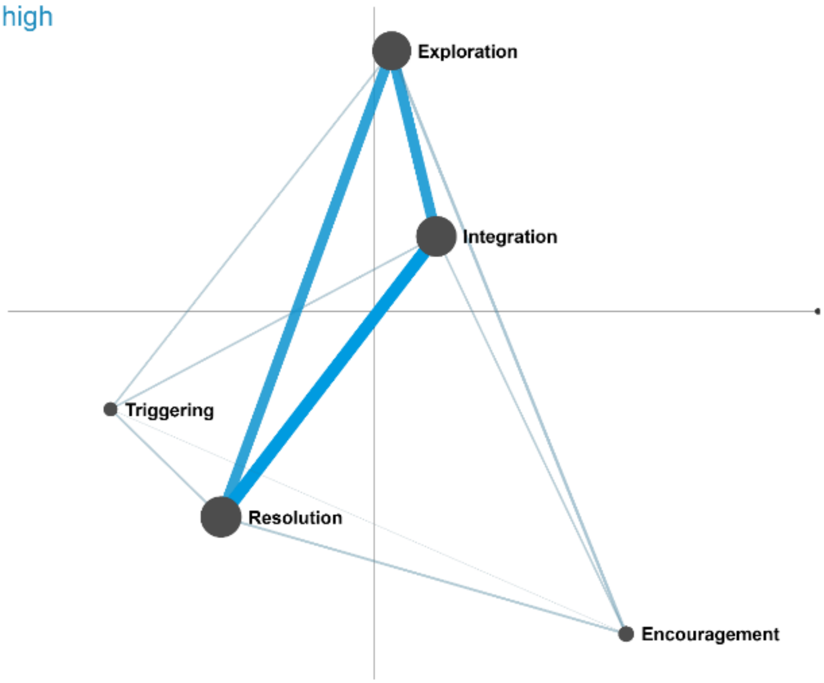
Note. HPG = high-performing group; LPG = low-performing group. ATD = apply, test, and defend; AWS = agreement; BO = building on; CL = clarification; CS = creating solutions; DIS = divergence; EA = encouraging answer; IS = information sharing; JH = justified hypothesis; LTC = leap to conclusion; OP = opinion; PN = personal narration; RE = restating; SA = supported agreement; SD = supported divergence; TE = thought experiment; WU = wrap-up. The figure represents the interactions between different behavior codes across HPGs and LPGs. The blue dots and blue dashed squares represent the HPGs, while the red dots and red dashed squares represent the LPGs. The grid lines in the figure correspond to the boundaries of the groups, separating high and low levels of interaction within the behavior codes. The thicker lines indicate stronger interactions between the behavior codes. The top red line is cut off due to the limits of the figure's display, but this is not an intentional aspect of the data representation.

According to Figure 4, which captures the interactions among the five core cognitive themes (i.e., triggering, exploration, integration, resolution, and encouragement), the HPG network demonstrated more balanced and diverse connections compared to the relatively limited and linear structure observed in the LPG. The LPG network lacks triggering, suggesting that LPGs may have entered directly into exploration without adequate cognitive preparation or question posing, potentially limiting the depth of subsequent discourse (Ba et al., 2023). In both networks, the strongest inter-theme connections are observed between exploration and integration, indicating that participants frequently transitioned from exploring ideas to building on and synthesizing them; the strong connections between integration and resolution and between exploration and resolution also suggest that ideas were carried forward into deeper reasoning and attempts at conclusion or application.

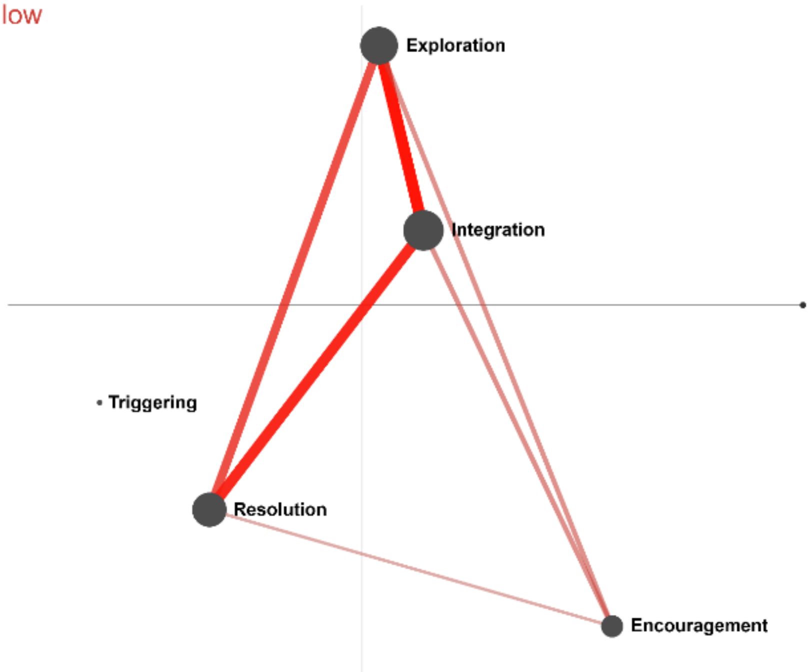
Figure 4

Comparison of Cognitive Theme Interaction Networks Between HPGs and LPGs

A high



B low

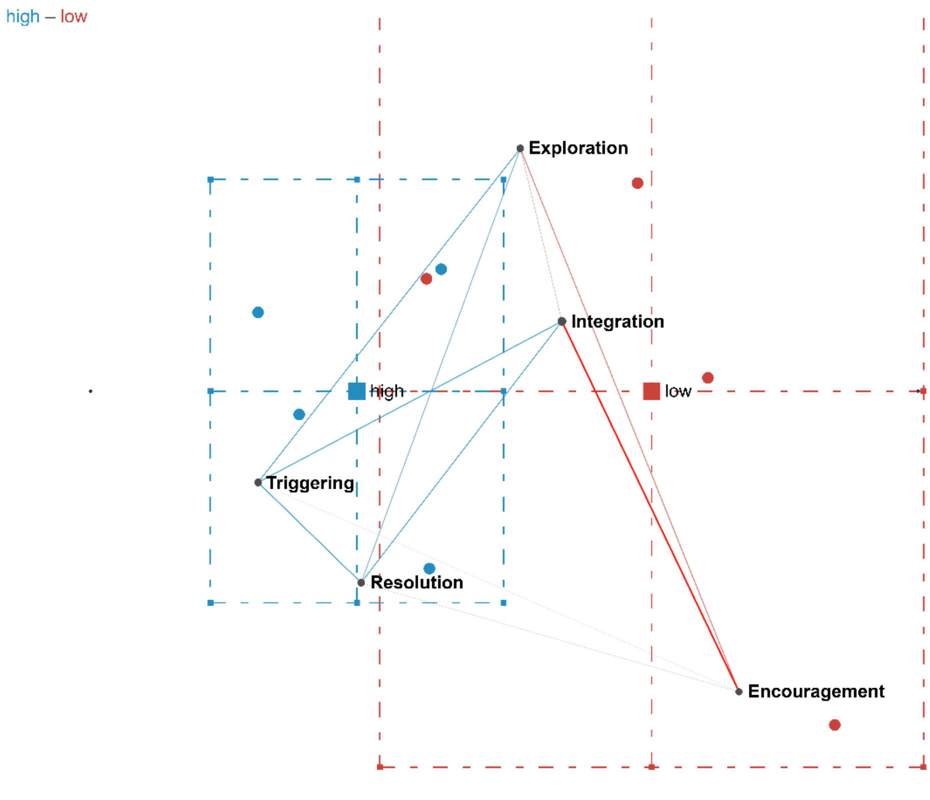


Note. HPG = high-performing group; LPG = low-performing group. Panel A: Interactions between cognitive themes in HPGs. Panel B: Interactions between cognitive themes in LPGs. Blue dots and lines represent the interactions within the HPG, while red dots and lines represent the interactions within the LPGs. The thicker lines indicate stronger interactions between cognitive themes. The grey axes in the background represent the structural framework of the ENA model, used for visualizing the relative positioning and interaction intensity between cognitive themes. These axes do not display actual data points but provide a reference for interpreting the connections and relationships in the network.

Figure 5 shows differential interaction patterns between HPGs and LPGs across the five principal cognitive themes. The stronger connection between integration and encouragement suggests that LPGs may rely more heavily on encouraging discourse when integrating ideas, possibly as a compensatory mechanism to sustain collaborative engagement. In contrast, several connections involving resolution and triggering appear more intense in the HPGs, reflecting a stronger engagement in initiating problem-based thinking and driving discussions toward meaningful conclusions. HPGs thus appear to engage more robustly in the initiation and resolution phases of discourse, whereas LPGs emphasize emotional or motivational support during integration, potentially at the expense of deeper knowledge construction processes.

Figure 5

Combined Cognitive Network for HPGs and LPGs Regarding Main Themes



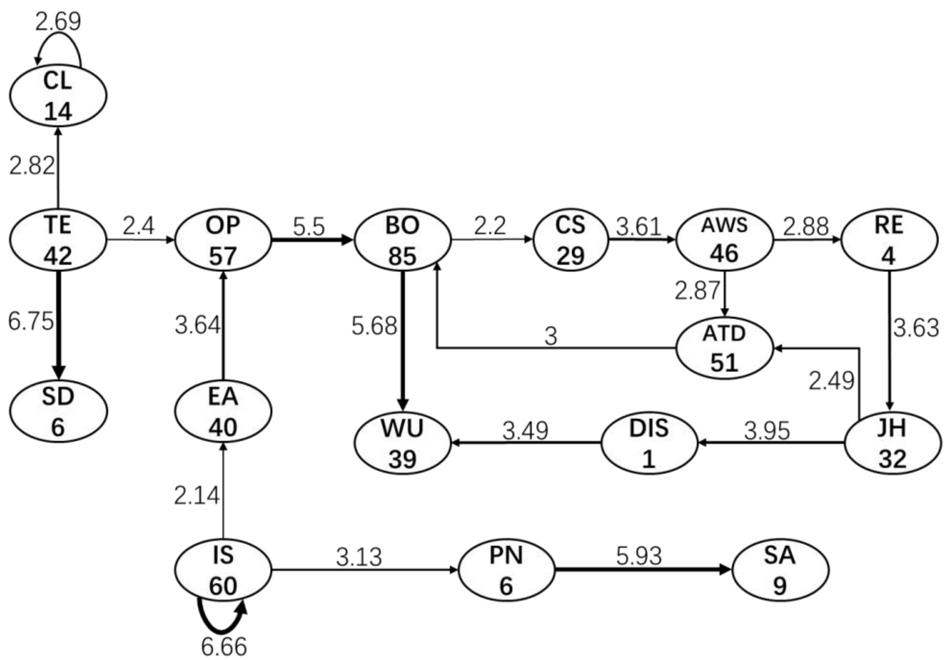
Note. HPG = high-performing group; LPG = low-performing group. The figure represents the interactions between different cognitive themes across HPGs and LPGs. The blue dots and blue dashed squares represent the HPGs, while the red dots and red dashed squares represent the LPGs. The grid lines in the figure correspond to the boundaries of the groups, separating high and low levels of interaction within the cognitive themes. The thicker lines indicate stronger interactions between the cognitive themes. The top red line is cut off due to the limits of the figure's display, but this is not an intentional aspect of the data representation.

Temporal Sequences of Cognitive Presence (RQ2)

Figure 6 presents the overall behavioral transition diagram. The self-cyclic transition of information sharing (IS; IS → IS, Z = 6.66) indicates sustained knowledge exchange among members, which likely involves understanding, integrating, and evaluating peers' contributions—processes that reflect higher-order thinking as emphasized in prior research on online discussion forums (Wu & Wu, 2021). Second, the significant transition from TE to supported divergence (SD; TE → SD, Z = 6.75) suggests that hypothetical questioning often prompted participants to express disagreements with reasoned justification, supporting deeper conceptual elaboration and group understanding (Cherbow & McNeill, 2022). Third, the behavioral shift from personal narration (PN) to supported agreement (SA; PN → SA, Z = 5.93) indicates personal anecdotes' contributions to collective understanding and group cohesion through reasoned affirmation. Additionally, the movement from OP to BO (OP → BO, Z = 5.50) highlights that personal viewpoint expressions facilitated elaboration by peers to enhance both interactive richness and cognitive depth.

Figure 6

Overall Transition Diagram Showing Changes in Behavior Codes Across All Groups



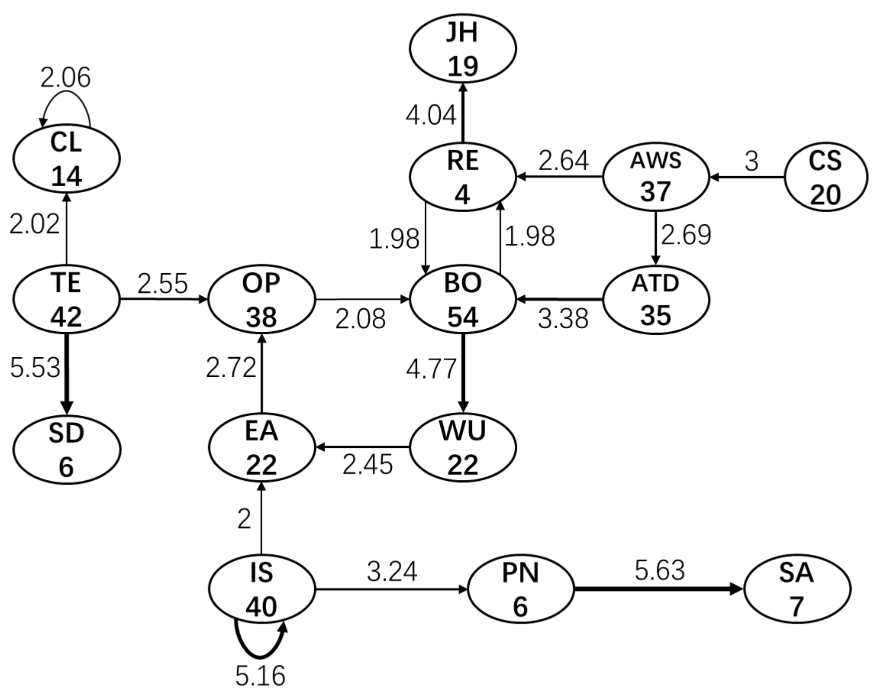
Note. Each node represents a behavior and directed arrows denote significant behavioral transitions, with line thickness reflecting the magnitude of the adjusted residual value (Z-score). ATD = apply, test, and defend; AWS = agreement; BO = building on; CL = clarification; CS = creating solutions; DIS = divergence; EA = encouraging answer; IS = information sharing; JH = justified hypothesis; LTC = leap to conclusion; OP = opinion; PN = personal narration; RE = restating; SA = supported agreement; SD = supported divergence; TE = thought experiment; WU = wrap-up.

Figures 7 and 8, which respectively visualize group-specific behavioral sequences for HPGs (18 significant transitions) and LPGs (9 significant transitions), highlight that HPGs engage in more dialogically rich and cognitively integrated discourse patterns, while LPGs show limited behavioral diversity and elaboration, often relying on more superficial conversational strategies. For example, for HPGs, BO emerges as a central node with high connectivity with restating (RE; Z = 1.98) and WU (Z = 4.77), indicating its pivotal role in facilitating extended discourse, with reflective and elaborative behaviors frequently culminating in summarization. In contrast, LPGs, with an absence of a dominant behavior and unidirectional transitions, suggest fewer reciprocal interactions and more linear and less dynamic discourse patterns.

Furthermore, a cognitively rich path (i.e., IS → PN → SA) in HPGs illustrates the progression from factual exchange to personal contextualization, ultimately reaching consensus; the TE-centered transitions (e.g., TE → OP → BO → WU and TE → SD) also indicate a robust engagement in challenge-response sequences that foster deeper reasoning. In LPGs, while sequences such as creating solutions (CS) → AWS → JH → divergence (DIS) → WU are structurally coherent, they often terminate without extensive elaboration or justification, suggesting a tendency to conclude discussions quickly at the expense of conceptual depth and collaborative refinement. Comparatively, HPGs, with a sequence of CS → AWS → restating (RE)/ATD → BO → WU, indicate continued cognitive development and discussion iteration post-conclusion.

Figure 7

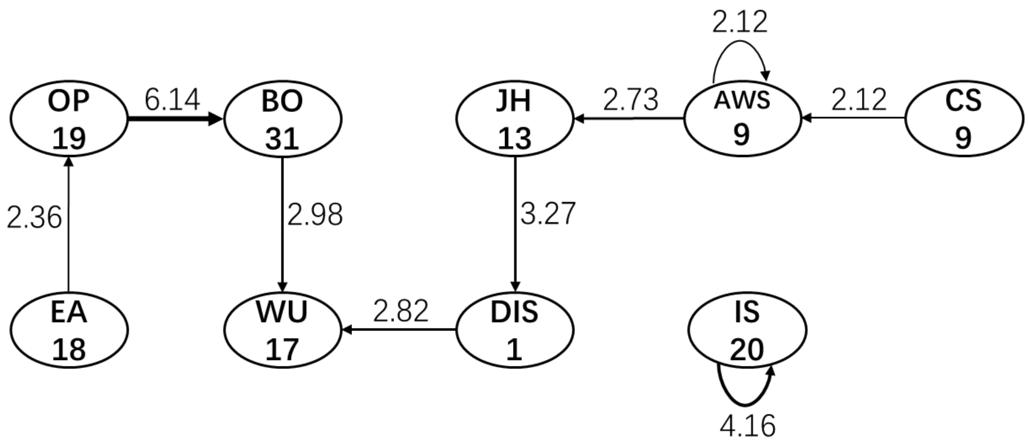
Transition Diagram Showing Changes in Behavior Codes Across HPGs



Note. HPG = high-performing group. Each node represents a behavior and directed arrows denote significant behavioral transitions, with line thickness reflecting the magnitude of the adjusted residual value (*Z*-score). ATD = apply, test, and defend; AWS = agreement; BO = building on; CL = clarification; CS = creating solutions; DIS = divergence; EA = encouraging answer; IS = information sharing; JH = justified hypothesis; LTC = leap to conclusion; OP = opinion; PN = personal narration; RE = restating; SA = supported agreement; SD = supported divergence; TE = thought experiment; WU = wrap-up.

Figure 8

Transition Diagram for LPGs Showing Changes in Behavior Codes Across LPGs



Note. LPG = low-performing group. Each node represents a behavior and directed arrows denote significant behavioral transitions, with line thickness reflecting the magnitude of the adjusted residual value (*Z*-score). ATD = apply, test, and defend; AWS = agreement; BO = building on; CL = clarification; CS = creating solutions; DIS = divergence; EA = encouraging answer; IS = information sharing; JH = justified hypothesis; LTC = leap to conclusion; OP = opinion; PN = personal narration; RE = restating; SA = supported agreement; SD = supported divergence; TE = thought experiment; WU = wrap-up.

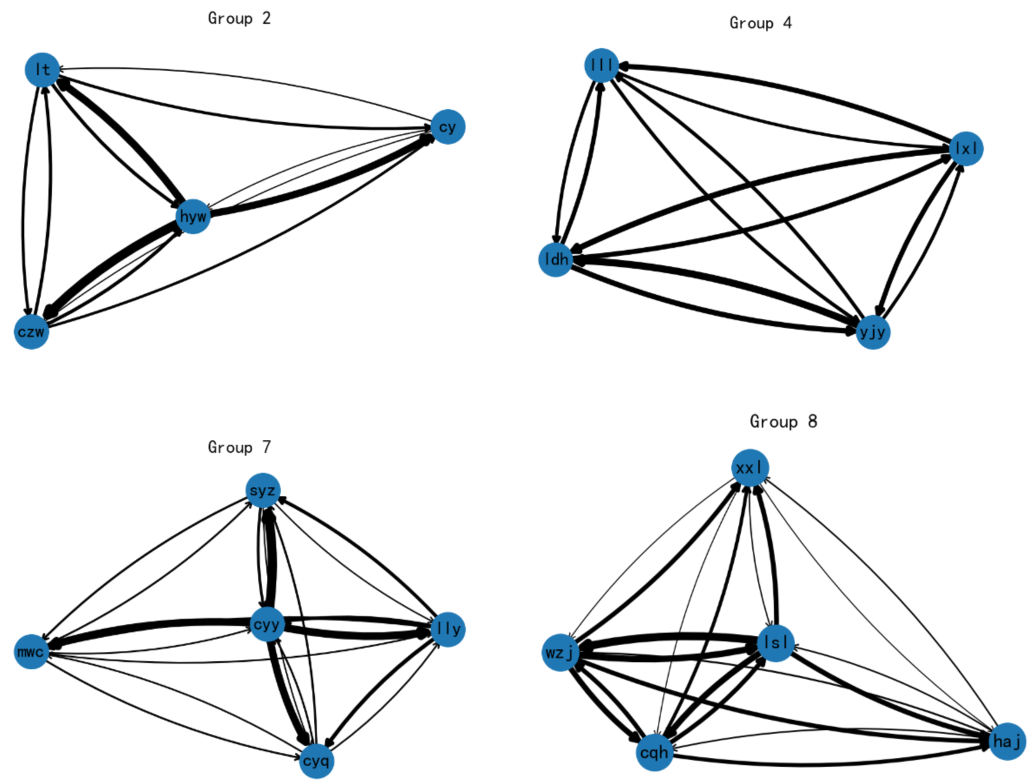
Social Patterns of Cognitive Presence (RQ3)

In figures 9 and 10, which present the interactional structures of the eight groups. While all four HPGs (G2, G4, G7, and G8) demonstrated a fully connected network with each member engaging with all others at least once (network density of 1.0), interaction patterns varied across groups. For example, G2 and G7, where one member (e.g., HYW in group 2) occupies a central position in the network, exhibit pronounced single-centrality structures, indicating a centralized communication flow where coordination was largely driven by one individual. G8, where two or three members were actively engaged in frequent communication, while others remained on the periphery, displays a polycentric structure, suggesting that leadership and participation were more distributed but still unequal. In contrast, G4 shows a more evenly distributed interaction pattern, with relatively uniform communication frequencies among members, suggesting a more collaborative group dynamic. Despite these differences, HPGs are marked by frequent and multidirectional exchanges, reflecting a high degree of interaction intensity and shared cognitive engagement.

For the LPGs, groups 1 and 3 lacked a clear central figure and showed weak connections among members, indicating poor coordination and limited leadership, which may have hindered group effectiveness, confirming the findings of Chen et al. (2022). In contrast, groups 5 and 6 had a central leader, but most other members contributed minimally, reflecting highly uneven participation that likely constrained opportunities for shared knowledge-building (see Liu et al., 2023).

Figure 9

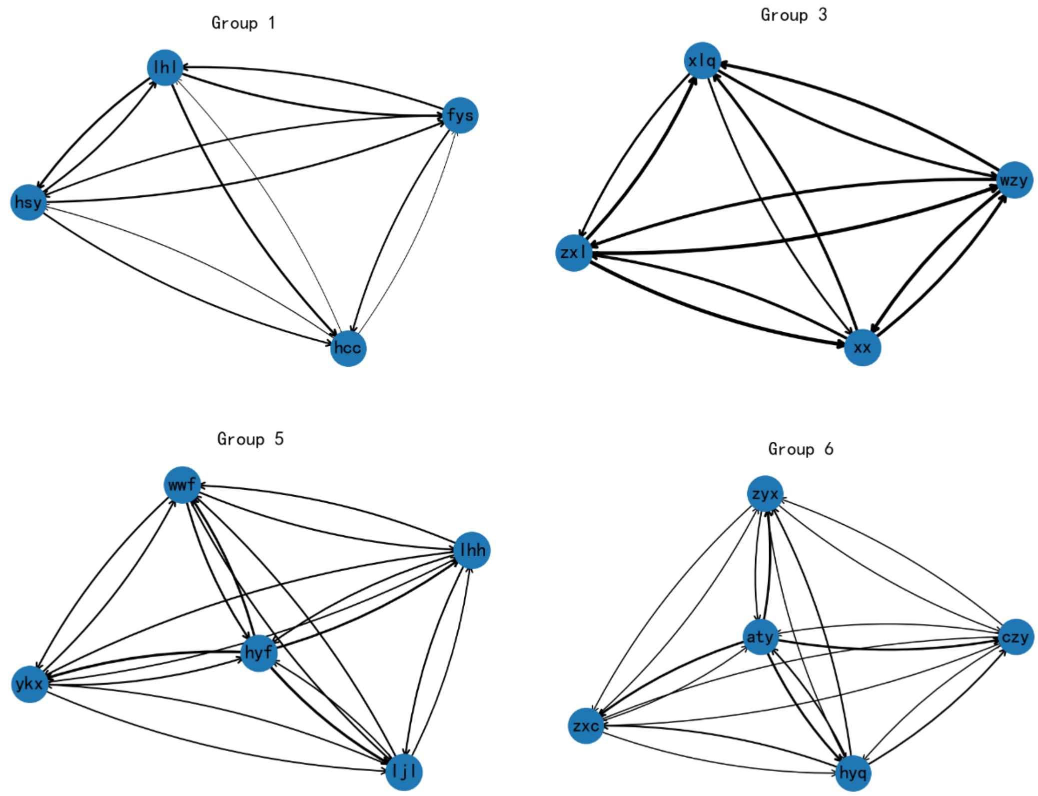
Interactional Structures for HPGs



Note. HPG = high-performing group. The initials on the nodes represent the first letters of each student's name. The thickness of the lines indicates the frequency of interactions between two students, and the arrows show the direction of communication (who is speaking to whom).

Figure 10

Interactional Structures for LPGs



Note. LPG = low-performing group. The initials on the nodes represent the first letters of each student's name. The thickness of the lines indicates the frequency of interactions between two students, and the arrows show the direction of communication (who is speaking to whom).

A MANOVA was conducted to examine differences in centrality indices between HPGs and LPGs (see Table 5). The Box's M test [$\chi^2(10, n = 36) = \text{Inf}$, $p < 2.2\text{e-}16$] and Pillai's Trace indicated a significant overall effect [Pillai's Trace = 0.582, $F(1, 36) = 8.80$, $p < .001$]. Univariate analyses showed that HPGs scored higher than LPGs on degree centrality [$F(1, 36) = 21.4$, $p < .001$] and betweenness centrality [$F(1, 36) = 5.12$, $p < .05$], indicating that HPG members were more actively connected and that some played key bridging roles in linking different parts of the network. These results suggest that successful groups not only had more evenly distributed participation but also stronger coordination roles, facilitating the flow of information across the group (see Liu et al., 2022).

Table 5

Results of Multivariate Analysis of Variance (MANOVA) Examining Differences in Centrality Indices Between HPGs and LPGs

Category of centrality	High-performing		Low-performing		Univariate Test		
	M	*SD*	*M*	*SD*	*F*	η2	Sig.
Degree	20.3	3.54	11.6	1.49	21.4***	.387	0.0000517
Betweenness	0.03	0.026	0	0	5.12*	.131	0.03
Closeness	0.055	0.004	0.105	0.01	1.36	.039	0.251
Eigenvector	0.466	0.017	0.457	0.013	0.205	.006	0.654

Note. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

Discussion

By employing a multi-analytical approach, this study provides a comprehensive, multi-dimensional understanding of cognitive presence in online collaborative knowledge-building, showing not only what learning processes occurred (e.g., discourse structures and cognitive behaviors) but also how sequential, behavioral, and social dynamics shaped effective collaboration.

Differences in Cognitive Presence Between HPGs and LPGs (RQ1)

A key finding of this study is that HPGs and LPGs differ fundamentally in demonstrating cognitive presence, or the extent to which learners construct and confirm meaning through dialogue. HPGs frequently initiated inquiry through questions (triggering), integrated and built upon peers' contributions (exploration and integration), and synthesized discussions into coherent solutions (resolution). In contrast, LPGs rarely initiated questions, remained at surface-level exploration, and seldom reached resolution. These patterns suggest that educators should design prompts, roles, or scaffolds to encourage all groups to generate questions, connect ideas, and apply discussions to problem-solving tasks (Liu et al., 2022; Wise & Hsiao, 2019).

From both behavioral and thematic perspectives, HPGs engaged in more conceptually rich and integrative dialogue, with dense and connected behaviors such as building on, thought experiment, and opinion reflecting movement through all phases of cognitive presence. LPGs, by contrast, relied on motivational prompts to sustain dialogue, with limited engagement in inquiry. Teachers can support LPGs by providing structured starter questions, rotating discussion leaders, or scaffolding to encourage elaboration and justification (Yang et al., 2018).

SPM analyses revealed that HPGs followed recursive discussion loops (e.g., TE → OP → BO → WU), reflecting iterative refinement of ideas (Hong & Choi, 2019), whereas LPGs displayed linear and fragmented patterns. This underscores the importance of requiring reflection checkpoints, peer feedback, or summary–revision cycles to encourage revisiting and refining ideas (Chen et al., 2022).

SNA results showed that HPGs exhibited multidirectional communication and, in some cases, distributed leadership, supporting sustained idea elaboration and synthesis. LPGs displayed fragmented communication and lacked facilitators, making them prone to topic drift, limited focus, and premature closure. The MANOVA confirmed that central actors in HPGs maintained cognitive presence by guiding and integrating discourse (Onrubia et al., 2022). Educators can address this by assigning peer facilitators or strategically stepping in as "more knowledgeable others" (Vygotsky, 1978) to keep discussions focused and collaborative.

In sum, HPGs succeeded by initiating questions, building on each other's ideas, synthesizing discussions, and actively sharing responsibility, while LPGs often stopped at opinion-sharing, lacked clear leadership, and failed to connect ideas meaningfully. Instructional strategies that encourage inquiry, support integration of ideas, and assign balanced roles can help all groups achieve deeper cognitive engagement.

Educational Implications

Based on the findings, this study has five implications for teachers and instructional designers.

First, foster cognitive presence by using open-ended prompts that require students to identify problems or pose questions.

Second, encourage integration and resolution by providing guiding questions that move students from information sharing to connecting and applying ideas.

Third, normalize idea revision by structuring activities that prompt students to revisit, critique, and refine contributions.

Fourth, support participation by assigning rotating group roles (e.g., leader, challenger, summarizer) to prevent marginalization and promote shared responsibility.

Fifth, leverage learning analytics (ENA, SPM, SNA) to detect groups engaged in surface-level discourse and provide timely interventions.

There are four implications for students.

First, engage beyond opinion-sharing by actively building on others' ideas, asking clarifying questions, and co-constructing meaning.

Second, rotate roles to take responsibility for different group functions and develop collaborative and metacognitive skills.

Third, reflect regularly through logs or peer feedback to monitor cognitive progress and refine strategies.

Fourth, value iterative learning by recognizing that disagreement, revision, and rethinking are essential for deep knowledge-building.

Overall, applying these guidelines will enable teachers to design collaborative learning environments that foster higher-order cognitive processes, ensure balanced participation, and support inclusion, while students actively take responsibility for co-constructing knowledge. In addition, integrating ENA, SPM, and SNA would provide practical tools for instructors to identify at-risk learners and implement timely scaffolds to enhance both individual and group outcomes.

Conclusion

This study examined cognitive presence in collaborative knowledge-building using a multi-analytical approach to capture structural, sequential, and social dynamics. HPGs exhibited richer, more interconnected discourse, recursive behavioral patterns, and denser, well-coordinated social networks, whereas LPGs showed surface-level engagement, linear or fragmented sequences, and weaker interaction structures. These findings indicate that strong cognitive presence emerges when learners not only share information but also ask questions, build on ideas, and sustain dialogue through coordinated participation. Practically, teachers should design prompts that trigger inquiry, scaffold the integration and resolution of ideas, and assign rotating roles to ensure equitable participation, while learning analytics can help identify groups at risk of remaining at surface-level discourse and enable timely interventions.

Limitations of this study include a small sample and reliance on coded discussion data, which may not capture affective or motivational aspects of learning. Future research should explore larger and more diverse contexts and integrate multimodal or machine learning methods to capture the full complexity of collaborative learning. In sum, fostering cognitive presence requires both cognitive scaffolds and social supports to help all students, not just high performers, engage in deeper, more meaningful knowledge-building.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 62307010) and the Philosophy and Social Science Planning Project of Guangdong Province of China (No. GD24XJY17).

References

Ba, S., Hu, X., Stein, D., & Liu, Q. (2023). Assessing cognitive presence in online inquiry-based discussion through text classification and epistemic network analysis. *British Journal of Educational Technology*, *54*(1), 247–266. https://doi.org/10.1111/bjet.13285

Baanqud, N. S., Al-Samarraie, H., Alzahrani, A. I., & Alfarraj, O. (2020). Engagement in cloud-supported collaborative learning and student knowledge construction: A modeling study. *International Journal of Educational Technology in Higher Education*, *17*, Article 56. https://doi.org/10.1186/s41239-020-00232-z

Bhutoria, A., & Aljabri, N. (2022). Patterns of cognitive returns to information and communication technology (ICT) use of 15-year-olds: Global evidence from a hierarchical linear modeling approach using PISA 2018. *Computers & Education*, *181*, Article 104447. https://doi.org/10.1016/j.compedu.2022.104447

Chen, X., Xie, H., Qin, S. J., Wang, F. L., & Hou, Y. (2025). Artificial intelligence-supported student engagement research: Text mining and systematic analysis. *European Journal of Education*, *60*(1), Article e70008. https://doi.org/10.1111/ejed.70008

Chen, X., Zou, D., Cheng, G., & Xie, H. (2022). Understanding classroom interaction using epistemic and social network analysis. In R. C. Li, S. K. S. Cheung, P. H. F. Ng, L. P. Wong, & F. L. Wang (Eds.), *Blended learning: Engaging students in the new normal era. ICBL 2022*. Springer. https://doi.org/10.1007/978-3-031-08939-8_14

Cherbow, K., & McNeill, K. L. (2022). Planning for student-driven discussions: A revelatory case of curricular sensemaking for epistemic agency. *Journal of the Learning Sciences*, *31*(3), 408–457. https://doi.org/10.1080/10508406.2021.2024433

Elmoazen, R., Saqr, M., Hirsto, L., & Tedre, M. (2024). Capturing temporal pathways of collaborative roles: A multilayered analytical approach using community of inquiry. *International Journal of Computer-Supported Collaborative Learning*, *20,* 41–77. https://doi.org/10.1007/s11412-024-09431-6

Fougt, S. S., Siebert-Evenstone, A., Eagan, B., Tabatabai, S., & Misfeldt, M. (2018, March). Epistemic network analysis of students' longer written assignments as formative/summative evaluation. In A. Pardo, K. Bartimote-Aufflick, & G. Lynch (Eds.), *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 126–130). ACM. https://doi.org/10.1145/3170358.3170414

Gao, Q., Zhang, S., Cai, Z., Liu, K., Hui, N., & Tong, M. (2022). Understanding student teachers' collaborative problem solving competency: Insights from process data and multidimensional item response theory. *Thinking Skills and Creativity*, *45*, Article 101097. https://doi.org/10.1016/j.tsc.2022.101097

He, Q., Borgonovi, F., & Paccagnella, M. (2021). Leveraging process data to assess adults' problem-solving skills: Using sequence mining to identify behavioral patterns across digital tasks. *Computers & Education*, *166*, Article 104170. https://doi.org/10.1016/j.compedu.2021.104170

Hong, Y.-C., & Choi, I. (2019). Relationship between student designers' reflective thinking and their design performance in bioengineering project: Exploring reflection patterns between high and low performers. *Educational Technology Research and Development*, *67*, 337–360. https://doi.org/10.1007/s11423-018-9618-6

Jiang, J.-P., Hu, J.-Y., Zhang, Y.-B., & Yin, X.-C. (2023). Fostering college students' critical thinking skills through peer assessment in the knowledge building community. *Interactive Learning Environments*, *31*(10), 6480–6496. https://doi.org/10.1080/10494820.2022.2039949

Liu, S., Kang, L., Liu, Z., Zhao, L., Yang, Z., & Su, Z. (2023). Exploring the relationships between students' network characteristics, discussion topics and learning outcomes in a course discussion forum. *Journal of Computing in Higher Education*, *35*(3), 487–520. https://doi.org/10.1007/s12528-022-09335-0

Liu, Z., Zhang, N., Peng, X., Liu, S., Yang, Z., Peng, J., Su, Z., & Chen, J. (2022). Exploring the relationship between social interaction, cognitive processing and learning achievements in a MOOC discussion forum. *Journal of Educational Computing Research*, *60*(1), 132–169. https://doi.org/10.1177/07356331211027300

Maranna, S., Willison, J., Joksimovic, S., Parange, N., & Costabile, M. (2022). Factors that influence cognitive presence: A scoping review. *Australasian Journal of Educational Technology*, *38*(4), 95–111. https://doi.org/10.14742/ajet.7878

Moon, J., McNeill, L., Edmonds, C. T., Banihashem, S. K., & Noroozi, O. (2024). Using learning analytics to explore peer learning patterns in asynchronous gamified environments. *International Journal of Educational Technology in Higher Education*, *21*, 45. https://doi.org/10.1186/s41239-024-00476-z

Morueta, R. T., López, P. M., Gómez, Á. H., & Harris, V. W. (2016). Exploring social and cognitive presences in communities of inquiry to perform higher cognitive tasks. *The Internet and Higher Education*, *31*, 122–131. https://doi.org/10.1016/j.iheduc.2016.07.004

Norz, L.-M., Dornauer, V., Hackl, W. O., & Ammenwerth, E. (2023). Measuring social presence in online-based learning: An exploratory path analysis using log data and social network analysis. *The Internet and Higher Education*, *56*, Article 100894. https://doi.org/10.1016/j.iheduc.2022.100894

Onrubia, J., Roca, B., & Minguela, M. (2022). Assisting teacher collaborative discourse in professional development: An analysis of a facilitator's discourse strategies. *Teaching and Teacher Education*, *113*, Article 103667. https://doi.org/10.1016/j.tate.2022.103667

Sharma, P., Akgun, M., & Li, Q. (2024). Understanding student interaction and cognitive engagement in online discussions using social network and discourse analyses. *Educational Technology Research and Development*, *72*(5), 2631–2654. https://doi.org/10.1007/s11423-023-10261-w

Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the community of inquiry framework for online learning. *Educational Psychologist*, *57*(3), 148–161. https://doi.org/10.1080/00461520.2022.2089989

Song, Y., Cheng, B., Zhu, J., & Hu, X. (2022). Exploring the collective process of classroom dialogue using sequential pattern mining technique. *International Journal of Educational Research*, *115*, Article 102050. https://doi.org/10.1016/j.ijer.2022.102050

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

Wilson, E. C., & Berge, Z. L. (2023). Educational experience and instructional design effectiveness within the Community of Inquiry framework. *The International Review of Research in Open and Distributed Learning*, *24*(1), 159–174. https://doi.org/10.19173/irrodl.v24i1.6751

Wise, A. F., & Hsiao, Y.-T. (2019). Self-regulation in online discussions: Aligning data streams to investigate relationships between speaking, listening, and task conditions. *Computers in Human Behavior*, *96*, 273–284. https://doi.org/10.1016/j.chb.2018.01.034

Wu, B., & Wu, C. (2021). Research on the mechanism of knowledge diffusion in the MOOC learning forum using ERGMs. *Computers & Education*, *173*, Article 104295. https://doi.org/10.1016/j.compedu.2021.104295

Yang, X., Li, J., & Xing, B. (2018). Behavioral patterns of knowledge construction in online cooperative translation activities. *The Internet and Higher Education*, *36*, 13–21. https://doi.org/10.1016/j.iheduc.2017.08.003

Zhang, S., Gao, Q., Sun, M., Cai, Z., Li, H., Tang, Y., & Liu, Q. (2022). Understanding student teachers' collaborative problem solving: Insights from an epistemic network analysis (ENA). *Computers & Education*, *183*, Article 104485. https://doi.org/10.1016/j.compedu.2022.104485

Zou, W., Hu, X., Pan, Z., Li, C., Cai, Y., & Liu, M. (2021). Exploring the relationship between social presence and learners' prestige in MOOC discussion forums using automated content analysis and social network analysis. *Computers in Human Behavior*, *115*, Article 106582. https://doi.org/10.1016/j.chb.2020.106582

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Innovating Interprofessional Continuing Professional Development: Applying the Community of Inquiry Framework to Digital Learning Platforms

Flavio Manganello[1] and Giuseppe Aleo[2]
[1]Institute for Educational Technologies, National Research Council, Genoa, Italy; [2]Faculty of Nursing and Midwifery, Royal College of Surgeons in Ireland, Dublin, Ireland

Abstract

This study investigated how the community of inquiry (CoI) framework can inform digital platform design for interprofessional continuing professional development (ICPD) in healthcare. Using a three-stage comparative methodology, we analyzed technological tools from foundational CoI literature (stage 1), conducted a rapid review of current digital ICPD practices (stage 2), and synthesized findings through matrix-based comparison (stage 3). The analysis of 10 foundational CoI studies and 11 digital interprofessional education studies revealed four distinct adaptation patterns: (a) technological convergence in core communication tools (asynchronous forums, Learning Management System (LMS) platforms); (b) evolutionary divergence in collaborative technologies (video conferencing, real-time document sharing); (c) implementation gaps in reflective and scaffolding tools; and (d) professional context adaptations addressing healthcare-specific needs. While current ICPD practices have demonstrated strong alignment with CoI principles in communication and collaboration tools, significant gaps exist in structured reflection mechanisms, automated feedback systems, and adaptive facilitation features. Critically, systematic CoI framework application in authentic ICPD contexts with practicing professionals has remained largely unexplored, with studies predominantly focused on pre-licensure interprofessional education. Current implementations have used CoI retrospectively as an analytical framework rather than proactively for design guidance. These findings suggest selective rather than comprehensive CoI integration in professional continuing education contexts. The study provided preliminary theoretical guidance for enhancing digital ICPD through CoI-informed design while highlighting the urgent need for empirical validation with practicing healthcare professionals.

Keywords: interprofessional continuing professional development, community of inquiry, distance education, collaborative learning, healthcare professionals

Introduction

The evolving healthcare landscape demands continuous professional development to uphold care quality and respond to emerging clinical challenges. Interprofessional continuing professional development (ICPD) has emerged as a critical strategy for fostering collaboration among healthcare professionals, supporting essential competencies for effective teamwork in complex care environments (Barr, 2009; Meleis, 2016). Through structured learning experiences, ICPD has enabled practicing professionals to develop communication, teamwork, ethical practice, and role understanding skills necessary for high-quality patient care (Dyess et al., 2019; Thigpen et al., 2023).

However, traditional face-to-face ICPD has encountered significant systemic barriers. Research consistently identifies geographical constraints, scheduling conflicts, and resource limitations as primary obstacles to effective interprofessional learning (Rawlinson et al., 2021). These barriers have intensified since the COVID-19 pandemic, with accelerated demands for scalable, effective ICPD solutions that maintain educational quality while accommodating diverse professional schedules. Digital platforms have offered promising solutions through flexible, scalable approaches that can overcome temporal and spatial constraints (MacNeill et al., 2014).

Yet the transition to digital ICPD has presented fundamental design challenges that extended beyond technological implementation. Research has revealed a critical gap between technological capability and pedagogical effectiveness, with platforms often prioritizing technical features over meaningful educational design (Shohel & Kirkwood, 2012). Studies examining digital interprofessional education have consistently highlighted the need for frameworks that systematically support interaction and collaboration (Evans et al., 2018), while current implementations frequently lack structured approaches to foster the collaborative inquiry essential for interprofessional learning. This disconnect between technological potential and educational outcomes suggested that successful digital ICPD requires more than platform functionality; it demands theoretically grounded frameworks that can guide design decisions and ensure pedagogical coherence.

The selection of appropriate educational frameworks for digital ICPD has represented a critical yet underexplored challenge. While multiple theoretical approaches could inform platform design, including the emphasis of connectivism on network-based learning (Siemens, 2013), the focus of activity theory on collaborative work systems (Engeström, 2009), and constructivist approaches emphasizing individual knowledge building (Fosnot, 2013), each framework offered distinct advantages and limitations for interprofessional contexts.

The community of inquiry (CoI) framework (Garrison et al., 2000) has emerged as particularly promising for digital ICPD design for several reasons. Unlike connectivism, which prioritized network connections over structured collaboration, CoI explicitly theorized the interplay among cognitive engagement, social interaction, and instructional facilitation—core requirements for interprofessional learning. Compared to activity theory's complex system focus, CoI has provided an accessible yet theoretically grounded approach with demonstrated effectiveness across diverse educational contexts (Robb & Spadaro, 2022; Zheng et al., 2023). While constructivist approaches have emphasized individual learning, the CoI community-centred model has aligned with the collaborative requirements of interprofessional education.

However, the application of the CoI to professional continuing education has revealed significant theoretical and practical limitations that remain underexplored. Originally developed for higher education contexts with sustained student engagement, the framework assumed participation patterns that may not align with practicing professionals' time constraints and established expertise. The emphasis of the framework on asynchronous discussion, while valuable for reflective inquiry, may inadequately address the real-time collaborative decision-making central to interprofessional practice. Furthermore, the development of the CoI for student populations requires adaptation for professionals who bring established expertise and seek immediate practice applications rather than foundational knowledge building.

These theoretical limitations have manifested empirically in a concerning disconnect between the potential of the CoI and practical ICPD implementation. While studies have demonstrated CoI applicability in synchronous online interprofessional education contexts (Tunningley et al., 2024), systematic analysis of how CoI-recommended tools have enhanced interprofessional collaboration remains limited. Current evidence suggests that digital ICPD platforms have addressed cognitive presence through content delivery and basic social presence through communication tools, yet structured support for collaborative inquiry and adaptive facilitation (core CoI principles) has remained underused (Evans et al., 2018). This implementation gap extends beyond technological considerations to fundamental questions about how collaborative inquiry has unfolded among experienced professionals versus traditional students.

The urgency of addressing this theory-practice gap has intensified with the accelerated adoption of digital health education and growing demands for evidence-based ICPD solutions. Current platforms have demonstrated selective implementation of CoI principles rather than systematic framework integration, suggesting missed opportunities to enhance interprofessional collaboration through theoretically informed design. Understanding how CoI-recommended tools compare with current ICPD practices, and identifying adaptation strategies for professional contexts, represents a critical step toward more effective digital interprofessional learning environments.

Objectives

This study investigated how the CoI framework can inform digital ICPD platform design through a systematic three-stage analysis to identify implementation gaps and develop recommendations for enhancing interprofessional collaboration. Our analytical framework encompassed technological affordances, CoI presence alignment, implementation contexts, and adaptation requirements for practicing professionals.

The research employed a systematic three-stage approach. Stage 1 mapped technological tools from foundational CoI literature to establish theoretical benchmarks for framework-based design. In stage 2 we conducted a rapid review of current digital ICPD practices following the population-concept-context framework (Aromataris et al., 2024), examining healthcare professionals (population), technological tools and collaborative mechanisms (concept), within digital interprofessional education platforms (context). Stage 3 synthesized findings through comparative analysis to generate CoI-informed recommendations that bridge theoretical principles with practical implementation needs.

The study explored three key research questions that built systematically toward actionable recommendations.

1. Which technological tools are recommended in foundational CoI literature to support the three presences?

2. Which technological tools and features are currently employed in digital ICPD platforms, and how do they relate to CoI framework constructs?

3. How do CoI-recommended tools compare with current ICPD practices, and what recommendations emerge for enhancing interprofessional collaboration?

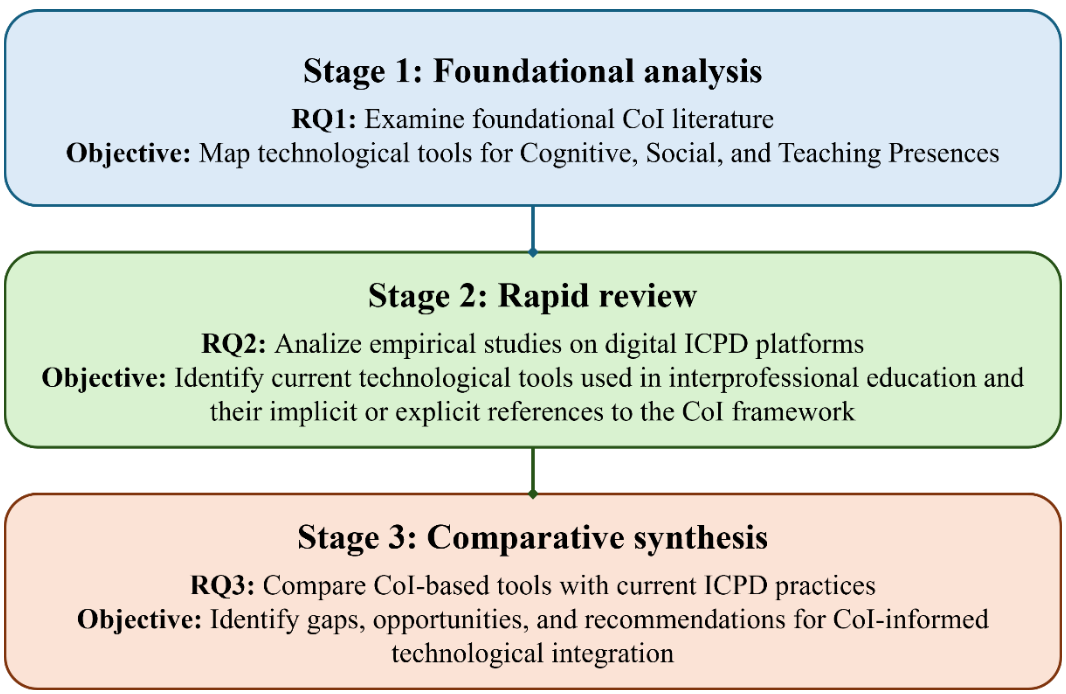
Methods

Protocol and Study Design

This study's three-stage methodology examined how the CoI framework can inform the design of digital ICPD platforms, integrating theoretical analysis with empirical evidence to develop evidence-based recommendations. The overall research process is illustrated in Figure 1, which outlines the three stages and their corresponding research questions and objectives.

Figure 1

Three-Stage Research Process: Methodological Approach, Research Questions, and Objectives



Stage 1 addressed our first research question through systematic content analysis of [foundational CoI research papers by the framework's developers](), mapping technological tools explicitly recommended to support cognitive, social, and teaching presence to establish theoretical benchmarks for CoI-based design.

Stage 2 addressed the second research question through a rapid review following PRISMA for rapid reviews (PRISMA-RR) guidelines (Tricco et al., 2015). This review analyzed empirical studies on digital ICPD platforms, identified technological tools and features supporting interprofessional education, regardless of explicit CoI references, and highlighted common practices and design strategies.

In stage 3, we addressed our third research question through a matrix-based comparative analysis of findings from the first two stages, identifying alignments, gaps, and opportunities between CoI-recommended tools and current ICPD practices.

Inclusion and Exclusion Criteria

In stage 1, we analyzed peer-reviewed journal articles from foundational CoI research papers by the framework's developers. Inclusion criteria required (a) empirical studies with explicit discussion of technological tools for implementing the CoI framework, (b) clear description of tool functions and CoI presence relationships, and (c) implementation in educational contexts. Exclusion criteria eliminated (a) book chapters, conference proceedings, dissertations; and (b) purely theoretical papers without technological focus; (c) studies lacking specific tool descriptions; and (d) secondary analyses or reviews.

Stage 2 included empirical studies (2010–2024) meeting the following criteria: (a) digital platforms involving healthcare professionals from multiple disciplines, (b) continuing professional development or interprofessional education focus, (c) specific technological tool descriptions, (d) reference to collaborative learning mechanisms regardless of explicit CoI mention, and (e) peer-reviewed articles and conference papers in English. Exclusion criteria removed (a) studies focusing solely on pre-licensure Interprofessional Education (IPE) without professional development components; (b) face-to-face learning without digital components; (c) conceptual papers without empirical data; (d) studies lacking detailed technological tool descriptions; and (e) secondary analyses, reviews, or book chapters.

Search Strategy

Stage 1 used foundational CoI research papers from the [official compilation of the developers](), gathered in September 2024. Stage 2 (conducted October 2024) employed a systematic search in Scopus and Web of Science (WoS) using three combined term sets:

- interprofessional education terms ("inter*professional continuing professional development" OR ICPD OR "inter*professional education" OR IPE);
- digital contexts ("electronic learning" OR e*learning OR "virtual learning" OR "technology*enhanced learning" OR "blended learning" OR "web-based learning" OR "distance learning" OR "remote learning" OR Internet OR "educational technology" OR online); and
- specific technological tools and CoI-related terms ("discussion forum*" OR "asynchronous discussion*" OR facilit* OR "computer*mediated communication" OR "computer*mediated conferenc*" OR

"text*based communication" OR synchronous* OR asynchronous* OR video*conf* OR "learning management system*" OR LMS OR "content management system*" OR CMS OR "facilitation tool*" OR "moderation tool*" OR "case-based learning" OR "reflective journal*" OR "peer feedback" OR "group project*" OR "automated feedback" OR "direct instruction tool*" OR "debate format" OR "small group discussion*" OR "role*play" OR "anchored instruction" OR "community of inquiry" OR "teaching presence" OR "social presence" OR "cognitive presence").

Searches were limited to peer-reviewed articles and conference papers in English (2010–2024).

Data Extraction and Analysis

Stage 1

Structured content analysis identified technological tools in foundational CoI literature using a custom extraction form capturing: (a) article metadata (author, year, context); (b) technological tool specifications (name, function, implementation mode); (c) CoI presence associations (cognitive/social/teaching); (d) implementation context (course type, duration, participant numbers); and (e) reported outcomes.

Coding categories included: (a) learning management systems (Blackboard, WebCT); (b) communication tools (forums, chat, e-mail); (c) conferencing systems (synchronous, asynchronous); (d) content tools (wikis, blogs, multimedia); and (e) assessment tools (quizzes, peer review systems). Both authors conducted calibration using three pilot articles before independent coding. Cohen's kappa achieved K = 0.82 for tool identification and K = 0.78 for presence categorization, indicating substantial agreement (Belur et al., 2021). Discrepancies were resolved through discussion, producing structured mapping of distinct technological tools categorized by CoI presence support.

Stage 2

Both authors independently screened titles and abstracts using Rayyan QCRI, achieving 85% initial agreement on study inclusion. Data extraction used standardized forms capturing: (a) study characteristics (design, participants, setting); (b) platform specifications (technology type, features, integration); (c) technological tools (synchronous/asynchronous communication, content delivery, collaboration features); (d) professional disciplines involved; and (e) implementation challenges and outcomes.

Coding categories for ICPD platforms included: (a) core communication (video conferencing, discussion boards); (b) collaboration tools (shared documents, virtual whiteboards); (c) content delivery (LMS, multimedia resources); (d) assessment methods (case studies, peer feedback); and (e) support features (facilitation tools, analytics). Inter-rater reliability achieved 91% agreement for technological tool identification and 87% for CoI construct mapping on calibration subset (n = 5), with disagreements resolved through discussion and consensus.

Quality assessment employed MMAT version 2018 criteria (Hong et al., 2018), evaluating study design appropriateness, sampling adequacy, data collection rigor, and reporting transparency. Individual criteria responses informed evidence synthesis with particular attention to studies that demonstrated stronger methodological rigor.

Stage 3

Matrix-based thematic comparison followed Braun and Clarke's (2022) reflexive thematic analysis. The analytical process involved four systematic steps.

1. Deductive coding. Technological tools from both datasets were coded using predetermined categories (i.e., communication, content, collaboration, assessment, support) derived from CoI framework elements

2. Pattern identification. Systematic comparison across datasets to identify convergent tools (present in both), divergent tools (unique to one dataset), and gaps (recommended but absent).

3. Thematic development. Clustering patterns into themes around technological convergence (shared tools), evolutionary divergence (new vs. traditional tools), implementation gaps (missing CoI-recommended features), and professional context adaptations.

4. Consensus building. Independent review by both authors of thematic classifications with consensus-building discussions for cases requiring refinement.

The comparative matrix systematically mapped CoI-recommended tools against ICPD platform tools, identifying convergent practices, framework-specific recommendations, platform-specific innovations, and implementation gaps where CoI principles remain underused in current ICPD practice.

Results

Selecting Sources of Evidence

In stage 1, we analyzed 35 foundational CoI articles (1999–2015) from the official CoI website, with 10 articles explicitly addressing technological tools included for analysis. Stage 2 identified 527 records from Scopus and WoS, with 11 studies ultimately included after screening and full-text assessment. Table 1 summarizes the selection process for the first two stages.

Table 1

Number of Studies Identified and Selected for Analysis in Each Stage

Selection process	CoI literature	2) ICPD platforms
Source and number of records identified	Foundational studies (*n* =35)	Scopus (*n* = 358) WoS (*n* = 169)
Duplicates removed	–	143
Records screened (title & abstract)	–	384
Full-text articles assessed for eligibility	35	130

Selection process	CoI literature	2) ICPD platforms
Full-text articles excluded	25	119
Final studies included	10	11

Research Question 1: Technological Tools in Foundational CoI Literature

The foundational CoI dataset included 10 articles (1999–2011), revealing a focused but evolving technological landscape. Table 2 provides an overview of the identified technologies and their relation to the three CoI presences.

Table 2

How Different Technologies Support Online Learning in Foundational CoI Studies

Study	Technological tool	Cognitive presence	Social presence	Teaching presence
Rourke et al. (1999)	Computer conferencing (FirstClass, WebCT)	Not analyzed, focus on social presence methodology	Framework for measuring affective/interactive/ cohesive responses through content analysis	Implications discussed; instructor monitoring and facilitation roles
Garrison et al. (2000)	Computer conferencing	Practical inquiry model phases	Supported cognitive presence through interaction; ability to project personal characteristics as real people	Instructional design, facilitation, direct instruction; design of educational experience and content delivery
Garrison et al. (2001)	Computer conferencing	Four-phase practical inquiry operationalization	Collaborative critical thinking support	Guided discourse, facilitated inquiry progression
Anderson et al. (2001)	Computer conferencing	Analysis not focused on cognitive presence	Facilitated student-instructor interaction	Categorized design, facilitation, instruction

Study	Technological tool	Cognitive presence	Social presence	Teaching presence
Akyol & Garrison (2008)	LMS (Blackboard) Synchronous conferencing (Elluminate)	Tracked inquiry phase engagement across time periods	Showed increased group cohesion over time, decreased affective expression	Evaluated instructor guidance effectiveness, student facilitation roles
Garrison & Akyol (2009)	Web 2.0 tools LMS (Blackboard)	Enabled collaborative information discovery and knowledge construction	Fostered interaction through personal blogs, social media, networking sites	Expanded design capabilities, enhances facilitation, supports direct instruction through diverse content forms
Akyol et al. (2009a)	LMS (Blackboard) Synchronous conferencing (Elluminate)	Tracked engagement across inquiry phases; compares online vs. blended integration patterns	Compared affective expression online vs. blended; higher group cohesion in blended format	Facilitation and organization patterns; direct instruction distribution between online/blended
Akyol et al. (2009b)	LMS (Blackboard) Synchronous conferencing (Elluminate)	Integration phase dominant in both contexts; significantly higher in blended course	Higher group cohesion in blended; more affective expression in online	Students assumed facilitation roles; blended course showed higher teaching presence perceptions
Akyol & Garrison (2011)	LMS (Blackboard), Synchronous conferencing (Elluminate)	Integration phase most active in both contexts; strong correlation with perceived learning (r = 0.67 - 0.81)	Distributed student facilitation; higher perceptions in blended course	Student-led facilitation model; teaching presence distributed among participants
Akyol et al. (2011)	LMS (Blackboard) Synchronous conferencing (Elluminate)	Integration phase higher in long-term (47.4%) vs short-term (35.6%); resolution also higher in long-term courses	Group cohesion higher in short-term; affective expression higher in long-term courses	Direct instruction higher in long-term; facilitating discourse higher in short-term courses; higher teaching presence perceptions in short-term

Computer conferencing systems dominated early CoI applications, serving as the primary vehicle for implementing all three presences. Learning management systems emerged as the core platform, with Blackboard predominant for content delivery, discussion, and assessment. Technological evolution was evident across the timeframe. Early studies (1999–2001) relied exclusively on computer conferencing for asynchronous communication. Synchronous conferencing tools (e.g., Elluminate) appeared in later studies (2008–2011), enabling real-time interaction and expanding social presence opportunities. Web 2.0 tools received limited exploration, representing underused potential.

Presence-specific technological patterns emerged clearly. Cognitive presence was consistently supported through structured discussion forums and implementation of the practical inquiry model, with tools facilitating progression through triggering events, exploration, integration, and resolution phases. Social presence relied primarily on communication features that enabled affective expression, open communication, and group cohesion development, with later studies showing distributed student facilitation roles. Teaching presence was operationalized through LMS instructional design capabilities, facilitation tools, and direct instruction features.

Critical findings revealed technological convergence around LMS-synchronous conferencing combinations in later studies, suggesting optimal tool pairing for CoI implementation. The integration phase of cognitive presence emerged as most frequently coded across multiple studies, indicating successful progression beyond initial exploration. Notably, distributed teaching presence through student facilitation became a recurring design pattern, representing a significant evolution from instructor-centered approaches. The timeframe (1999–2011) captured the transition from basic computer conferencing to more sophisticated blended technological environments, establishing foundational patterns for future CoI implementations.

Research Question 2: Technological Tools and CoI Constructs in Digital Interprofessional Learning

The digital ICPD dataset included 11 studies (2011–2023) that represented the next evolutionary phase beyond foundational CoI implementations (Table 3). During this period, there was dramatic technological diversification, with video conferencing platforms emerging as the dominant synchronous communication tool, replacing basic computer conferencing systems. Learning management systems maintained their central role while expanding beyond Blackboard to include OpenEdX and Moodle.

Collaborative platforms emerged as transformative innovations unavailable during the foundational CoI period. Google Docs, Padlet, and Miro enabled real-time collaborative content creation, supporting cognitive and social presence through shared knowledge construction previously impossible with earlier technologies. Multimedia tools gained prominence, with FlipGrid and comic-style animations enhancing engagement beyond text-based interactions.

Professional context adaptations were evident in tool selection and implementation. Case-based discussion platforms (e.g., asynchronous discussion boards focused on patient scenarios) became prevalent, reflecting healthcare education's emphasis on practical application. Synchronous patient case conferences via video conferencing addressed real-time collaborative decision-making needs specific to interprofessional healthcare practice.

Despite technological advancement, structured reflective tools (e.g., e-portfolios, learning journals, reflection prompts) were notably absent across all ICPD studies, representing a significant departure from the emphasis of CoI on individual reflection within collaborative inquiry. Automated feedback mechanisms and peer moderation tools were similarly underused, limiting teaching presence optimization.

Geographic and temporal patterns revealed Australian and Canadian leadership in CoI-informed IPE research, with implementation spanning undergraduate health professions programs exclusively. No studies examined CoI application with practicing healthcare professionals, highlighting the fundamental gap between theoretical framework development and real-world ICPD implementation.

Table 3

Overview of IPE Studies using Digital Platforms

Study	Country	Population	Description of CoI-based implementation	Description of technological implementation	Outcome measures
Billings et al. (2022)	USA	350 health professional educators, learners, and administrative staff	CoI framework for shared experiences, interaction, critical inquiry	Zoom, FlipGrid, Google Docs, Qualtrics	4,000+ ideas generated, 92% satisfaction
Bluteau (2020)	UK	Pre-registration health/social care students	CoI with therapeutic presence emphasis	Moodle, case scenarios, comic animations	Successful therapeutic presence, positive evaluations
Evans et al. (2017)	Australia	14 health profession facilitators	Teaching presence focus (discourse, instruction, design)	Asynchronous discussion boards	64 messages on average: 50.2% discourse, 32.8% instruction, 17.2% design
Evans et al. (2020)	Australia	7 health profession facilitators	Teaching and social presence indicators	Asynchronous discussion boards	17 of 19 CoI indicators used; new feedback on assessment tasks indicator
Evans & Perry (2023)	Australia	118 students, 21 facilitators	CoI comparison: synchronous vs. asynchronous	Discussion boards, Zoom conferences	Higher facilitation strategy ratings in synchronous vs. asynchronous sessions
Hanna et al. (2013)	Canada	7 trained IPE facilitators	CoI for teaching/social presence	Audio-video conferencing, chat	Four themes: technology as dynamic force, reduced non-verbal cues, group process evolution, co-facilitation importance
Hayward et al. (2021)	Canada	Health professional students	CoI for social, cognitive, teaching presence	Discussion boards, video introductions, Zoom, collaborative tools	Four principles: facilitator development, modelling IPC, meaningful content, psychological safety

Study	Country	Population	Description of CoI-based implementation	Description of technological implementation	Outcome measures
Lazinski et al. (2021)	USA	54 PT/PA students	CoI-based HIPE model integration	Asynchronous (Padlet, video) and synchronous (Web-conferencing) phases	Improved understanding of interprofessional practice, teamwork, communication; asynchronous format preferred over synchronous
MacNeill et al. (2014)	Canada	15 multidisciplinary practicing healthcare professionals	CoI framework for cognitive, social, teaching presence in interprofessional collaboration	Asynchronous e-modules and synchronous build-a-case exercise; group vs. individual learning	Group learners: deeper understanding, peer feedback; individual learners: flexibility but lower motivation
Waterston (2011)	Canada	323 students from six healthcare disciplines	CoI for cognitive, social, teaching presence	Mixed-mode: Blackboard discussions, Web-based communication	High participation groups more successful; positive teams showed greater interaction, effective facilitator use, meaningful discourse
Zheng et al. (2023)	Hong Kong	110 undergraduate students from five disciplines	Extended CoI framework incorporating self-regulation and co-regulation as learning presence	OpenEdX platform and Miro for collaborative concept mapping	Co-regulation significantly predicted cognitive presence and self-regulation; both self-regulation and co-regulation had significant positive effects on perceived progress

Research Question 3: Comparing Foundational CoI Tools and Current ICPD Practices

To address our third research question, we used comparative analysis to examine the alignment between technological tools from foundational CoI literature (RQ1) and those used in digital ICPD platforms (RQ2). Table 4 presents a matrix mapping the tools from both datasets, highlighting shared practices, missing elements, and areas for future development to support CoI-informed ICPD platforms.

Table 4

Comparing Technologies Used in CoI Research Versus Current IPE Platforms

Technological tool/feature	Foundational CoI literature	Digital ICPD practices	Multiple affordances	Comparative notes
Asynchronous discussion forums	Extensive use	Widely used	Strong continuity; supported multiple presences	Asynchronous discussion forums
LMS platforms	Core platform	Core platform	Full alignment across contexts	LMS platforms
Synchronous conferencing	Emerging use	Dominant tool	Expanded for real-time collaboration	Synchronous conferencing
Video conferencing	Minimal use	Dominant tool	Professional adaptation; ubiquitous in ICPD	Video conferencing
Collaborative document tools	Minimal use	Common usage	Growth in co-construction activities	Collaborative document tools
Case-based platforms	Not present	Widely used	Professional context innovation	Case-based platforms
Reflective tools	Recommended	Largely absent	Significant implementation gap	Reflective tools
Web 2.0 tools	Limited exploration	Rarely used	Underused potential	Web 2.0 tools
Assessment tools	Recommended	Limited use	Implementation gap in professional contexts	Assessment tools
Multimedia tools	Not present	Emerging use	Innovation in engagement strategies	Multimedia tools

The comparative analysis revealed four distinct patterns of technology-framework alignment across foundational CoI literature and current digital ICPD practices.

The first was technological convergence, which emerged in core tools where both datasets demonstrated strong alignment. Asynchronous discussion forums and LMS platforms represented the stable foundation of CoI-informed learning, maintaining central roles across contexts. These technologies demonstrated functional multiplicity, supporting multiple CoI presences simultaneously. Discussion forums primarily supported cognitive presence through structured inquiry while fostering social presence through community building and teaching presence through guided facilitation.

The second pattern, evolutionary divergence, appeared where current ICPD practices extended beyond foundational CoI recommendations. Synchronous conferencing evolved from an emerging social presence tool in foundational literature to the dominant communication platform in current ICPD practice. Collaborative document tools (e.g., Google Docs, Padlet, Miro) emerged as significant innovations absent from foundational CoI literature, supporting real-time co-construction activities that bridged cognitive and social presence. Video conferencing platforms became ubiquitous in ICPD contexts despite minimal presence in foundational CoI work, reflecting the emphasis of professional education on real-time collaboration and decision-making.

A pattern of implementation gaps revealed critical disconnects between CoI theoretical recommendations and ICPD practice. Reflective tools, strongly recommended in foundational CoI literature for supporting cognitive presence and individual meaning-making, were largely absent from current ICPD implementations. This constituted a significant departure from emphasis of the CoI on personal reflection within collaborative inquiry. Similarly, Web 2.0 tools remained underused across both contexts, suggesting missed opportunities for enhancing social presence through user-generated content and community interaction.

Finally, professional context adaptations highlighted how ICPD practices evolved to meet healthcare professionals' specific needs. Case-based discussion platforms emerged as a professional education innovation, adapting asynchronous forums to focus on patient scenarios and clinical decision-making. Synchronous patient case conferences addressed real-time collaborative needs central to interprofessional healthcare practice, extending beyond traditional educational applications.

Our analysis revealed that current ICPD practices have demonstrated selective implementation of CoI principles rather than comprehensive framework adoption. While communication and collaboration tools aligned well with CoI recommendations, the absence of structured reflection mechanisms suggests incomplete realization of the framework's cognitive presence potential in professional continuing education contexts.

Discussion

This study investigated the technological foundations of the CoI framework and its application in digital interprofessional continuing development, revealing both continuity and divergence between theoretical underpinnings and current practical implementation. The comparative analysis identified four distinct patterns: technological convergence, evolutionary divergence, implementation gaps, and professional context adaptations. These findings contributed to understanding CoI framework adaptation in distributed learning environments, with implications extending beyond health education to broader professional continuing education contexts.

The analysis of foundational CoI literature confirmed a technological ecosystem anchored in computer conferencing systems, learning management systems, and emerging synchronous tools, reflecting the framework's origins in higher education contexts. The technological evolution from 1999 to 2011 demonstrated how distance education frameworks developed technological recommendations that responded to available capabilities while maintaining theoretical coherence. The CoI framework's strength lies in its technological flexibility, providing presence-based criteria that can accommodate technological evolution rather than prescribing specific tools.

Current digital ICPD practices have demonstrated significant adaptation beyond foundational CoI recommendations, reflecting the specific needs of practicing healthcare professionals. The emergence of case-based platforms, synchronous patient conferences, and real-time collaborative tools represents professional education innovations extending traditional applications. These adaptations addressed fundamental differences between student-focused and professional-focused environments, including time constraints, established expertise, and immediate practice applications.

Critically, CoI constructs have been frequently used retrospectively as analytical lenses rather than proactively guiding instructional design. This suggests that while theoretical constructs provided valuable evaluation criteria, their translation into design principles for professional continuing education required substantial adaptation.

Our comparative analysis revealed that technological tools in professional continuing education have demonstrated functional multiplicity, supporting multiple CoI presences simultaneously rather than serving single-purpose functions. Asynchronous discussion forums exemplified this multiplicity, primarily supporting cognitive presence through structured inquiry while fostering social presence through community building and teaching presence through guided facilitation.

Technological convergence emerged in core communication and platform tools, suggesting robust foundational alignment in basic infrastructure requirements for distributed collaborative learning. However, evolutionary divergence appeared in collaborative document tools, video conferencing platforms, and multimedia innovations that became dominant in professional contexts despite minimal presence in foundational CoI work.

Our analysis identified critical implementation gaps between CoI theoretical recommendations and ICPD practice, particularly regarding structured reflective tools and automated feedback mechanisms. The absence of reflective tools represented a significant departure from CoI's emphasis on individual meaning-making within collaborative inquiry, potentially limiting cognitive presence development in professional contexts.

These findings revealed clear opportunities for innovation in ICPD platform design. Integrating structured reflective tools directly into collaborative learning environments could reinforce cognitive presence while accommodating professional learners' preferences for practice-integrated reflection. Intelligent facilitation systems capable of providing adaptive feedback and targeted scaffolding could enhance teaching presence while addressing scalability challenges in professional continuing education.

The selective implementation pattern observed in ICPD contexts—where communication and collaboration tools align with framework recommendations while reflective and scaffolding tools remain underused—may represent a common adaptation strategy when frameworks transition from formal educational to professional development contexts. This study demonstrated how comparative analysis can illuminate both successful adaptations and missed opportunities, providing guidance for framework evolution that serves diverse learning populations in distributed learning environments.

Contribution to Open and Distributed Distance Education (ODDE)

This study has made distinctive contributions to the ODDE field beyond the specific healthcare education context, through advancing understanding of how established pedagogical frameworks adapt when transitioning from formal educational to professional continuing education contexts. While the CoI framework is one of the most extensively cited theories in distance education, our research represented the first systematic analysis of framework evolution across educational contexts, revealing adaptation patterns with broader theoretical and practical implications.

The comparative methodology demonstrated here offers a replicable approach for examining pedagogical framework translation across diverse learning populations. The four adaptation patterns identified (i.e., technological convergence, evolutionary divergence, implementation gaps, professional context adaptations) have provided a conceptual framework for understanding how educational technologies and pedagogical approaches evolve when serving different learner populations. These insights extend beyond healthcare to inform framework application in other professional continuing education domains, including engineering, business, education, and legal professions, where practitioners require ongoing professional development through distributed learning environments.

Critically, this research revealed that framework adaptation involves more than technological substitution. The finding that current implementations use CoI retrospectively as an analytical framework rather than proactively for design guidance illuminates a broader challenge in distance education, namely the gap between theoretical frameworks developed for student populations and their application in professional contexts. This disconnect suggests that successful framework translation requires systematic adaptation rather than direct application, with implications for how the distance education field approaches framework implementation across diverse learning contexts.

The identification of selective implementation patterns, where communication and collaboration tools aligned with framework recommendations while reflective and scaffolding mechanisms remained underused, may represent a common adaptation strategy when frameworks transition between educational contexts. This insight contributes to distance education theory by suggesting that framework evolution follows predictable patterns that can inform proactive adaptation strategies rather than reactive implementation challenges.

Furthermore, this study demonstrated how technological affordances interact with pedagogical frameworks across different learning populations. The emergence of real-time collaborative tools and case-based platforms in professional contexts, despite their absence from foundational framework literature, illustrates how learner characteristics and contextual demands drive technological innovation beyond original framework specifications. This technological-pedagogical evolution pattern likely applies across professional education domains, informing distance education practice in diverse continuing education contexts.

The methodological contribution of systematic comparative analysis between foundational framework literature and current implementation practices has provided a model for evidence-based framework adaptation that addresses both theoretical coherence and practical implementation needs. This approach offers distance education researchers and practitioners a systematic method for understanding framework evolution and optimizing implementation across diverse learning contexts, advancing both distance education theory and practice beyond specific disciplinary boundaries.

Limitations and Future Directions

This study was associated with several critical limitations affecting scope and generalizability that require careful consideration when interpreting findings and recommendations.

Absence of Authentic ICPD Contexts

Most critically, we found no direct evidence of systematic CoI framework application in authentic ICPD contexts involving practicing healthcare professionals. While the stage 2 review identified one study (MacNeill et al., 2014) involving practicing professionals, this study used CoI only as background theoretical foundation rather than systematic design framework, without analyzing CoI presence constructs or implementing CoI-recommended tools. The remaining studies focused exclusively on pre-licensure interprofessional education with student populations, creating a fundamental gap between theoretical analysis and real-world ICPD implementation. This absence represented a significant limitation because practicing professionals differ substantially from students in terms of time constraints, established expertise, immediate practice applications, and learning motivations. Without empirical evidence from systematic CoI implementation in actual continuing education contexts, our recommendations remain theoretical propositions requiring validation rather than evidence-based directives for immediate implementation.

Generalizability Constraints From Student to Professional Populations

The predominant focus on pre-licensure education in current research, with only one study (MacNeill et al., 2014) involving practicing professionals without systematic CoI implementation, limits transferability of findings to practicing professionals. Students and professionals represent fundamentally different learning populations with distinct characteristics: students engage in sustained, structured learning programs while professionals require flexible, practice-integrated development opportunities. Students build foundational knowledge while professionals adapt existing expertise to evolving challenges. This population difference suggests that CoI presence patterns, technological tool effectiveness, and collaborative learning dynamics may function differently in professional contexts, requiring empirical validation before confident generalization.

Temporal and Methodological Limitations

Additional constraints included temporal mismatch between datasets, with foundational CoI literature (1999–2011) reflecting earlier technological capabilities while current ICPD practices (2011–2023) use more advanced tools. The rapid review methodology limited analytical depth compared to full systematic reviews, while geographic concentration in Australia and Canada limited generalizability to other healthcare systems and cultural contexts. These methodological constraints affected the comprehensiveness of evidence synthesis and cross-cultural applicability of recommendations.

Preliminary Nature of Recommendations

Given these limitations, our recommendations should be viewed as preliminary theoretical guidance requiring empirical validation rather than evidence-based directives for immediate implementation. The lack of authentic ICPD data means that proposed CoI adaptations remain largely hypothetical until tested with practicing professionals in real-world continuing education contexts. Implementation recommendations necessarily rely on theoretical extrapolation from student-focused research rather than direct professional education evidence.

Implications for Framework Adaptation Research

These limitations collectively suggest that our study represents an initial step toward understanding framework adaptation across educational contexts rather than definitive guidance for professional education design. The findings illuminated both adaptation patterns and critical research gaps, highlighting the need for empirical investigation of how established pedagogical frameworks function in professional continuing education environments.

Despite these constraints, this study has made distinctive contributions to the ODDE field by providing the first systematic analysis of how established pedagogical frameworks transition from higher education to professional continuing education contexts. The comparative methodology demonstrated here can inform future research examining framework adaptation across educational contexts, while the four adaptation patterns identified also provide a conceptual framework for understanding educational technology evolution beyond specific disciplinary boundaries.

Future Research Priorities

Empirical testing of CoI-informed designs with practicing healthcare professionals in authentic ICPD settings represents the most critical research need. Longitudinal studies examining CoI presence patterns between student and professional populations could validate theoretical assumptions underlying framework adaptation and inform design principles specific to professional continuing education contexts. Cross-cultural validation studies and investigation of framework adaptation principles applicable beyond healthcare contexts represent important directions for advancing distance education theory and practice.

The need for design-based research approaches that systematically develop and test CoI-informed technologies while documenting adaptation strategies could contribute both to practical design guidelines and theoretical insights about framework evolution, advancing both professional education practice and distance education theory in distributed learning environments.

Conclusion

This study revealed a partial integration of CoI principles within digital ICPD practices through four distinct adaptation patterns: (a) technological convergence in core communication tools, (b) evolutionary divergence in collaborative technologies, (c) implementation gaps in reflective mechanisms, and (d) professional context adaptations addressing healthcare-specific needs. Current implementations have demonstrated selective rather than comprehensive framework adoption, with significant opportunities for innovation through structured reflective tools, intelligent facilitation systems, and co-regulation mechanisms.

For the broader distance education field, this research has demonstrated how comparative analysis can illuminate framework evolution across learning contexts, revealing that successful adaptation requires systematic modification rather than direct application when transitioning from student to professional populations. The methodological approach and adaptation patterns identified provide a replicable framework for understanding pedagogical theory translation across diverse continuing education domains.

Future developments must prioritize empirical validation with practicing professionals in authentic ICPD settings to transform these theoretical insights into evidence-based guidance for professional continuing education in distributed learning environments.

References

Akyol, Z., & Garrison, D. R. (2008). The development of a community of inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. *Journal of Asynchronous Learning Networks*, *12*, 3–22. https://doi.org/10.24059/olj.v12i3-4.1680

Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250. https://doi.org/10.1111/j.1467-8535.2009.01029.x

Akyol, Z., Garrison, D. R., & Ozden, M. Y. (2009a). Development of a community of inquiry in online and blended learning contexts. *Procedia-Social and Behavioral Sciences*, *1*(1), 1834–1838. https://doi.org/10.1016/j.sbspro.2009.01.324

Akyol, Z., Garrison, D. R., & Ozden, M. Y. (2009b). Online and blended communities of inquiry: Exploring the developmental and perceptional differences. *The International Review of Research in Open and Distributed Learning*, *10*(6), 65–83. https://doi.org/10.19173/irrodl.v10i6.765

Akyol, Z., Vaughan, N., & Garrison, D. R. (2011). The impact of course duration on the development of a community of inquiry. *Interactive Learning Environments*, *19*(3), 231–246. https://doi.org/10.1080/10494820902809147

Anderson, T., Rouke, L., Garrison, D. R., & Archer, W. (2001). Assessing teaching presence in a conferencing context. *Journal of Asynchronous Learning Networks, 5*(2), 1–17. https://doi.org/10.24059/olj.v5i2.1875

Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (Eds.). (2024). *JBI manual for evidence synthesis*. JBI. https://synthesismanual.jbi.global

Barr, H. (2009). An anatomy of continuing interprofessional education. *Journal of Continuing Education in the Health Professions*, *29*(3), 147–150. https://doi.org/10.1002/chp.20027

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Billings, H., Malin, T., Allen, J., Bergene, A., Cornelius, K., Craft, S., Hartzheim, L., Johnson, S. A., Kallay, J., Kreuter, J., Raths, M., Renner, A. S., & Hall, E. (2022). Reimagining learning spaces of the future: An interprofessional, virtual workshop utilizing rapid idea generation and lean startup methodologies. *MedEdPORTAL*, *18*, 11217. https://doi.org/10.15766/mep_2374-8265.11217

Bluteau, P. (2020). The good enough facilitator: Exploring online interprofessional therapeutic facilitation in times of COVID-19. *Journal of Interprofessional Care*, *34*(5), 647–654. https://doi.org/10.1080/13561820.2020.1807919

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage.

Dyess, A. L., Brown, J. S., Brown, N. D., Flautt, K. M., & Barnes, L. J. (2019). Impact of interprofessional education on students of the health professions: A systematic review. *Journal of Educational Evaluation for Health Professions*, *16*. https://doi.org/10.3352%2Fjeehp.2019.16.33

Engeström, Y. (2009). The future of activity theory: A rough draft. In A. Sannino, H. Daniels, & K. Gutierrez (Eds.), *Learning and expanding with activity theory* (pp. 303–328). Cambridge University Press.

Evans, S., Knight, T., Walker, A., & Sutherland-Smith, W. (2020). Facilitators' teaching and social presence in online asynchronous interprofessional education discussion. *Journal of Interprofessional Care*, *34*(4), 435–443. https://doi.org/10.1080/13561820.2019.1622517

Evans, S., & Perry, E. (2023). An exploration of perceptions of online asynchronous and synchronous interprofessional education facilitation strategies. *Journal of Interprofessional Care*, *37*(6), 1010–1017. https://doi.org/10.1080/13561820.2023.2213718

Evans, S. M., Ward, C., & Reeves, S. (2017). An exploration of teaching presence in online interprofessional education facilitation. *Medical Teacher*, *39*(7), 773–779. https://doi.org/10.1080/0142159X.2017.1297531

Evans, S., Ward, C., & Reeves, S. (2018). Online interprofessional education facilitation: A scoping review. *Medical Teacher, 40*(7), 762–771. https://doi.org/10.1080/0142159X.2018.1460656

Fosnot, C. T. (2013). *Constructivism: Theory, perspectives, and practice* (2nd ed.). Teachers College Press.

Garrison, D. R., & Akyol, Z. (2009). Role of instructional technology in the transformation of higher education. *Journal of Computing in Higher Education*, *21*, 19–30. https://doi.org/10.1007/s12528-009-9014-7

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. https://doi.org/10.1080/08923640109527071

Hanna, E., Soren, B., Telner, D., MacNeill, H., Lowe, M., & Reeves, S. (2013). Flying blind: The experience of online interprofessional facilitation. *Journal of Interprofessional Care*, *27*(4), 298–304. https://doi.org/10.3109/13561820.2012.723071

Hayward, K., Brown, M., Pendergast, N., Nicholson, M., Newell, J., Fancy, T., & Cameron, H. (2021). IPE via online education: Pedagogical pathways spanning the distance. *Journal of Interprofessional Education & Practice*, *24*, 100447. https://doi.org/10.1016/j.xjep.2021.100447

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Lazinski, M. J., Ross, L., Wolf, S., Finck, M., & Cherry, L. (2021). Interprofessional education at a distance: The hybrid interprofessional education model. *Internet Journal of Allied Health Sciences and Practice*, *19*(3), 15. https://doi.org/10.46743/1540-580X/2021.2011

MacNeill, H., Telner, D., Sparaggis-Agaliotis, A., & Hanna, E. (2014). All for one and one for all: Understanding health professionals' experience in individual versus collaborative online learning. *Journal of Continuing Education in the Health Professions*, *34*(2), 102–111. https://doi.org/10.1002/chp.21226

Meleis, A. I. (2016). Interprofessional education: A summary of reports and barriers to recommendations. *Journal of Nursing Scholarship*, *48*(1), 106–112. https://doi.org/10.1111/jnu.12184

Rawlinson, C., Carron, T., Cohidon, C., Arditi, C., Hong, Q. N., Pluye, P., Peytremann-Bridevaux, I., & Gilles, I. (2021). An overview of reviews on interprofessional collaboration in primary care: Barriers and facilitators. *International Journal of Integrated Care, 21*(2), 32. https://doi.org/10.5334/ijic.5589

Robb, M., & Spadaro, K. C. (2022). Exploration of online Doctor of Nursing practice students' perceptions of effective teaching methods using the critical incident technique. *Nurse Educator*, *47*(6), 328–331. https://doi.org/10.1097/NNE.0000000000001217

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *The Journal of Distance Education/Revue de l'ducation Distance, 14*(2), 50–71. Retrieved September 3, 2024 from https://www.learntechlib.org/p/92000/

Shohel, M. M. C., & Kirkwood, A. (2012). Using technology for enhancing teaching and learning in Bangladesh: Challenges and consequences. *Learning, Media and Technology*, *37*(4), 414–428. https://doi.org/10.1080/17439884.2012.671177

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Thigpen, J., Naccarini, D., Moore, S. L., & Culhane, N. S. (2023). Incorporating interprofessional education into a required introductory pharmacy and nursing didactic course. *American Journal of Pharmaceutical Education*, *87*(9), 100123. https://doi.org/10.1016/j.ajpe.2023.100123

Tricco, A. C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., Perrier, L., Hutton, B., Moher, D., & Straus, S. E. (2015). A scoping review of rapid review methods. *BMC Medicine*, *13*, 1–15. https://doi.org/10.1186/s12916-015-0465-6

Tunningley, J. M., Zucchero, R. A., & Hooker, E. A. (2024). Comparing quantitative outcomes of synchronous online versus in-person interprofessional symposium. *Journal of Interprofessional Care*, *38*(1), 113–120. https://doi.org/10.1080/13561820.2023.2241506

Waterston, R. (2011). Interaction in online interprofessional education case discussions. *Journal of Interprofessional Care*, *25*(4), 272–279. https://doi.org/10.3109/13561820.2011.566647

Zheng, B., Ganotice, F. A., Lin, C. H., & Tipoe, G. L. (2023). From self-regulation to co-regulation: Refining learning presence in a community of inquiry in interprofessional education. *Medical Education Online*, *28*(1), 2217549. https://doi.org/10.1080/10872981.2023.2217549

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Usability Testing for an Open Educational Resource to Teach Language and Culture

Heejin Chang[1] and Scott Windeatt[2]
[1]University of Southern Queensland, Australia; [2]Newcastle University, UK

Abstract

This study focused on procedures for creating, testing, and developing a set of reusable online resources for use in English for academic purposes programmes. The aim of the materials was to help migrants and refugees develop the linguistic and cultural skills, knowledge, and understanding they would need to engage, interact, and collaborate effectively in a multicultural context. Development of the materials involved an iterative process using a three-stage approach:

1. Expert review: Experts in relevant fields worked through the first version of the materials and provided critical feedback, which guided initial revisions.

2. Usability testing groups: Small groups of target users (students and teachers) used the revised materials in workshop settings, and data were gathered from observations, interviews, and written comments.

3. Wider evaluation: Larger-scale use and evaluation of the materials (which is ongoing, beyond the scope of this paper).

This article reports on the second stage.

Keywords: open educational resource, OER, usability testing, English for academic purposes, EAP, language and culture

Introduction

Over the past two decades, the integration of technology into language education has accelerated, supported by the increasing availability of digital open educational resources (OER) created and shared by the global academic community. Engagement with OER in language learning courses can facilitate the acquisition of language and content knowledge by providing information, and instructional and elicitation activities (Burrows et al., 2022). Equally important, however, it can also promote a broad range of learning practices by encouraging learners to experience and explore language through exposure to resources on the Internet and by fostering collaboration among learners (Brook, 2011; Olivier, 2019).

This is the philosophy which underlies the design of an OER book titled *Communication Across Cultures* (Chang et al., 2025), which has the aim of tackling linguistic, cross-cultural, and digital literacy issues, initially for refugees and asylum seekers participating in English for academic purposes programmes in an Australian university. While the resource provides opportunities to practise and develop relevant language skills, its primary concern lies in supporting cross-cultural communication. This is achieved through activities that prompt reflection on users' own cultural identities and their perceptions of others, thereby cultivating intercultural awareness and competence (Liddicoat & Scarino, 2013). Recognising that levels of digital literacy among users will vary, opportunities for practising, acquiring, and developing aspects of digital literacy are also embedded in the OER design (Getenet et al., 2024). Intercultural understanding and digital literacy are central to equipping learners for meaningful participation in increasingly globalised and digitally mediated environments.

However, the processes involved in the creation, publication, re-use, and discovery of such an OER nevertheless require a significant engagement with technology and an understanding of how to make use of the digital environment for educational purposes (Borthwick & Gallagher-Brett, 2014; Olivier, 2019). The development of effective OER involves an iterative process of design, creation, assessment, and revision—activities that require continuous feedback from both students and instructors (Tomlinson, 2012). Despite this, there are relatively few studies that report on such analysis at the earlier stages of material creation, such as techniques to identify the effect of presentation, instructions, content, and technology on the usability of materials.

This study addressed this gap by investigating the early design stages of an OER (the textbook *Communication Across Cultures*). In doing so, we incorporated usability testing to assess how learners interacted with the materials and how these interactions informed iterative revisions. Such analysis and evaluation at these preliminary stages is crucial to identify and correct issues which might impose an unnecessary cognitive load on the learners (Sanders & Lafferty, 2010).

Literature Review

Open Educational Resources (OER)

Open educational resources (OER) are defined as "learning, teaching, and research materials in any format and medium that reside in the public domain or are under copyright that have been released under an open license, that permit no-cost access, reuse, repurpose, adaptation, and redistribution by others" (UNESCO, 2019, p. 5). They are characterised as *big OER* when referring to whole courses of connected materials and *little OER* when referring to individual shared items such as handouts, images, or presentations (Weller, 2010). Once these materials are made public, OER may be revised, substantially altered, or used for a purpose or by a target audience in a way that may vary to a greater or lesser degree from what was originally intended. This adaptability is central to the concept of OER-enabled pedagogy, which Wiley and Hilton (2018) defined as teaching and learning practices that are only possible or practical in the context of the 5R permissions (retain, reuse, revise, remix, and redistribute). Within this framework, even minor design enhancements—such as interface adjustments—can play a critical role. For instance, a small addition to an OER platform can balance the needs of diverse educational contexts, improving initial usability without overwhelming users who may not require advanced features, as suggested by Sanders and Lafferty (2010). Such design considerations align with the ethos of OER-enabled pedagogy by supporting flexible, learner-centred experiences.

Usability Testing

Usability testing is a means of evaluating how well design principles have been implemented in practice in creating materials. It involves observation of those materials in use to identify how the intended users react, how they carry out the intended tasks, whether they encounter any problems, and what might be the likely cause of any difficulties (Barnum, 2020). It is considered a critical stage in producing and revising materials and is usually repeated throughout the development cycle. For example, Doubleday et al. (2011) conducted usability tests on an online resource for teaching anatomy, implemented changes based on the feedback, and then tested again before releasing the materials for general use. Sugar (1999), although his work is dated, emphasised that designers should attribute any user difficulties to flaws in the design and should pay attention to even atypical user behaviours since these reflect individual differences that need to be accommodated. After a testing session, designers are advised to avoid rushing to implement the first obvious fix; instead, they should determine the underlying cause of each problem and consider multiple possible solutions, reflecting and consulting with others before deciding on changes. Barnum (2020) highlighted the importance of considering specified users, goals, and context when evaluating usability.

In the realm of language and culture learning and teaching, these considerations were applied in this study to determine whether users could effectively use the book, whether the material's goals (language and intercultural communication) were met without overwhelming or confusing learners, and whether the context of use (classroom vs. independent study) could affect how much guidance the material needs to provide. This latter point was vital: a resource used autonomously should have more self-explanatory instructions and supports, whereas in a teacher-mediated classroom, some details can be handled by the instructor.

A significant number of learners today are proficient with digital technologies, especially smartphones, which many use as their primary computing devices. This familiarity influences their expectations and how they interact with educational content. However, while many students may be tech savvy, the technological backgrounds of students can vary widely, particularly among migrants and refugees, who may not have the same level of access or familiarity with digital technologies. This diversity requires that educational technologies be versatile and accessible in order to cater to a broad range of learners. Kessler and Plakans's

(2001) rationale for student input—recognising learners as stakeholders with individual needs who are influenced by their environment—also underpinned our inclusion of both student and teacher feedback in the design process. By involving these stakeholders in usability testing, we gathered insights on how real users interacted with the OER and how it could be improved to support their learning more effectively. This study also included expert input to further enhance the quality and applicability of the material. It ensured that materials were user-centred and adaptable, which is particularly important for open resources intended for diverse audiences. Guided by these insights, this study posed the following research questions.

R1. How easy do students find it to navigate through the materials?

R2. What features do they find most useful?

R3. What difficulties, if any, do they encounter?

R4. What improvements might they suggest?

Methodology

Background

This study forms part of a systematic process for designing and developing open-access materials in the form of an online open textbook, a format which has been adopted in order to exploit the range of affordances the medium offers beyond that of a printed textbook. The initial development involved a collaborative, iterative approach: after the first draft of the materials was created, it underwent expert evaluation, followed by user testing with target learners, prior to a large-scale rollout. This study examined the application of a usability testing framework to the second stage, involving two usability testing groups to identify issues related to the format and content that needed addressing before wider release, using a co-discovery approach (see Procedure).

This stage in the study was concerned with the application of a usability testing framework to investigate issues that might need addressing before releasing the materials to a wider audience. Following a critical evaluation of the materials by expert educators, users representative of the target audience worked through a sample of the amended materials, and their reactions were recorded by means of observation, interviews, and written comments, following which possible improvements to the materials were identified.

Materials

The materials were designed and developed to be used as the intercultural communication component in English for academic purposes and other language courses. They were hosted on the open Pressbook platform (https://pressbooks.com), which provides features not commonly found in traditional printed textbooks used for teaching and learning English as a foreign or second language.

The content was composed of three modules, and the aim was for the learners to enhance their knowledge and skills in three areas: language proficiency, cultural knowledge, and digital literacy. Each module had 7 or 8 video, audio, or reading tasks which involved watching, listening, and reading, and the production of written or recorded spoken responses. Sixty-eight activities were designed to facilitate interactivity through collaborative learning experiences where students interacted with peers to complete tasks, solve problems, and communicate in the target language.

To mirror real-world language use, contexts, and interactions, the materials included tasks that involved learners cooperating to produce both spoken and written answers, using their individual experiences. They incorporated videos, articles, and images to create personalised multimodal learning experiences.

The development of the materials in the book progressed through three stages. The first stage was characterised by an iterative process of drafting, revising, and refining to ensure the content's quality, pedagogical soundness, and alignment with the project's goals. This phase involved close collaboration over the course of a year with a multidisciplinary team that included two researchers in computer-assisted language learning (CALL), one specialist in cross-cultural communication, and two librarians with expertise in OER. In the initial four months, the team held regular meetings to collaboratively shape the content and design of the materials. Then, building on the feedback and structure established in the first module, the same process was applied to the development of the remaining two modules—each undergoing its own cycle of drafting, circulation, and revision to maintain consistency and quality across the full set of materials.

The second stage involved usability testing with small groups of target users, which included two end-user groups: students and teachers. Feedback was collected through two group test sessions with six students as the target audience and two language teachers. Further details on these sessions can be found in the Procedure section. This study focused on the second stage that occurred prior to a broader assessment (stage three).

Steps in Usability Testing

Participants

The materials were developed for students and English as a second language (ESL) and English as a foreign language (EFL) teachers engaged in English for academic purposes (EAP) programmes.

Student Participants

In usability testing, practical considerations suggest that involving around five to six participants provides the optimal balance, resulting in meaningful insights without encountering diminishing marginal returns (Nielsen & Landauer, 1993). For this study, recruitment was announced in the classroom, and six students—forming two groups of three—volunteered to take part in the usability testing process. These six were EFL students at an Australian university. They were from a variety of first language (L1) backgrounds (Congo, Iraq, Bhutan, Japan, and Columbia), studying for 10 hours per week on a 10-week academic skills course as a part of an EAP programme. Four of these students were refugees, and one was a migrant, all with interrupted prior education, while one was an international student. Their level of spoken English ranged from pre-intermediate to intermediate level (CEFR B1 to B2, according to the Common European Framework of Reference for languages). By involving a mix of participants, we ensured that feedback was obtained from those who would truly benefit from improvements—including learners who might struggle with digital literacy or cultural content due to their backgrounds.

Teacher Participants

To gather feedback from educators experienced in teaching ESL students, two teachers were contacted and invited to review the materials independently at their convenience. Both were experienced in teaching EAP or ESL to international and refugee students. One teacher had taught exclusively in Australia, while the other had taught in multiple countries. They had more than 10 years of experience each in teaching English to ESL/EFL students and were familiar with using technology in their teaching. They were not given specific evaluation criteria, to elicit honest, holistic impressions similar to how an instructor might appraise a new resource when considering it for adoption. Their feedback would help identify any issues from an instructor's viewpoint (e.g., suitability of content, clarity of instructions for classroom use, and potential challenges for lower-level students) and suggest ways to make the material more teacher-friendly.

Procedure

The usability testing procedure was adapted from a number of approaches (Barnum, 2020; Dumas & Redish, 1999; Kessler & Plakans, 2001, p. 17; Rubin & Chisnell, 2008, p. 306) and employed a variety of methods to gather data on user experience:

- think aloud protocol
- co-discovery (group interaction)
- observation
- self-reporting
- interview
- teacher review

All student activities took place in a 1-hour workshop session. Students were briefed at the start to use the materials as if they were learning from it, while thinking aloud and discussing with their group. Each group of three students shared a computer, which encouraged collaboration (a co-discovery approach where they could help each other and react together). We circulated at a distance, observing but not interfering, except to note any obvious sticking points (e.g., trouble finding a section). This setup allowed us to capture spontaneous-use behaviours and peer discussions about the material. Among the group members, roles such as note-taker, materials navigator, and discussion leader were assigned and emerged organically rather than through explicit instruction. This approach aligns with Barnum's (2020) suggestion to avoid dominance by any one participant and to encourage balanced participation within the group.

After approximately 40 minutes of exploration and a group discussion, students spent 10–15 minutes writing individual comments (guided by a couple of open-ended questions about what they liked, disliked, found easy or hard, and suggestions for improvement). Immediately afterward, we conducted brief (~10 minute) interviews with the two students who opted to speak rather than write, covering the same questions verbally.

The two teachers conducted their review, and they each went through the materials at their convenience. Table 1 provides an overview of the data sources and which methods each participant group contributed.

Table 1

Data Sources by Participant Type in Stage Two of the Study (Small Target Users)

Participants	Method	
	Think-aloud protocol, co-discovery, and self-reporting log	Observation
6 EAP students (in 2 groups of 3)	Written feedback (4) Follow-up interview (2) Group feedback (2)	Observation notes (2)
2 ESL teachers	Written feedback (2)	Not applicable

Note. EAP = English for academic purposes; ESL = English as a second language; Numbers in parentheses = the number of responses provided by participants.

Data Analysis

The data were analysed using a usability matrix framework, categorising emerging themes under three foci: design, navigation, and content (Kessler & Plakans, 2001, p. 18) to investigate how the materials were viewed by first-time participants as potential users. The elicited themes were allocated into three core categories, which were further developed into sub-categories as shown in Table 2.

Table 2

Summarising the Three Core Categories and Sub-Categories

Core category	Sub-categories
Design	Visual appeal and layout consistency Interactive elements and multimedia integration
Navigation	Ease of use/user-friendly interface Functionality (search, bookmark, note-taking, progress tracking) Links Cross-device compatibility
Content	Educational value and relevance (including suitability for autonomous vs. classroom use) Interaction Cultural sensitivity and inclusivity Task instruction

All data were labelled with these codes. For example, a comment such as “the text is too dense in sections” was tagged under design—visual layout, whereas “I wasn’t sure where to click first” was tagged under navigation—ease of use. Some feedback touched multiple areas and was coded accordingly. We iteratively refined the coding scheme: initial codes were drawn from the research questions and expected issues (e.g., navigation difficulty, corresponding to question R1), but additional themes were added as needed (e.g., several comments mentioned notetaking, which we grouped under navigation—functionality).

To enhance reliability, the analysis incorporated triangulation across data sources. We compared what was observed (behaviour evidence) with what the participants said. For instance, if observation notes indicated a student struggled to find a module, we checked if the student or others mentioned navigation in their comments. In most cases, there was alignment, but triangulating helped verify that certain issues were genuine (noticed both by users themselves and by observer). The diverse sources (written, oral, and observed) allowed cross-validation of key findings. While we did not employ an independent coder, we discussed the coding framework to ensure the categories made sense and captured the data without major biases.

The analysis prioritised identifying recurring themes (mentioned by multiple participants) as well as notable unique feedback. Given the modest sample size, we were cautious not to overgeneralise, but if both groups of students and a teacher all pointed to the same issue (e.g., need for more instructions), we considered that a strong indication that design change was needed. Conversely, if one participant had a singular suggestion (e.g., integrating a particular tool), we noted it but weighed it in context (perhaps as a lower priority or a future enhancement).

Finally, to address the research questions explicitly, we mapped the themes and sub-categories back onto the individual questions. R1 (ease of navigation) corresponded mostly to the navigation category results; R2 (useful features) corresponded to positive findings in design and content (what they liked, e.g., multimedia); R3 (difficulties) corresponded to any negative issues across all categories; and R4 (suggested improvements) was drawn from the suggestions participants made, often directly tied to the difficulties.

The results are organised below by the three core categories (design, navigation, and content) for coherence. Direct quotes from participants are provided to illustrate each theme.

Results

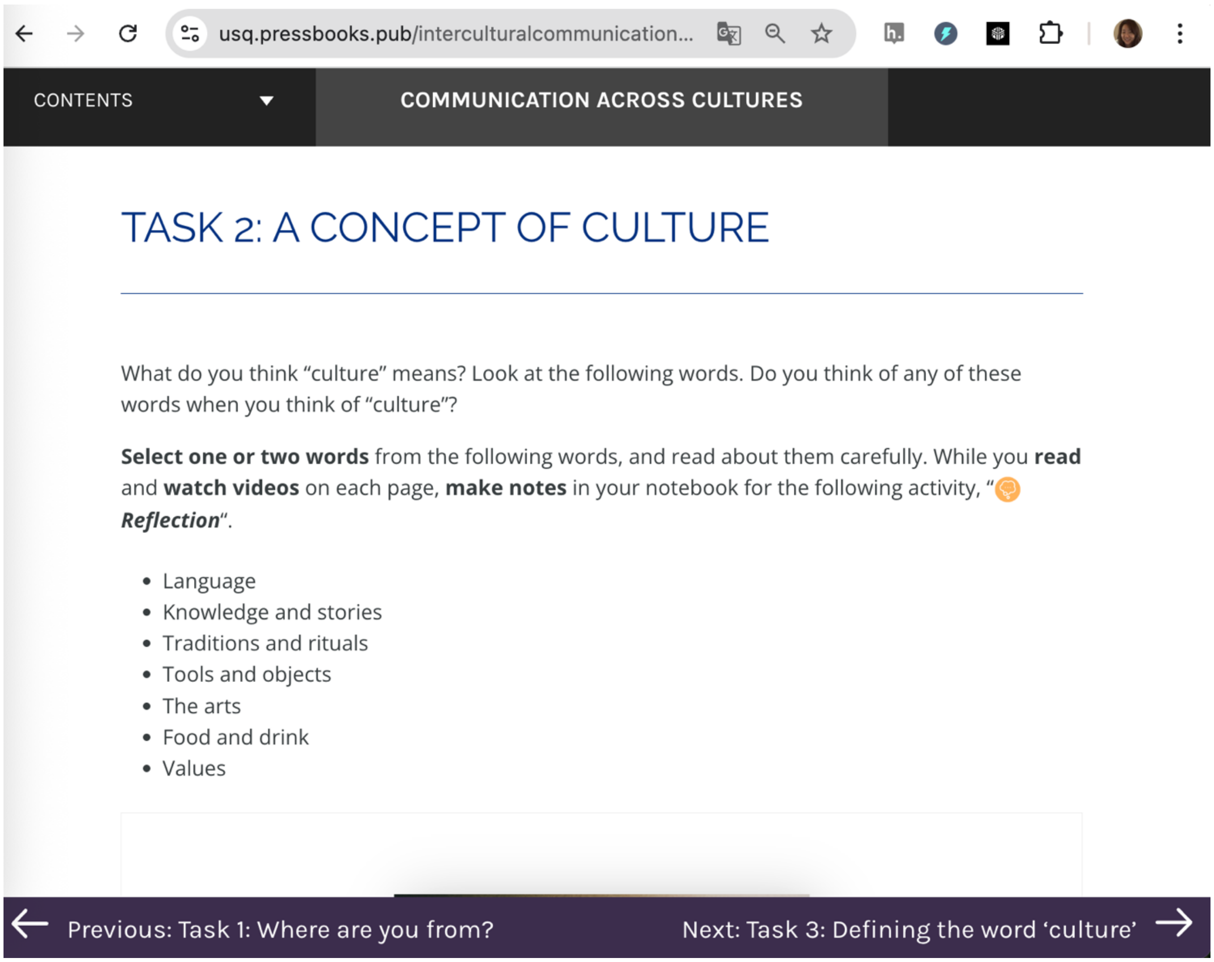
Design

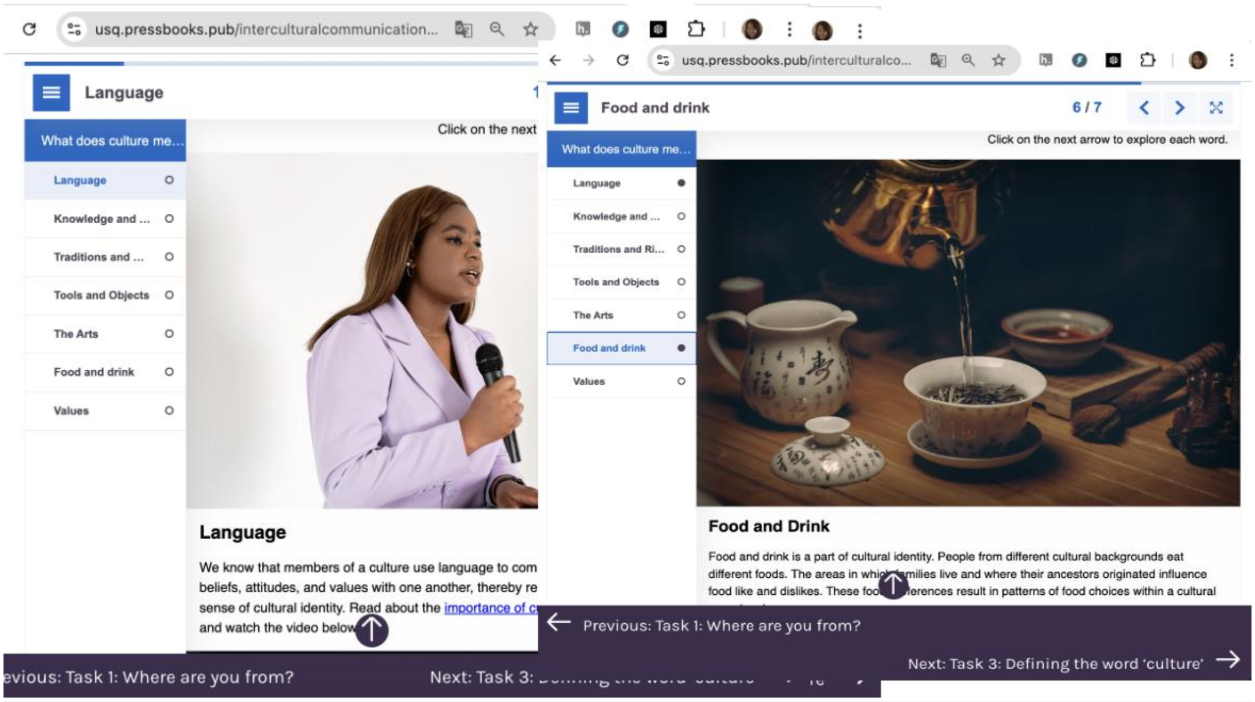
Visual Appeal and Layout Consistency

One teacher participant commented that, “The use of headings, subheadings, and bullet points helps to create a clear visual hierarchy.” Students likewise appreciated the structured appearance. However, a few sections were described as “somewhat crowded,” and participants in one usability testing group found the text too dense in certain sections. They suggested incorporating more visuals or breaking up text with videos and/or images to maintain interest. This suggestion was later incorporated into a revised version of the materials, which expanded a single page of text into multiple pages incorporating multimedia activities, as shown in Figure 1.

Figure 1

Example of Revised OER Materials: From Single-Page Text to Interactive Multi-Page Design





Note. Top panel lists the topics to be covered in Task 2: Language and Culture. The smaller panels are examples of activities covering two of those topics – Language and Food and Drink.

The participants (students and teachers) responded very positively to the presence of icons and images that signalled different activities (reading, listening, reflection, etc.). They found them effective, noting, "I like to pictures [icons], very useful and look nice" and "It's beneficial to see different images, as they make it easier for the students to understand what to do." However, one area for improvement in design consistency was identified: some icons were not used uniformly. The tasks involving speaking/recording and the end-of-module learning journal lacked an icon. A teacher noted, "I noticed that each module ends with a learning journal. Including an image similar to those used for other tasks would help readers recognise the consistent structure."

Interactive Elements and Multimedia Integration

All participants were positive about the interactive elements. One teacher said, "Quizzes and multimedia content enhance engagement and understanding." Students in both groups echoed this: they described the interactive quizzes as challenging but motivating, and found the video clips "very interesting." Three students stated they would like to see more of these kinds of material. During the workshop, it was observed that both student groups attempted the embedded quiz questions while using online dictionaries and discussing answers among themselves. Students also commented favourably on tasks that let them record themselves speaking: "Recording many times is good," and "Listening my recording is good to know what mistake I made." Additionally, another student, who had taken part in video recording activities in the EAP course, suggested that "video recordings would be even more beneficial like what we did, VoiceThread." They saw potential for even richer interactive practice (VoiceThread is a platform for video and voice discussions) beyond what was included. These reactions suggest that the integration of multimedia (video, audio) and interactive tools (H5P activities, recording, and note-taking) was a strength of the design.

Navigation

Ease of Use/User-Friendly Interface

Feedback on the overall navigation structure of the materials was generally positive, with some caveats for first-time users. Several commenters praised the interface. For example, a student noted, “There is a table of contents [on the left], which makes the book [structure] easy to access.” A teacher similarly appreciated that the modules and tasks were “easily accessible” via the sidebar menu and hyperlinks. Those remarks suggest that once users noticed the navigation menu and links, they found it intuitive to jump between sections.

However, challenges were observed, particularly at the beginning of the workshops. Two students seemed unsure about how to start navigating. One student clicked around randomly on various headings, possibly overwhelmed by the options. This was corroborated by a student: “At the beginning, I wasn’t sure what to do, what to click. It’ll be much easier, if we work with the teacher.” This indicates that first-time users might need a brief orientation. In practice, one group’s confusion was resolved after a few minutes as they discussed among themselves and realised how the content was structured. In contrast, a student in the second group, and the teachers, described the interface as “intuitive.” This difference could be due to individual familiarity with e-books. Some users will immediately understand a digital book layout, while others will benefit from guidance.

Results that address R1 (ease of navigation) show that generally, the navigation was easy, but initial orientation could be a challenge. A potential solution (suggested implicitly by the feedback) would be to include a brief tutorial or guide at the very start of the book. In a classroom setting, a teacher could carry out this introduction.

Functionality (Search, Bookmark, Taking a Note, Progress Tracking)

The search function was useful for finding specific topics or keywords. Despite the positive reception, however, feedback from a student and a teacher suggested that the addition of highlighting tools would further improve the effectiveness of these resources. A few students experienced issues with “saving and accessing their notes” and a teacher pointed out that: “[recording and written answer] files [for the tasks] are not automatically saved, which is a nuisance because it requires manual saving on their computer.” This is indeed a limitation of using the OER platform outside a learning management system (LMS)—data is not stored between sessions. These comments highlight that while the existing functionality was appreciated, enhancements could greatly improve usability. A teacher suggested features such as a bookmark to remind students where they left off or a progress indicator for completed sections. These are common in a LMS, but Pressbooks does not have such features. Although implementing progress-tracking enhancements may be beyond the scope of content development, acknowledging them is important.

One specific positive observation was how some students discovered the link “Listen to this section,” which is an audio option in one of the tasks (a clickable accessibility feature to hear the text read aloud). They pointed out that listening to the correct pronunciation of words was helpful for improving their listening and speaking practice. This suggests that it would be worth exploring other facilities for enhancing interactivity (e.g., text-to-speech or speech-to-text) to better meet users' needs and improve usability.

Links

There were no reports of broken links. However, during the observation, one group had trouble with a video in module 3, task 4. The student who was navigating the book on behalf of two other students clicked on the link several times, apparently unsure whether there was a problem with slow downloading or whether the link was broken. Slow downloading turned out to be the problem rather than an issue with the link itself. We realized that an embedded loading indicator or a note such as "Video may take a moment to load" could alleviate confusion. It is a minor point, but it affected the flow for that group until they figured it out.

Cross-Device Compatibility

The ability to access materials on a variety of devices, including desktops, tablets, and smartphones, was praised as it provided flexibility. One student commented, "I accessed the book with my phone while I was on the bus [after the workshop]. It was good to watch the videos and do some tasks." This spontaneous feedback was encouraging—it means the responsive design worked, and students appreciated the flexibility of being able to study on-the-go. No one reported any layout breakages or difficulties on mobile, so it appears the responsive design considerations (stage one actions) were effective.

Content

Educational Value and Relevance

Both students and teachers regarded the content of the book as valuable and relevant to their needs. Participants noted that the materials covered a range of cultural topics from general concepts (e.g., definitions of culture, culture shock) to specific real-world scenarios (e.g., challenges). The students in the workshops claimed that as in the examples in the materials, they had had experiences of being treated differently by people, which made them "angry" or led to "frustration," or "missing [their] home [countries]." One student said in the workshop, "She [a character in a video] is in a similar situation .... I had the same experience." This indicates the content successfully prompted learners to make connections to their own lives, a key goal for intercultural learning. They also commented that, as a result, they acquired some strategies for coping in such situations, stating, "It is good to know how to deal with. I fought when I had similar experience in the school," which implies previously they might have responded aggressively.

One teacher found the topics appropriate, though another cautioned that some topics could be "quite sensitive" to handle in class depending on students' language proficiency. For beginner-level learners, discussing complex intercultural issues might be challenging since they may lack the vocabulary to express their opinions. The teacher added, "Honestly, I'm not sure how to start discussing this issue, and I don't believe it's my job." This comment reflects a concern some teachers have: facilitating intercultural discussions can be delicate, especially if issues stray beyond language teaching into personal territory.

Participants also reported that the embedded links provided in various tasks in each module were useful. The students expected, according to their comments, that these would help them "expand" their knowledge of the topics. The teachers pointed out that the materials could "offer alternative sources" to meet the needs of different cohorts. Regular updates and expansion of these resources could further benefit users. This affirms that including ease of updating and external resources is a strength of OER, as long as technical access is smooth. Also, this flexibility is indeed one advantage of OER—teachers or learners can pick and choose or extend content as needed.

The students in the workshops highlighted the videos and recording tasks as providing effective opportunities for them to practice strategies for improving their English. Encouraging students to research definitions on their own was also considered a good strategy for enhancing academic skills. One student stated, "I like this task, *research*, anyway we have to do in university study, so it is helpful."

Some challenges were identified by the participants. For example, the quiz in module 2 was reported to be quite difficult to understand and required more explanation. Also, task 1 in module 1 was noted to be overly "technical," which might be challenging for users with limited digital literacy.

Interaction

The tasks that included videos were the most engaging. Participants commented positively on tasks that included real-life examples and personal stories, which tended to enhance their interest and motivation. During the workshop, students interacted and communicated with other students, sharing their opinions and experiences. One student confessed to the others, "[In the video], she is in a similar situation, yeah, I think she is right. I had the same experience." They appeared to appreciate these kinds of material, as they clicked and watched until the video finished. This interactive approach may not only provide a more dynamic learning environment but also encourage students to relate the material to their own lives, which may help deepen their understanding of and interaction with the content.

Overall, multimedia elements appeared to be highly effective in aiding understanding, with videos and interactive tasks receiving particularly positive feedback. Students commented that they "like videos, easy to understand" because they are "more interesting" than just reading or listening. During the workshop, we noted that most of the videos were watched and provoked some discussion and while they were watching them, students clicked subtitles to read, paused, and returned to watch again.

The use of multimedia is an important feature of the materials and proved popular with users. Further expanding these elements and ensuring they are seamlessly integrated into the content is therefore likely to enhance learning outcomes.

Cultural Sensitivity and Inclusivity

An important aspect of content for this project was ensuring cultural inclusivity. The content seemed to be generally culturally sensitive and inclusive, recognising and respecting diverse perspectives. Participants expressed satisfaction when they saw videos and reading materials that represented their cultural backgrounds and perspectives. One student commented, "The stories and examples are from different cultures and my own. And that is true in general what they said." A teacher also commended the selection of videos and images that represented "various races." Such positive feedback is crucial, as one risk in the development and use of intercultural materials lies in the inadvertent privileging of one culture or the misrepresentation of cultural perspectives. This inclusivity likely enhanced interaction as well. When learners see themselves or their experiences validated in learning materials, it can increase motivation and trust in the material. The feedback underscores the importance of incorporating diverse cultural perspectives in educational materials to enhance relatability and interaction among students (e.g., migrants, international students, and refugees).

Task Instruction

Clear task instructions are essential for students and teachers, whether using this material independently or in the classroom. Feedback on the instructions was mixed. Some participants (both students and one teacher) commented that there was a need for more detailed instructions. A teacher indicated that some words with multiple meanings, such as "demonstrate," might confuse learners. The recommendation was to use simple, common language to avoid misunderstandings, particularly if this book is to be used for self-study. Another teacher suggested including examples under the instructions to guide students on how to start their tasks. On the other hand, a few students cautioned that overly detailed instructions might cause readers to "lose interests" and become "demotivated" before even attempting the task. This reveals a tension: lower-proficiency learners might need step-by-step guidance, whereas others prefer brevity to maintain motivation.

Additionally, while "making notes in a notebook" is beneficial for some students, others find "typing answers is easier" because it allows them "to copy, delete, and check spelling." Writing on paper has certain advantages, but it is also "easy to lose." This comment actually pertained to task format (writing by hand vs. typing in the digital book), but it also connects to the instructions, which encouraged students to take notes on paper. The first ones followed that instruction literally and liked it, whereas the second group found their own way, typing digitally, implying the instructions might need to accommodate multiple modes or at least not mandate one way.

Despite these variances, most agreed that the language of the instructions was generally clear and understandable (no one reported not understanding what to do). The main issues were about scope and how much guidance to include. A suggestion made by one teacher to address this was to include a simple example for complex tasks. For instance, under a reflection question, provide a one-sentence example answer to illustrate what is expected. This idea was positively received by the researcher and implemented to help users get started without over-explaining in the main text.

Discussion

The usability testing answered our research questions and provided actionable insights. While there was general agreement about the materials among participants, differences highlight the importance of considering both the context in which educational materials are used and the individual learners and teachers when interpreting feedback and making subsequent revisions.

R1: Ease of Navigation

The study found that once familiar with the interface, students and teachers could navigate the materials with relative ease. The table of contents and module structure were effective in guiding users by providing multiple navigation pathways. However, initial confusion experienced by some students needs to be taken into account. This highlights a key point from the feedback: the *context* in which materials are used influences perceived usability. If a teacher is present (classroom context), they can introduce the resource and thus mitigate navigation issues. In an independent study context, the material itself should provide that orientation. This consideration of context underscores the importance of aligning the design of educational materials with their intended use (Laurillard, 2012).

Because our materials as OER might be used either with or without a teacher, we need to ensure it serves both. In practice, we have considered including a quick-start tutorial for independent users. This minor addition could balance the needs of students in different contexts, thus improving initial ease of use without overloading the interface for those who do not require it, as noted by Sanders and Lafferty (2010).

R2: Useful and Interactional Features

Participants valued the interactive and multimedia features of the resource. The positive reactions to videos, quizzes, and recording tasks confirm that these elements enhanced interactions and learning. This posits that well-integrated verbal and visual materials can improve understanding. The open-book format enabled these features, illustrating the advantage of OER-enabled pedagogy (Wiley & Hilton, 2018) that leverages openness and technology to create more engaging learning experiences. Student collaboration around quizzes and discussion of video content, for example, demonstrates active learning, which Kessler and Plakans (2001) emphasised by involving learners as stakeholders in materials evaluation. The findings support the idea that learners, when given interactive content, often take the initiative (e.g., by using dictionaries or replaying videos) to deepen their understanding. This aligns with the concept of self-directed learning in OER (Olivier, 2019) and shows that digitally literate learners will use the tools at hand to aid their learning.

One interesting point is that certain features had a dual nature: for example, the note-taking method. The resource allowed typing notes, but some students chose to write on paper. This highlights individual preferences in how students interact with content. As Mueller and Oppenheimer (2014) found, writing notes by hand can have cognitive benefits for some, while others prefer the convenience of digital notes. Given these differing preferences, it may be advantageous to encourage learners to experiment with both methods to discover what works best for them. Additionally, providing materials in multiple formats—such as printable worksheets for those who prefer to write by hand—could help accommodate these varying preferences, especially if a digital book is used in a classroom setting.

R3: Difficulties Encountered

As is often the case in deliberative settings, one of the notable aspects of the comments was the presence of disagreements and apparent contradictions among participants. For example, while some experienced difficulty in processing some of the written information in particular, others appreciated the clarity that such details provided. This dichotomy suggests that there is no one-size-fits-all approach to the design of instructional materials. The challenge for educators and course designers lies in balancing these opposing needs. Ideally, providing tiered instructions—where learners and teachers can choose between a basic outline or a more detailed guide, depending on their comfort level and prior knowledge—could address these differences. The effect of these individual differences is directly mirrored in the work of Kessler and Plakans (2001) who pointed out that learners are individuals and are affected by their environment.

The usability issues encountered often traced back to individual differences rather than universally poor design. This does not mean we should dismiss them; rather, it means the design should strive for flexibility. The positive side is that because this OER is in a digital format, we have more leeway to adapt and iterate than with a print textbook. We can refine instructions, add alternative pathways, or provide additional resources to cater to different needs.

Another difficulty was technical friction, such as the lack of a note-saving feature and the impression of a broken link due to slow loading. These are reminders that technical usability (e.g., interface design, performance) is as important as content usability. Slow Internet or lack of an auto-save feature are external to content but internal to user experience. For our project, this suggests that when deploying an OER, considering the technological infrastructure of our learners (e.g., Internet speed, device usage) is crucial. In contexts where connectivity is an issue, common for some refugee learners, we should provide low-bandwidth alternatives, such as downloadable PDFs, to ensure those technical difficulties do not impede learning. Sanders and Lafferty (2010) emphasised that effective e-learning requires attention to such usability details to avoid cognitive overload with the medium itself.

R4: Suggested Improvements

Participants were not short on suggestions, and importantly, many suggestions were actionable and lined up with best practices. The call for features such as highlighting, bookmarking, and progress tracking indicates that users expect a digital learning experience to have personalisation capabilities. Moreover, the desire for such features shows a level of engagement where learners want to interact more deeply with the materials. This is encouraging because if we supply the means, learners are inclined to actively study the content, not just passively read it.

Another improvement suggestion, though implicit, was to keep content updated and consider alternative content if something does not work. This speaks to the sustainability of OER materials. Unlike a traditional textbook that might go unchanged for years, open materials benefit from continuous improvement. Our plan to periodically update links, add new case studies or videos, and incorporate user feedback is aligned with the OER ethos of iterative enhancement (Wiley & Hilton, 2018). Regular updates can also address any future usability issues that arise as technology or student expectations change.

Conclusion and Recommendations

This study provides an example of applying usability testing to the development of open educational resources and is consistent with other studies emphasising the importance of incorporating usability testing into resource design prior to full implementation (e.g., Doubleday et al., 2011; Kessler & Plakans, 2001).

As curricula and course details undergo modification, it is likely that instructors will seek to modify existing online resources to be more closely aligned with changing course goals or formats. Many classrooms may also increasingly rely on online resources for reinforcement or delivery of content, and priority must be given to testing resource interfaces and to assessing a student's ability to quickly and appropriately navigate through a system. A follow up to this study will investigate the effectiveness of student learning while using the resources.

As more educators attempt to develop technology-based materials, it is important to ensure the appropriateness and usability of these materials, especially to reduce the considerable cognitive load imposed on students by language, content, and technology. The involvement of students in the development process proved to be valuable, and the testing procedures employed in this study represent one of many approaches that may be used to involve students in assessing the usability of materials.

Acknowledgements

This project is funded by the Open Education Practice Grant, part of the Learning and Teaching Grants Program at the University of Southern Queensland.

References

Barnum, C. M. (2020). *Usability testing essentials: Ready, set … test!* (2nd ed.). Morgan Kaufmann.

Borthwick, K., & Gallagher-Brett, A. (2014). "Inspiration, ideas, encouragement": Teacher development and improved use of technology in language teaching through open educational practice. *Computer Assisted Language Learning*, *27*(2), 163–183. https://doi.org/10.1080/09588221.2013.818560

Brook, J. (2011). The affordances of YouTube for language learning and teaching. *Hawaii Pacific University TESOL Working Paper Series, 9*(1, 2), 37–56. https://hpu.edu/research-publications/tesol-working-papers/2011/9_1-2_Brook.pdf

Burrows, K. M., Staley, K., & Burrows, M. (2022). The potential of open educational resources for English language teaching and learning: From selection to adaptation. *English Teaching Forum, 60(*2), 2–9. https://americanenglish.state.gov/files/ae/60_2_pg02-09_cx2_corrected.pdf

Chang, H., Windeatt, S., & Stockwell, E. (2025). *Communication across cultures*. University of Southern Queensland. https://usq.pressbooks.pub/interculturalcommunication/

Doubleday, E. G., O'Loughlin, V. D., & Doubleday, A. F. (2011). The virtual anatomy laboratory: Usability testing to improve an online learning resource for anatomy education. *Anatomical Sciences Education, 4*(6)*,* 318–326. https://doi.org/10.1002/ase.252

Dumas, J. S., & Redish, G. (1999). *A practical guide to usability testing* (rev. ed.). Intellect Books.

Getenet, S., Cantle, R., Redmond, P., & Albion, P. (2024) Students' digital technology attitude, literacy and self-efficacy and their effect on online learning engagement. *International Journal of Educational Technology in Higher Education, 21*, Article 3. https://doi.org/10.1186/s41239-023-00437-y

Kessler, G., & Plakans, L. (2001). Incorporating ESOL learners' feedback and usability testing in instructor-developed CALL materials. *TESOL Journal*, *10*(1), 15–20. https://doi.org/10.1002/j.1949-3533.2001.tb00012.x

Laurillard, D. (2012). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.

Liddicoat, A. J., & Scarino, A. (2013). *Intercultural language teaching and learning* (2nd ed.). Wiley-Blackwell. https://doi.org/10.1002/9781118482070

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science, 25*(6), 1159–1168. https://doi.org/10.1177/0956797614524581

Nielsen, J., & Landauer, T.K. (1993). A mathematical model of the finding of usability problems. *Proceedings of the INTERCH93: Conference on Human Factors in Computing Systems*, The Netherlands, 206-213. https://doi.org/10.1145/169059.16916

Olivier, J. (2019) Towards a multiliteracies framework in support of self- directed learning through open educational resources. In E. Mentz, J. de Beer, & R. Bailey (Eds.), *Self-directed learning for the 21st century: Implications for higher education* (pp. 167–201). AOSIS Scholarly Books. https://doi.org/10.4102/aosis.2019.BK134.06

Rubin. J., & Chisnell., D. (2008). *Handbook of usability testing: How to plan, design, and conduct effective tests* (2nd ed.). Wiley. https://content.e-bookshelf.de/media/reading/L-571931-cbc66f3178.pdf

Sanders, J., & Lafferty, N. (2010). Twelve tips on usability testing to develop effective e-learning in medical education. *Medical Teacher, 32*(12), 956–960. https://doi.org/10.3109/0142159X.2010.507709

Sugar, W. (1999). Novice designers' myths about usability sessions: Guidelines to implementing user-centered design principles. *Education Technology, 36*(6), 40–44. https://www.jstor.org/stable/44428569

Tomlinson, B. (2012). Materials development for language learning and teaching. *Language Teaching*, *45*(2), 143–179. https://doi.org/10.1017/S0261444811000528

UNESCO. (2019, November 25). *Recommendation on open educational resources (OER)*. UNESCO. https://www.unesco.org/en/legal-affairs/recommendation-open-educational-resources-oer

Weller, M. (2010). Big and little OER. In *Open Ed 2010 Proceedings*. Universitat Oberta de Catalunya, Open Universiteit Nederland, & Brigham Young University. https://hdl.handle.net/10609/4851

Wiley, D., & Hilton, J. L., III. (2018). Defining OER-enabled pedagogy. *The International Review of Research in Open and Distributed Learning*, *19*(4). https://doi.org/10.19173/irrodl.v19i4.3601

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

The Role of Open and Distance Education in Reducing the Educational Gap in Indonesia

Yuyun Alfasius Tobondo
Universitas Kristen Tentena, Indonesia

Abstract

This study analyzed the role of open and distance education (ODE) in reducing the educational gap in Indonesia, particularly between urban and rural areas. The research method was a literature review that collected and analyzed various articles and reports related to ODE in Indonesia and other developing countries. The findings indicated that ODE had great potential to enhance access to education in remote areas; however, this potential has been constrained by uneven technological infrastructure, lack of teacher training, and educational policies that are not yet inclusive. This study recommended improvements in policies to support the development of digital infrastructure and continuous teacher training. Practical recommendations include providing subsidies for devices and Internet access for students in rural areas. While ODE has provided solutions to similar educational gaps in other developing countries, such as India and Nigeria, the implementation of more inclusive policies and enhanced teaching capacity is necessary to optimize its benefits in Indonesia.

Keywords: educational access, educational policies, digital literacy, distance education, technological infrastructure

Introduction

Open and distance education (ODE) is an educational model that allows students to access learning without the need to be physically present in a classroom. It offers significant flexibility, as students can engage in the learning process anytime and anywhere, provided they have access to the necessary technological devices, such as computers or smartphones. ODE typically uses information and communication technology to deliver educational content, enabling students from various social backgrounds to access education, even if they live in remote locations or do not have the time to participate in traditional formal education.

In Indonesia, ODE has developed rapidly in recent decades, with universities such as Universitas Terbuka (UT) playing a crucial role in providing access to higher education for students from various regions, particularly in rural areas where educational facilities are limited. ODE has become a primary solution in bridging the educational gap, offering learning opportunities to those previously hindered by physical, social, or economic constraints.

However, despite ODE offering greater opportunities for educational access, significant challenges remain, particularly regarding the disparity in education quality between urban and rural areas, technological limitations, and the preparedness of teachers to manage online learning. Therefore, research on ODE is essential to understand how this system can be more effective in addressing the educational gap in Indonesia.

In Indonesia, the educational gap between urban and rural areas is striking. According to data from Kurnia et al. (2023), children living in rural areas often face limited access to quality educational facilities, including a shortage of trained teachers and adequate learning resources. In urban areas, students tend to have better access to technology, a more modern curriculum, and higher-quality educators. For example, in Indonesia, 70% of students attending school in urban areas have regular access to the Internet, while in rural areas, only about 35% have stable access to such services (Fagbamigbe et al., 2021). This data suggests that, although ODE offers great potential to reach students in rural areas, limited access to digital infrastructure remains a major barrier that needs to be addressed.

This gap is also reflected in health, where higher rates of illness in rural areas, such as diarrhea, affect student attendance and concentration in school (Fagbamigbe et al., 2021). Additionally, economic factors in rural areas often limit parents' ability to support their children's education, both in terms of tuition fees and access to devices necessary for participating in distance learning.

The phenomenon of the educational gap between urban and rural areas is not unique to Indonesia. In other developing countries such as India and Nigeria, this gap is also quite evident. In India, more than 60% of the rural population lacks adequate access to technology for distance education, while in Nigeria, many students in remote areas are unable to access online learning materials due to internet limitations (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2022; World Bank, 2021). This highlights that, although ODE offers significant potential, disparities in digital infrastructure and teaching quality remain key challenges that need to be addressed.

This study aimed to provide an original contribution by identifying the main challenges faced by Indonesia in implementing ODE as a tool to reduce the educational gap. One of the key contributions of this research is to offer policy recommendations that support the development of more equitable digital infrastructure and enhance teachers' competencies in managing distance learning. These recommendations also emphasize the importance of policies that target vulnerable groups, such as students in rural areas, while considering the economic, social, and health factors that affect their access to quality education.

Several studies have shown that the proper implementation of technology can enhance the effectiveness of ODE. For example, the use of learning management system (LMS) platforms such as Moodle and Google Classroom has proven effective in increasing student engagement in distance learning (Salwa, 2021). In Indonesia, Universitas Terbuka (UT) has utilized LMS to support students in remote areas, although challenges related to limited infrastructure persist (Heryanto, 2021). A study by Heryanto (2021) also indicated that the use of educational technologies such as LMS can improve learning outcomes when supported by adequate training for both teachers and students.

However, the effectiveness of this technology is highly dependent on the readiness of digital infrastructure and teaching competencies. For instance, in some areas, even though there is access to LMS, Rahmi Rivalina (2013) suggested that many teachers were not skilled in using the technology optimally. Therefore, policies supporting digital training for both teachers and students are crucial.

The success of ODE not only relies on the technology used but also on supportive policies and appropriate pedagogical approaches. As noted by Savitri et al. (2023), educational policies that focus solely on providing technology platforms without considering the readiness of the entire educational ecosystem, such as teacher training and infrastructure development, have often been ineffective. Therefore, a holistic approach that combines inclusive policies, technology development, and adaptive teaching methods is necessary to ensure the success of ODE.

In Indonesia, although the government launched the Merdeka Belajar policy aimed at decentralizing the education system and facilitating technology-based learning, the implementation of this policy has remained limited to certain regions only (Irhamsyah, 2023). This policy did not place sufficient emphasis on strengthening technological infrastructure in remote areas, as well as providing digital training for both teachers and students to ensure equitable access to education.

This study analyzed the role of ODE in addressing the educational gap in Indonesia, particularly in rural areas. The main focus of this research was to explore the challenges and opportunities present in the implementation of ODE, considering factors such as technology readiness, teacher competencies, and policies that support distance learning. With an evidence-based approach, this study also aimed to provide policy recommendations that support the development of ODE as a tool to enhance equitable and inclusive access to education across Indonesia.

Overall, while ODE holds great potential in expanding access to education in Indonesia, significant challenges still remain, particularly regarding the technology gap and varied levels of teaching competence. This study identified that the success of ODE depends heavily on the readiness of digital infrastructure, the quality of teaching, and policies that support distance learning in an inclusive and sustainable manner. Therefore, more integrated policies are needed, encompassing the development of technological infrastructure, digital training for teachers, and strengthening support for students from diverse social and economic backgrounds.

Research Method

This study conducted a literature review to identify and analyze previous research relevant to the topic of ODE and the educational gap between urban and rural areas in Indonesia. The literature review collected and synthesized articles, policy reports, and scientific studies focused on the implementation of ODE in Indonesia and other developing countries. This approach allowed the author to examine the factors that have influenced the success or failure of ODE and to identify gaps in existing research that need to be addressed.

Research Steps

Data Collection

Data collection was conducted by accessing and analyzing various journal articles, research reports, and academic publications discussing ODE, particularly those related to Indonesia. The search process examined leading academic databases such as Google Scholar, Scopus, and ERIC. Relevant articles on ODE and the education gap were selected based on keywords related to the topic, such as open and distance education, Open and Distance Education, online learning, education gap, and rural-urban education inequality.

Articles included in this study were those published in indexed journals (e.g., Scopus, Web of Science) that directly addressed ODE, educational gaps in Indonesia or other developing countries, and the challenges and solutions in implementing ODE. Articles that were not relevant to the topic of ODE, lacked empirical data, or provided only superficial analysis were excluded. As well, articles that did not use clear or reliable methodologies were excluded from the literature review.

Analysis Method

This study used narrative-based qualitative analysis. Selected articles were examined to identify key findings related to the implementation of ODE, including aspects of policy, technology, pedagogy, and the challenges faced. The main focus of this analysis was to understand how ODE can reduce the educational gap between urban and rural areas and to provide policy recommendations that can enhance its effectiveness.

Synthesis of Findings

Once analyzed, the collected data was synthesized to explore patterns and consistent findings in previous research. These findings were categorized into several key themes, including:

- Quality of education and access in rural versus urban areas
- The impact of digital infrastructure on the effectiveness of ODE
- Policies that have supported and/or hindered the implementation of ODE
- Successes and challenges in using educational technologies such as LMS and massive open online courses (MOOCs)
- The role of teachers in implementing ODE and the training needed to improve online learning quality

Evaluation of Limitations in Previous Research

This study also evaluated the limitations in previous research, such as the lack of studies addressing the long-term impact of ODE on students' academic achievements and social mobility, as well as limitations in understanding the integration of educational policies with technological infrastructure.

Analysis

Based on the analysis of various studies relevant to the topic of ODE in Indonesia, several key findings were identified, which can be explained across different themes. These themes included (a) educational access, (b) quality of learning, (c) technological infrastructure, (d) educational policies, and (e) challenges faced by various stakeholders in the implementation of ODE.

Educational Access in Urban and Rural Areas

One of the key findings was the educational access gap between urban and rural areas in Indonesia. According to Kurnia et al. (2023), students in urban areas have had better access to quality educational facilities, including technological infrastructure and trained teachers. On the other hand, rural areas have faced limitations in terms of access to technology and adequate educational resources. According to Fagbamigbe et al. (2021), approximately 70% of students in major cities in Indonesia had regular access to the Internet, while only about 35% in rural areas had reliable Internet access. While ODE can offer greater learning opportunities in remote areas, digital infrastructure constraints remain a significant barrier to its implementation.

Additionally, although ODE can reduce the educational gap, limited access to digital devices, such as computers or smartphones, has long been a major challenge in rural areas. Savitri et al. (2023) noted that current policies did not fully accommodate the needs of students in areas with limited access to technological devices and the Internet, resulting in disparities in the quality of education students received.

Quality of Online Learning and Distance Education

The quality of online learning was also a key concern in this study. Heryanto (2021) found that the quality of online learning is uneven, especially in rural areas. Although various LMS platforms have been used to support online learning, many students and teachers lack the skills to manage these technologies optimally. This has led to low student engagement in learning activities and inadequate academic achievement. Rahmi Rivalina (2013) showed that while LMS platforms like Moodle and Google Classroom can improve access and student engagement, the lack of training and technical support for teachers in rural areas has hindered effective implementation.

Moreover, a study by Tsara Ayuninggati et al. (2023) found that one of the biggest challenges in distance learning is the low motivation and engagement of students, which was particularly noticeable among students from lower economic backgrounds. The lack of direct interaction with instructors and classmates made students feel isolated, which impacted their understanding of the material. Therefore, their study emphasized the importance of more interactive technologies, such as the use of blended learning, which combines online and face-to-face learning to improve quality and engagement in education.

Technological Infrastructure and Digital Readiness

Technological infrastructure is one of the key determinants of the success of ODE. Research by Mizal et al. (2021) revealed that uneven infrastructure is the biggest barrier to the implementation of ODE in Indonesia. While some major cities have adequate digital infrastructure, many in rural areas still face difficulties accessing high-speed Internet. In addition, the limited availability of digital devices for students poses another challenge that exacerbates the inequality in distance education. Research by Savitri et al. (2023) showed that more than 50% of students in rural areas relied on outdated devices to access online learning materials, which affected the quality of their learning experience.

The availability of technology has also affected the quality of teaching. Heryanto (2021) indicated that many teachers were not trained in using online platforms optimally, so despite the availability of technology, its effectiveness was hindered by a lack of training and understanding of how technology can support pedagogy. Therefore, this study emphasized the importance of enhancing teacher capacity, especially in managing online learning, to improve the quality of education delivered through ODE.

Educational Policies and Government Support

Educational policies that support the implementation of ODE are crucial for the success of this model. Research by Jusas et al. (2021) highlighted the importance of policies that cover all aspects of education, including the development of technological infrastructure, teacher training, and support for students from disadvantaged economic backgrounds. Although the Indonesian government has launched the Merdeka Belajar policy, aimed at introducing flexibility in education and strengthening technology-based learning, this policy has not yet been fully implemented across Indonesia, especially in rural areas (Irhamsyah, 2023).

More inclusive policies, which address infrastructure inequalities and ensure equitable access to technology, are essential for supporting the success of ODE. For example, policies that provide subsidies for technological devices or Internet access for financially disadvantaged students would greatly help reduce the access gap and ensure more equitable education (Savitri et al., 2023).

Synthesis of Research Findings

Overall, the findings of this study indicated that while ODE has great potential to reduce the educational gap in Indonesia, many challenges remain. The technology access gap between urban and rural areas, the low quality of online learning, and uneven digital infrastructure emerged as the biggest barriers to be addressed. Furthermore, the digital readiness of teachers and policies that support inclusive and sustainable distance learning are crucial to ensuring the success of ODE.

This study emphasized the need for more comprehensive policies that not only focus on providing technology platforms but also on developing equitable infrastructure and improving teaching capacity to support online learning. Additionally, the role of more interactive technologies and the use of blended learning models can enhance the quality of distance education in Indonesia, especially for students in remote areas who face access limitations.

As part of the synthesis of research findings, Table 1 presents a summary of the key findings from various articles relevant to the topic of ODE and the educational gap in Indonesia. This table provides a clear overview of the various studies that have been conducted, including the challenges and opportunities in the implementation of ODE in Indonesia, particularly in rural areas. Each study offered a different perspective on the factors that influence the effectiveness of ODE, such as the quality of learning, educational policies, and technological infrastructure.

Table 1

Summary of Findings From Literature Review

Research title	Citation	Research focus	Key findings	Research limitations
Implementation of Home Learning Policy During COVID-19	Ministry of Education Indonesia (2020)	Online education policy during the COVID-19 pandemic	The policy worked well despite challenges related to signal issues and varying student engagement	Limited to one location and does not cover all policies in Indonesia
Development and Validation of Online Classroom Learning Environment Inventory	Rahayu et al. (2021)	Developed and validated an instrument to assess student readiness for online learning	The instrument was valid for assessing access, interaction, and faculty support affecting student readiness	Did not account for Internet infrastructure in rural areas

Research title	Citation	Research focus	Key findings	Research limitations
Mind the Gap: What Explains the Rural-Nonrural Inequality in Education?	Fagbamigbe et al. (2021)	Educational inequality between rural and urban areas	Educational inequality higher in rural areas, influenced by economic and health factors	Does not directly link to distance education
Models for Administration to Ensure the Successful Implementation of Distance Learning	Jusas et al. (2021)	Administrative models for successful ODE implementation	The importance of strategic planning and infrastructure development to support ODE implementation	Did not cover policies and technology in depth
Growth and Collaboration in Massive Open Online Courses: A Bibliometric Analysis	Wahid et al. (2020)	Collaboration in MOOCs and its impact on open education	MOOCs have grown rapidly and can expand access to education, but quality challenges persist	Focused more on MOOCs, not on localized ODE implementation in Indonesia
Economic Zones and Local Income Inequality	Hornok & Raeskyesa (2023)	The impact of economic zones on income inequality and educational access	Economic zones exacerbate income inequality, affecting education	Did not investigate direct effects on ODE implementation in rural areas
The Role of Open and Distance Education in Bridging the Learning Gap	Belawati et al. (2020)	The role of ODE in reducing the educational gap	ODE provides broader access but is limited by infrastructure in remote areas	Did not fully integrate educational policies supporting ODE
Economic Inequality and Access to Education: Rural vs. Urban	Kurnia et al. (2023)	Educational gap between rural and urban areas in Indonesia	Educational access more limited in rural areas, influenced by economic and infrastructure factors	Did not discuss ODE implementation specifically

Table 1 illustrates several key findings from studies relevant to the implementation of ODE in Indonesia. By analyzing various aspects such as policies, infrastructure, technology, and teaching quality, it can be seen that although ODE has great potential in reducing the educational gap, there are still significant challenges that must be addressed. Therefore, more integrated policy recommendations that focus on infrastructure development and teacher training are necessary to ensure the success of ODE in Indonesia.

Discussion

This study analyzed the role of ODE in bridging the educational gap in Indonesia, particularly between urban and rural areas. Based on findings from the literature, several important aspects needed to be analyzed, namely (a) educational access, (b) the quality of online learning, (c) technological infrastructure, (d) educational policies, and (e) the challenges and opportunities faced in implementing ODE. In this discussion, each research question will be addressed by linking the key findings to the literature presented earlier.

Research Question 1: The Main Challenges in Implementing ODE in Indonesia

Based on the findings from the research by Fagbamigbe et al. (2021) and Kurnia et al. (2023), the main challenge faced in implementing ODE in Indonesia is the unequal distribution of technological infrastructure between urban and rural areas. Students in urban areas tend to have better access to technology and quality educational facilities, while those in rural areas often struggle to access high-speed Internet; some areas lack adequate Internet access altogether. According to Fagbamigbe et al. (2021), as few as 35% of students in rural areas have stable Internet access, while in urban areas, this rate is over 70%.

In addition, the quality of teaching is also a major challenge in implementing ODE. Heryanto (2021) showed that many teachers in rural areas were not trained in using online learning technologies such as LMS. The result was low-quality interaction between students and teachers, which greatly hampered the distance learning process. Isolated students with insufficient motivation often experience difficulties in understanding the material (Savitri et al., 2023).

Another challenge faced when implementing ODE is related to existing policies. Many educational policies do not fully accommodate the needs of students in remote areas. Research by Jusas et al. (2021) showed that more integrated policies, which included aspects of infrastructure development and teacher training, were essential to improving the effectiveness of ODE. The Merdeka Belajar policy launched by the Indonesian government, although focused on distance learning, has not been able to reach all regions of Indonesia equally, especially areas with limited access to technology (Irhamsyah, 2023).

Research Question 2: Factors That Influence the Success of ODE in Reducing the Educational Gap in Indonesia

The success of ODE in reducing the educational gap in Indonesia depends on several factors, the most significant of which are technological infrastructure, teaching quality, and supportive policies. According to Mizal et al. (2021) and Heryanto (2021), the success of ODE relied heavily on the readiness of digital infrastructure, particularly stable Internet access and adequate technological devices. In urban areas, such infrastructure has been sufficient to support the implementation of ODE, but in rural areas, this remains a major barrier.

In addition, teacher competence in managing online learning is also a crucial factor in determining the success of ODE. According to Rahmi Rivalina (2013) and Tsara Ayuninggati et al. (2023), teachers trained in using educational technology enhanced student engagement in the learning process and achieved better academic outcomes. On the other hand, teachers with less training in educational technology faced difficulties designing and delivering material effectively on online platforms, which ultimately reduced the quality of learning.

Policy factors are also important. The Indonesian government, through the Merdeka Belajar policy, has made efforts to support the development of distance education (Irhamsyah, 2023). However, as shown by Savitri et al. (2023) and Jusas et al. (2021), this policy has not fully addressed the access inequalities between urban and rural areas. Therefore, to ensure the success of ODE, more inclusive educational policies that focus on infrastructure development and digital training for teachers are essential.

Research Question 3: How ODE Can Reduce the Educational Gap

ODE has great potential in reducing the educational gap between urban and rural areas. According to Belawati et al. (2020), ODE provided an opportunity for students in remote areas to access education that may not be conventionally available in their regions. By using technology, ODE can reach students in rural areas who were previously isolated from the traditional education system. However, for ODE to be truly effective in bridging this gap, supportive policies and the development of equitable infrastructure are required. Kurnia et al. (2023) indicated that the government must address the inequality in access to technology and the Internet in rural areas. One solution is to provide subsidies for technological devices and Internet access for students in areas with limited access. Additionally, teacher training in the use of educational technology should also be an integral part of educational policies to ensure that the quality of teaching is maintained even in an online learning environment.

Significance of the Research Findings

The findings of this study have provided a deeper understanding of the challenges and factors affecting the effectiveness of ODE in reducing the educational gap in Indonesia. This research showed that, while ODE has the potential to improve access to education in remote areas, significant challenges related to infrastructure, teaching quality, and existing policies must be addressed to achieve this goal. Therefore, this study has not only provided a clear picture of the existing issues but also offers policy recommendations that could help optimize ODE.

Contribution to the Field

This research has contributed to the field of education by introducing a new perspective on the challenges and opportunities faced in the implementation of ODE in Indonesia. The main contribution of this study is providing more integrated policy recommendations that can support the development of technological infrastructure and teacher training in rural areas. This study has also enriched the literature on distance education by bringing together various factors influencing the success of ODE implementation, including policies, technology, and teaching quality.

Implications and Limitations of the Study

This study has emphasized the importance of developing more inclusive and equitable policies to support the implementation of ODE across Indonesia. Policies that include providing better access to technology in rural areas, digital training for teachers, and strengthening educational infrastructure should be prioritized. Additionally, this study has highlighted the need for continuous evaluation of existing policies to ensure that ODE can effectively reduce the educational gap.

However, this study had several limitations. One was the lack of available data on the implementation of ODE in specific areas, particularly those not covered by existing research. Furthermore, this study relied on secondary literature, meaning that empirical data obtained from surveys or direct interviews with students and teachers in rural areas was not fully considered. Further research involving primary data would be valuable to gain a deeper understanding of the effectiveness of ODE in the field.

Figure 1 provides a visual representation of the key factors influencing the successful implementation of ODE in Indonesia. It illustrates how various aspects, such as technological infrastructure, teacher competence, educational policies, and the quality of learning, all interact within the context of ODE. The challenges to be faced and recommendations needed to enhance the effectiveness of ODE in Indonesia are also clearly reflected in this diagram.

Figure 1

Visualization of the Factors Influencing the Implementation of ODE in Indonesia

Figure 1 clearly shows the relationships among factors that influence the success of ODE, including the uneven infrastructure between urban and rural areas, as well as the importance of teacher training and supportive educational policies. By understanding these factors, stakeholders can more easily design policies that can improve the implementation of ODE across Indonesia. This image emphasizes the importance of a holistic approach that integrates infrastructure, policies, and teaching to achieve more equitable and quality educational outcomes.

Conclusion

This study analyzed the role of ODE in reducing the educational gap in Indonesia, focusing on the differences between urban and rural areas. The main findings of this research showed that ODE has great potential to address the educational access gap, particularly for students in rural or underserved areas. However, the effectiveness of ODE in achieving this goal is still hindered by several challenges, including unequal technological infrastructure, limited teacher training, and uneven educational policies. Although ODE can improve access to education, improvements in learning quality, the provision of adequate technology, and the development of teaching capacity are necessary to achieve optimal results.

These findings make a significant contribution to the field of education by clarifying the relationships among policy, technology, and pedagogy in the context of distance education. This research highlighted the importance of strong infrastructure support and ongoing teacher training as key factors for the success of ODE. Additionally, the study emphasized that inclusive policies focused on reducing socio-economic inequality, such as providing device and Internet access subsidies, can play an important role in maximizing ODE's potential as a tool for reducing the educational gap in Indonesia.

The findings of this study aligned with the community of inquiry (CoI) model, which emphasizes the importance of interaction between three key elements in online learning: teaching, cognition, and social community (Martin, Wu, Wan, & Xie, 2022). In the context of ODE, the quality of interaction between students and instructors, as well as student engagement in the learning community, is crucial to supporting the success of distance learning. However, this study also challenges several assumptions in the CoI theory, particularly regarding the readiness of infrastructure and the digital competencies of teachers in rural areas. The lack of equitable access to technology hinders the formation of an effective learning community, which should be an integral part of the ODE model.

This research opens opportunities for further studies in several specific areas. First, further research on student experiences in online learning is essential. A qualitative study exploring the perceptions and challenges faced by students in rural areas will provide deeper insights into how ODE can be optimized to improve student engagement and learning outcomes. Second, research on the impact of digital literacy training for teachers and students is needed to evaluate whether such training can enhance the effectiveness of online learning and reduce educational access gaps.

Based on the findings of this study, several practical policy recommendations can be proposed. First, the government needs to provide subsidies or Internet access assistance to students in remote areas who lack adequate connectivity. Similar policies have been established in other developing countries, such as India, where the government provides internet subsidies and digital devices to students from low-income families to ensure more equitable access to education (UNESCO, 2022). Furthermore, the government and educational institutions should prioritize training for teachers in the use of educational technology, particularly to manage effective online learning. This training should include both technical skills and pedagogies that support distance learning.

In the implementation of ODE, socio-economic factors and cultural context significantly influence the success of the program. Students in rural areas often face socio-economic challenges, such as financial limitations and health issues, which hinder their access to education. Therefore, policies supporting ODE should consider not only technological aspects but also the socio-economic and cultural factors that affect students' lives. For example, the introduction of programs that support low-income families, such as subsidies for purchasing digital devices, could help address existing socio-economic barriers.

The importance of stakeholder involvement, including the government, educational institutions, the private sector, and the community, in the implementation of ODE cannot be overstated. Collaboration among these parties is necessary to ensure that ODE operates effectively and equitably across Indonesia. For example, collaboration between the government and technology companies to provide affordable digital devices for students could accelerate the equalization of educational access. Additionally, community and parental involvement in supporting their children's learning is also crucial to creating an environment that fosters the success of ODE.

In conclusion, it is important to remember that ODE is not just an alternative form of education but also an urgent need to reduce the educational gap in Indonesia. It is critical that the government and other stakeholders take immediate action by implementing more inclusive policies, improving digital infrastructure, and providing adequate training for teachers. Only through a holistic and sustainable approach can ODE achieve its goal of reducing the educational gap between urban and rural areas and improving the quality of education for all students in Indonesia.

References

Belawati, T., Daryanto, S., & Sumitro, T. (2020). The role of open and distance education in bridging the learning gap in Indonesia. *Distance Education, 41*(3), 341–358. https://doi.org/10.1080/01587919.2020.1770872

Fagbamigbe, A. F., Oyinlola, F. F., Morakinyo, O. M., Adebowale, A. S., Fagbamigbe, O. S., & Uthman, A. O. (2021). Mind the gap: what explains the rural-nonrural inequality in diarrhoea among under-five children in low and medium-income countries? A decomposition analysis. *BMC Public Health*, 21(1), 575. https://doi.org/10.1186/s12889-021-10615-0

Heryanto, O. (2021). The utilization of information technology as a learning medium for teachers in distance education (PJJ) during COVID-19. *Avant Garde, 9*(2), 1371–1380. https://doi.org/10.36080/ag.v9i2.1371

Hornok, Cecília & Raeskyesa, Dewa Gede Sidan. (2023). Economic zones and local income inequality: Evidence from Indonesia. *The Journal of Economic Inequality*. 22. 1-32. https://doi.org/10.1007/s10888-023-09581-x

Irhamsyah, T. (2023). *General policy on the implementation of merdeka belajar*. Jurnal At-Tabayyun. Retrieved from https://api.semanticscholar.org/CorpusID:267131840

Jusas, Vacius & Butkienė, Rita & Venčkauskas, Algimantas & Burbaite, Renata & Gudoniene, Daina & Grigaliunas, Sarunas & Andone, Diana. (2021). Models for administration to ensure the successful transition to distance learning during the pandemic. *Sustainability*. 13. https://doi.org/10.3390/su13094751

Kurnia, A., Sari, R., & Dewi, M. P. (2023). Literacy data on urban vs. rural education gap in Indonesia. *International Journal of Education and Practice, 14*(2), 102–115. https://doi.org/10.24956/ijep.14.2.2023

Martin, Florence & Wu, Tong & Wan, Liyong & Xie, Kui. (2022). A meta-analysis on the community of inquiry presences and learning outcomes in online and blended learning environments. *Online Learning*. 26. 325-359. https://doi.org/10.24059/olj.v26i1.2604

Ministry of Education Indonesia. (2020). *Policy report on learning from home during the COVID-19 pandemic*. Retrieved September 30, 2025, from https://www.kemdikbud.go.id/main/

Mizal, B., Basith, R. I., & Tathahira, T. (2021). Critical thinking through distance learning: An analysis of Indonesian Open University. *International Journal of Education, Language, and Religion, 3*(1), 25–38. https://doi.org/10.35308/ijelr.v3i1.3667

Rahayu, W., Putra, M. D. K., Faturochman, Meiliasari, Sulaeman, E., & Koul, R. B. (2022). Development and validation of Online Classroom Learning Environment Inventory (OCLEI): The case of

Indonesia during the COVID-19 pandemic. *Learning Environments Research*, 25(1), 97–113. https://doi.org/10.1007/s10984-021-09352-3

Rahmi Rivalina, R. R. (2013). Why open and distance education (PTJJ)? *Jurnal Teknodik, 1*(1), 93–107. https://doi.org/10.32550/teknodik.v0i0.93

Salwa, A. (2021). Adaptation of the UT online tutorial learning model during the Covid-19 emergency learning period. *Jurnal Pendidikan Terbuka Dan Jarak Jauh*. https://doi.org/10.33830/ptjj.v21i1.1159.2020

Savitri, E. N., Amaliap, A. V., Listiaji, P., Atunnisa, R., & Fibriana, F. (2023). Virtual experiments of DNA isolation and hybridoma techniques to improve students' critical thinking ability. *Biosaintifika Journal of Biology & Biology Education, 15*(1), 15–26. https://doi.org/10.15294/biosaintifika.v15i1.43847

Tsara Ayuninggati, R. R., Sihotang, S. V., Mustopa, M. I., & Wahyuningsih, W. (2023). Research on students' extracurricular interest in music during remote learning in the 4.0 era. *Jurnal Mentari Manajemen Pendidikan Dan Teknologi Informasi*. https://doi.org/10.33050/mentari.v1i2.245

United Nations Educational, Scientific and Cultural Organization. (2022). *The impact of digital learning in education: Global trends and lessons learned*. United Nations Educational, Scientific and Cultural Organization. Retrieved September 30, 2025, from https://www.unesco.org/en/digital-learning-impact

Wahid, R., Ahmi, A., & Alam, A. F. (2020). Growth and collaboration in massive open online courses: A bibliometric analysis. *The International Review of Research in Open and Distributed Learning*, 21(4), 292–322. https://doi.org/10.19173/irrodl.v21i4.4693

World Bank. (2021). *The role of education technology in addressing education gaps in developing countries*. Retrieved September 30, 2025, from https://www.worldbank.org/en/topic/edutech

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Analyzing Middle School Students’ Distance Education Experiences in COVID-19 via Sentiment Analysis and Topic Modeling

Ekrem Bahçekapılı[1], Bülent Kandemir[2], and Elif Baykal Kablan[3]
[1]Department of Management Information Systems, Karadeniz Technical University; [2]Department of Educational Sciences, Ordu University; [3]Department of Software Engineering, Karadeniz Technical University

Abstract

This study investigated middle school students’ experiences with emergency remote education during the COVID-19 pandemic using natural language processing (NLP), sentiment analysis, and topic modeling techniques. A total of 2,739 valid responses from Turkish students (ages 9–15) were collected through open-ended survey questions regarding the perceived advantages and disadvantages of distance learning. Sentiment classification was performed using a semi-supervised machine learning approach, combining TF-IDF, Word2Vec, and FastText vectorization with five classification algorithms. The TF-IDF + support vector machines (SVM) combination yielded the highest performance (F1 = 0.85). Results show a total of 1,867 positive and 2,542 negative opinions, indicating that students generally adopted a more critical view of distance education. To explore the thematic structure of opinions, topic modeling was applied with six topics. Positive sentiments clustered around themes such as educational continuity, health protection, time savings, flexible scheduling, self-regulated learning, and digital literacy. Negative sentiments were dominated by themes including limited interaction, screen fatigue, perceived low quality, technical barriers, and structural inequalities. Findings suggest that while students appreciated the safety and flexibility of remote learning, they also faced significant pedagogical, physical, and technological challenges. The study contributes methodologically by demonstrating the effectiveness of AI-based text analysis and offers practical implications for designing more equitable and student-centered digital education models. These results underscore the importance of integrating NLP and machine learning tools into educational research to uncover deeper insights from student-generated content at scale.

Keywords: sentiment analysis, topic modeling, student emotion, student perception, COVID-19, NLP

Introduction

The COVID-19 pandemic has affected education systems worldwide, producing marked changes for K–12 learners. Reliant on structured, face-to-face classrooms and characterized by age-specific developmental needs, this cohort encountered serious challenges (Adams et al., 2024; Tomaszewski et al., 2023). The abrupt shift to remote instruction exposed critical infrastructure gaps especially in rural and socioeconomically disadvantaged regions, thereby amplifying preexisting inequalities. As the digital divide widened, meaningful participation and learning opportunities became increasingly constrained (Dhawan, 2020; Hurling et al., 2024). To interpret these effects, we first distinguish emergency remote teaching (ERT) from pedagogically designed online learning: ERT prioritizes continuity through rapid, temporary solutions under crisis constraints. This distinction frames our interpretation of students' experiences during the crisis.

Remote instruction, made compulsory by the pandemic, generated both benefits and obstacles. Digital platforms afforded flexibility and continuity in learning, showing potential for enhancing engagement and achievement (Lo et al., 2023). Yet the rapid migration online also revealed significant shortfalls in meeting individual learning needs (Diz-Otero et al., 2023). Students reported technical difficulties, loss of motivation, and a lack of collaborative environments, all of which contributed to learning deficits (Sandvik et al., 2024). At the middle school stage, these design and implementation constraints intersect with developmental needs. In particular, early adolescents are reorganizing motivation, belonging, and self-efficacy (Eccles & Roeser, 2011). Consistent with this, from a self-determination perspective, threats to autonomy, competence, and relatedness predict lower engagement in online settings (Ryan & Deci, 2000).

The pandemic likewise exerted profound social and emotional pressures on middle schoolers. Loneliness and mental health issues increased globally (Geulayov et al., 2024). Social media sustained peer communication but also spread misinformation and panic (Radwan et al., 2020). Interruptions to peer and teacher relationships elevated stress levels, especially at the middle school stage (Albert, 2024). The crisis highlighted the centrality of digital infrastructure for equitable access, especially in economically disadvantaged contexts (Hurling et al., 2024). From a Community of Inquiry (CoI) perspective, disruptions to social and teaching presence would have been expected to erode belonging and perceived support, magnifying stress and disengagement in early adolescence (Garrison, 2016; Garrison et al., 2000). Beyond connectivity, second- and third-level aspects of digital divide skills, and the translation of use into academic benefit, which varies widely across contexts, help explain uneven socioemotional outcomes under remote learning (van Dijk, 2006). These dynamics are likely to be pronounced in early adolescence.

Artificial intelligence (AI) technologies have emerged as powerful tools for making learning more effective and personalized. AI-enhanced applications analyze students' emotional states and adapt content accordingly (Lin & Chen, 2024). Facial-recognition and affect-detection systems have enabled real-time monitoring of learners' emotions (Fang et al., 2023). In addition, natural language processing (NLP) and machine learning techniques identify cognitive and affective barriers, offering timely interventions (Martínez-Comesaña et al., 2023). Personalized feedback has been shown to bolster motivation and emotional resilience (An et al., 2023), although concerns persist about excessive surveillance and equity (Lin & Chen, 2024). Aligned with the CoI framework and transactional distance, analytics-enabled, timely feedback can strengthen teaching and social presences and reduce structure–dialogue imbalances that fuel disengagement in remote contexts (Garrison, 2016; Moore, 1993). In practice, dashboarded prompts and micro-goal check-ins can scaffold self-regulated learning processes involving planning, monitoring, and self-evaluation, particularly salient for early adolescents in flexible online settings (Pintrich, 2004; Zimmerman, 2002). Large-scale NLP analysis therefore offers an innovative alternative to traditional survey-based studies.

Against this backdrop, this study aimed to examine middle school students' perceptions of emergency remote teaching by (a) quantifying their positive versus negative sentiment and (b) extracting latent thematic structures through advanced NLP techniques, offering a scalable alternative to traditional survey-based approaches, thereby delivering theoretical, methodological, and practical contributions. As noted above, because these data were generated under ERT, we interpreted all patterns in light of crisis-driven, rapidly deployed solutions; accordingly, claims have been bounded to ERT conditions rather than fully designed online learning.

We framed our research with two questions and two sub questions:

RQ1. How are middle school students' perceptions of remote education distributed across positive and negative sentiment categories?

RQ2. What latent themes underlie middle school students' perceptions of distance education?

RQ2a. Which themes emerge from students' positive opinions?

RQ2b. Which themes emerge from students' negative opinions?

Literature Review

AI now supports learning on two fronts; it serves students directly through adaptive tutors and affect-aware dashboards, while also supplying teachers and instructional designers with indirect, data-rich insights into learners' emotions and thought processes (Lin et al., 2024; Uçar et al., 2024). NLP and machine learning pipelines analyze large corpora of student text, enabling educators to identify individual needs and tailor instructional strategies. For example, large language models (LLMs) can detect affective states in real time and summarize formative feedback, thus reducing teachers' analytical workload (Xu et al., 2024). Techniques such as sentiment analysis and text mining further feed design analytics that help instructors construct more personalized and responsive learning environments (Al Husaeni et al., 2022; Bittencourt et al., 2024). However, most prior studies rely on small-sample surveys or focus on university settings; as a result, we still know little about how middle school pupils felt during ERT. Filling that gap was the central aim of RQ1 in this study.

Recent K–12 research has shown that NLP can act both as a diagnostic lens and as an adaptive support in online or hybrid learning contexts. Using network analysis combined with topic modelling, Xing et al. (2025) found that "extra-periphery" participants in a large asynchronous mathematics forum could achieve the highest mathematical-literacy scores when interaction quality was high. At the message level, sentiment and tone analysis applied to middle school teacher feedback by Baral et al. (2023) revealed patterns that shape student motivation, while an emoji-based interface designed by Zarkadoulas and Virvou (2024) captured pupils' emotional states and highlighted gender differences in expression. From the teacher perspective, an NLP-driven virtual facilitator for professional development produced significant gains in student performance in a randomized trial reported by Copur-Gencturk et al. (2024), aligning with the "Turing Teacher" attributes outlined by Pelaez et al.

(2022). Inclusive angles are also emerging: a mixed qualitative–sentiment study by Tzimiris et al. (2023) documented layered psychological and technical barriers faced by students with functional diversity during ERT. Finally, automated text analysis developed by Žitnik and Smith (2024) flagged off-topic posts in fourth-grade book-club discussions with 90% accuracy, pointing toward real-time analytics that could keep young learners on track.

Various AI techniques are employed to analyze student opinions in remote education. One of the most prominent NLP approaches is topic modelling, which uncovers latent themes within a text. For example, Mujahid et al. (2021) used latent Dirichlet allocation (LDA) to analyse student comments on Twitter concerning online learning during the COVID-19 period. Their study revealed the main challenges of online learning, particularly infrastructure deficiencies and lack of technical support. However, LDA coherence degrades sharply with short, grammar-sparse comments, typical of K–12 data (Gallagher et al., 2017). In another study, researchers conducted a large-scale thematic analysis of social media data, allowing them to map broad perspectives on online education (Mishra et al., 2021). Waheeb et al. (2022) likewise analyzed social media datasets from the pandemic period, combining LDA with ontology-based approaches to capture both thematic and sentiment trends. Information-theoretic models such as correlation explanation (CorEx) achieve higher semantic consistency on short documents and allow researchers to anchor domain-specific keywords (Gallagher et al., 2017), yet have rarely been applied to open-ended responses from middle school students, particularly in non-English contexts.

Another major application of machine learning and NLP is sentiment analysis, which seeks to identify emotions embedded in text. Akhmedov et al. (2021) employed the joint sentiment topic (JST) model to integrate thematic and sentiment analysis, enabling a word-level examination of student opinions; this approach provided a more comprehensive understanding of perceptions related to online education. Nevertheless, JST requires large, labelled corpora resources that remain scarce for Turkish middle school datasets. Lin et al. (2024) used large databases to apply sentiment-analysis techniques that identified the key drivers of student satisfaction and negative emotions in online learning. In a separate investigation, researchers integrated machine learning and deep learning methods to classify emotions in online-course reviews with high accuracy. Such studies offer valuable insights for developing student-centered instructional strategies (Onan, 2021).

In sum, prior research has established the value of topic and sentiment modelling but left unanswered how middle school learners themselves articulate advantages and disadvantages of ERT in their own language; addressing this gap constituted the focus of RQ2, RQ2a, and RQ2b in our study.

Methodology

Convenience sampling allows researchers to obtain data quickly and easily from suitable sources. This method is particularly useful when probability sampling techniques are not feasible, enabling researchers to focus on subjects that are accessible and relevant to the study's purpose (Marshall, 1996). Although convenience sampling may introduce bias by limiting the representativeness of the sample, it was considered appropriate in this study due to the extraordinary circumstances of the COVID-19 pandemic. During this period, accessibility and timeliness were critical, and voluntary participation from available students provided a practical means to capture authentic perceptions of ERT.

The study's participants consisted of middle school students (ages 9–15, mean age = 11.95, *SD* = 1.2; a critical developmental stage and a relatively less-studied group) enrolled in the 2019–2020 academic year in Türkiye. During this period, these students were educated entirely through remote learning methods (live classes and asynchronous learning resources) for the first time. Following ethical committee approval (E-81614018-000-330), necessary information was conveyed to the students' families through the Ministry of National Education, and a survey was distributed via an online form. Data collection was conducted anonymously. A total of 2,890 students voluntarily participated in the study. In the preliminary analysis, responses with missing data, participants who answered all questions identically, and consecutive identical responses were excluded. A total of 2,739 valid responses (female = 56.7%, male = 43.3%) were collected and accepted for further data analysis.

Data were obtained through an online form. The data collection instrument included two open-ended questions. The first question was, "What do you think are the advantages of distance education?" and the second was, "What do you think are the disadvantages of distance education?"

Data Analysis

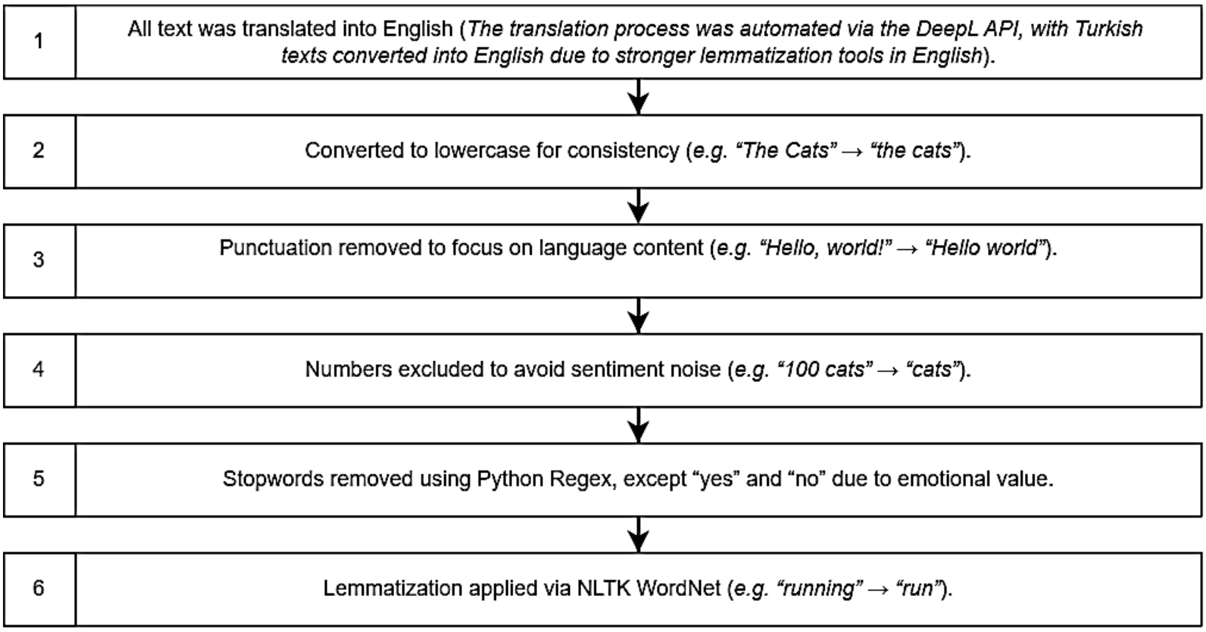
Research data were analyzed using NLP and machine learning. After text preprocessing, sentiment analysis employed TF-IDF, Word2Vec, and FastText with classifiers to estimate student emotions. Topic modeling explored opinions on distance learning. The process is explained in detail under the following subheadings.

Data Preprocessing

Data preprocessing involves a series of operations performed on a dataset prior to the sentiment extraction phase. Our preprocessing steps, illustrated in Figure 1, aimed to prepare the data for model training and testing, ensuring the model would be both understandable and reliable.

Figure 1

Data Preprocessing Steps



Sentiment Analysis

Sentiment analysis is a natural language processing technique used to determine the emotional content within texts (Devika et al., 2016). It aims to ascertain whether a piece of text contains positive, negative, or neutral sentiments. In this work, we tackled sentiment classification with a dataset of 4,439 user-generated opinions, of which only 1,961 instances were manually labeled by experts (positive: 858 vs. negative: 1,103). Given the scarcity of labeled data, we adopted a semi-supervised learning approach to leverage the remaining unlabeled examples.

After preprocessing, phrase vectorization was applied to convert text into numerical formats suitable for machine learning. Common methods include TF-IDF, Word2Vec, and FastText. TF-IDF generates vectors based on word frequency across documents. These methods do not consider word order or meaning—they only focus on word frequencies (Khanna & Coumans, 2023). Word2Vec, based on neural networks, also reflects the context and meaning of words into vectors (Garcia, 2021). FastText, on the other hand, converts words into vectors and also uses the sub-parts (n-grams) within the word (Patil et al., 2023).

After vectorization, the model was trained using five classification algorithms: naive Bayes, logistic regression, support vector machines (SVM), decision tree, and random forest (RF). Combined with three vectorization methods, this means 13 models were tested. These algorithms aim to accurately and efficiently categorize data into predefined classes (Alweshah, 2019). Classification algorithms are used in a wide range of applications, from text classification to medical diagnosis, and from image recognition to financial matters. Traditional machine learning models, such as k-nearest neighbors (k-NN), naive Bayes, SVM, decision trees, and RF, have historically been successful in sentiment analysis (Rodríguez-Ibánez et al., 2023). These models rely on statistical algorithms and predefined features to classify text sentiments such as positive, negative, or neutral.

In the following section, the classification algorithms used in this study are discussed in detail.

Topic Modeling

After the sentiment analysis, unlabeled sentiments were predicted with the best model, and the opinions in two categories (advantage and disadvantage of distance education) were analyzed separately with topic modeling. Topic modeling is a NLP method that aims to analyze large amounts of text data to uncover hidden thematic patterns (Gerlach et al., 2018). While topic modeling provides a systematic and scalable way to identify themes in large datasets, it differs from traditional qualitative thematic analysis, which relies on manual coding and interpretive validation. In this study, themes were derived directly from model outputs and researcher interpretation, without triangulation from additional qualitative methods. This is a particularly useful way of analyzing large amounts of student opinion. Topic modeling has been successfully applied in many areas, and topics in large datasets have been revealed (Ozyurt & Ayaz, 2022).

While determining the ideal number of topics and the modeling algorithms for topic modeling analysis, various metrics were used in addition to the interpretability-explainability of the topics that emerged. In this study, while defining the ideal model, LDA and CorEx topic modeling algorithms were tested by calculating coherence (C_V), homogeneity, precision (micro), recall (macro), and F1 score values. The algorithms were run separately to create topics between one and twenty and graphed (Figures 2 and 3), and both the algorithm and the number of topics were decided in this way. While consistency measures the degree of semantic similarity between high-scoring words on a topic, it is a frequently used metric for the balance between human interpretability and computational efficiency (Mifrah & Benlahmar, 2020). Homogeneity evaluates the similarity of documents assigned to the same topic and the harmony of topics in different documents and is very important to ensure that topics are not overly broad or ambiguous (Amaro & Bacao, 2024). Micro precision assesses model accuracy per topic, macro recall measures capturing all relevant terms, and the F1 score balances both measures (Virtanen & Girolami, 2019).

There are many algorithms developed for this method, notably LDA. Although LDA is commonly preferred (Vayansky & Kumar, 2020), it is noted that it does not always provide the best result. It is particularly insufficient in short texts (Özyurt & Akcayol, 2020) and in revealing the relationships between words. Correlation explanation (CorEx) is also used for topic modeling. CorEx offers a compelling alternative to LDA by leveraging an information-theoretic framework that learns maximally informative topics without the need for detailed assumptions or complex hyperparameter specification (Gallagher et al., 2017). This approach not only simplifies model complexity but also facilitates the incorporation of human input through anchor words, enhancing topic separability and representation with minimal intervention.

Results

In this section, results are organized according to the research questions. The findings related to the first research question, being foundational, were prioritized and used to inform the analysis of the findings for the second research question.

RQ1: How are middle school students' perceptions of remote education distributed across positive and negative sentiment?

The model performances of the vectorization and classification algorithm pairs are presented in Table 1. When the performance data in Table 1 is examined, the SVM model trained with TF-IDF feature extraction stands out as the most successful method in terms of both accuracy (85.2%) and F1 score (85.1%).

Table 1

Model Performances of Vectorization and Classification Algorithm Pairs

Vectorization	Model	Accuracy	F1 Score
TF-IDF	Naive Bayes	0.811545	0.807016
TF-IDF	Logistic regression	0.823430	0.820996
TF-IDF	**SVM**	**0.852292**	**0.851196**
TF-IDF	Decision tree	0.784380	0.785191
TF-IDF	Random forest	0.825127	0.825088
Word2Vec	Logistic regression	0.723260	0.718766
Word2Vec	SVM	0.735144	0.727256
Word2Vec	Decision tree	0.650255	0.650678
Word2Vec	Random forest	0.760611	0.755425

Vectorization	Model	Accuracy	F1 Score
FastText	Logistic regression	0.672326	0.648586
FastText	SVM	0.563667	0.406379
FastText	Decision tree	0.602716	0.601008
FastText	Random forest	0.706282	0.696931

Note. The most successful pairing—TF-IDF and SVM—is shown in bold. SVM = support vector machines.

When the labeled data was estimated with the machine learning model created with this pair, it was observed that a significant portion of students stated there was no advantage to the advantage question in the preprocessing step, and similarly, they gave answers to the disadvantage question stating there was an advantage. The results are consistent with our observation; since students expressed their opinions without focusing on the question heading, some positive responses under "disadvantage" aligned with negative responses to "advantage," and vice versa.

A summary presented in Table 2 gives the total number of opinions, positive comments, and negative comments for each question. Although a total of 2,217 opinions were stated for the advantage, 509 of those were negative, i.e., saying there was no advantage. Similarly, 170 of the 2,222 total disadvantage opinions did not contain positive, i.e., disadvantage-oriented opinions.

In sum, positive opinion was 1,878 and negative opinion was 2,561.

Table 2

Summary of Total, Positive, and Negative Opinions for Each Question

Question	Total comments *N*	Positive comments *N*	Negative comments *N*
What is the advantage of distance learning?	2,217	1,708	509
What is the disadvantage of distance learning?	2,222	170	2,052

The results indicate that many students questioned or did not perceive any advantages of distance education, reflecting a skeptical view of its benefits and a clearer, more negative attitude toward its drawbacks.

RQ2: What latent themes underlie middle school students' perceptions of distance education?

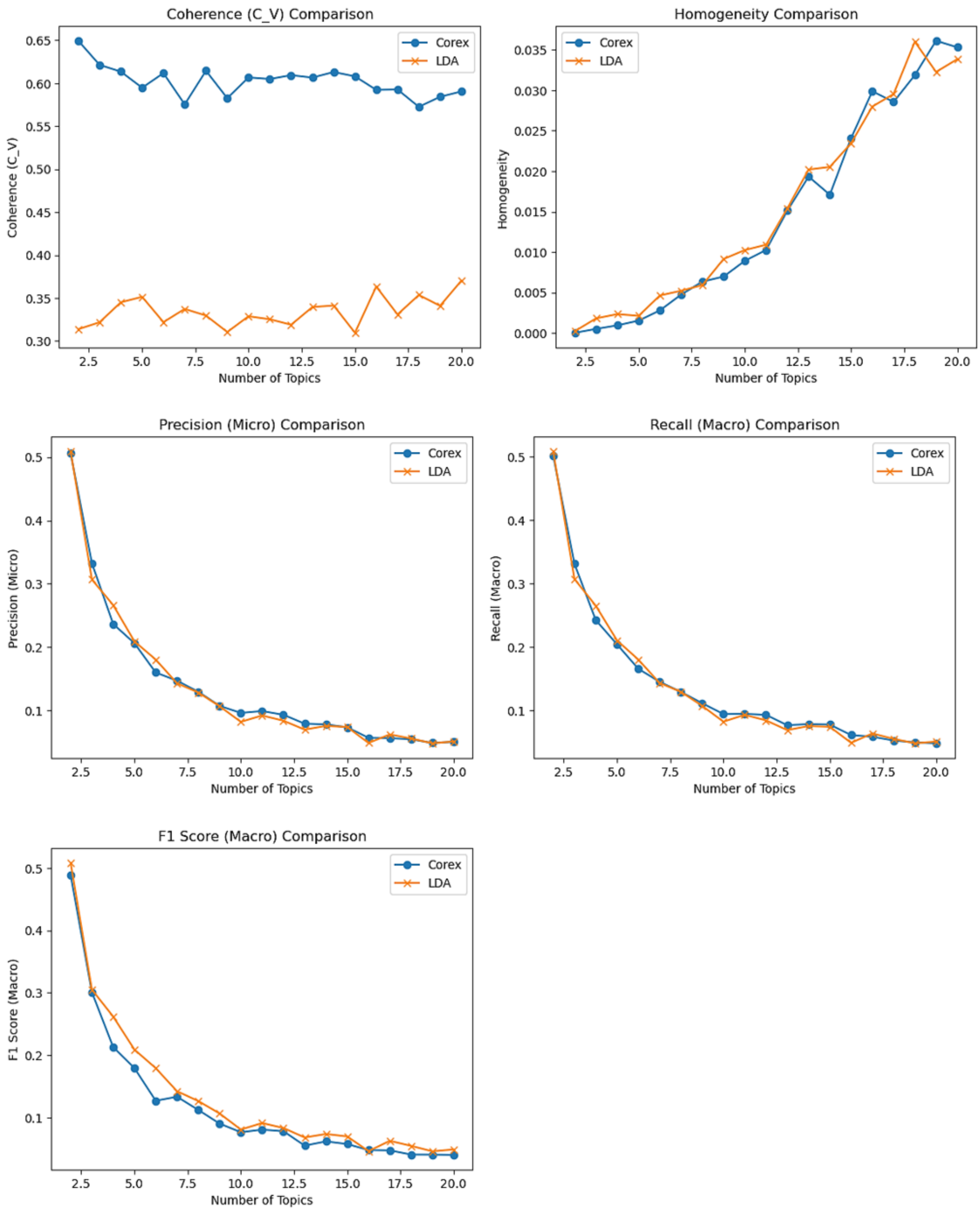
CorEx and LDA were compared to identify latent topics, evaluated by coherence, homogeneity, precision, recall, and F1 (see figures 2 and 3). Detailed findings are explained in the following sections.

RQ2a: Which themes emerge from students' positive opinions?

Figure 2 shows CorEx achieved optimal coherence (~0.60) with six to eight topics, exceeding the meaningful threshold of 0.40, unlike LDA, which peaked early and dropped sharply (~0.33). While LDA showed slightly better F1 at two topics, its interpretability suffered. CorEx maintained better F1 at six topics (~0.13), balancing accuracy and readability. At high topic counts, both models show fragmentation, lowering insight value. CorEx-6 offered thematically distinct, interpretable topics, aligning with qualitative goals. Thus, CorEx with six topics was chosen as optimal, and results are detailed in Table 3.

Figure 2

CorEx and LDA Performance Metrics for Positive Comments



Note. LDA = latent dirichlet allocation.

Table 3

Topics and Keywords for Positive Comments

Topic	Theme	Keywords
1	Continuity of education and health protection	education, distance, away, stay, continue, pandemic, period, technological, process, receive, class, close, virus, thank, tool
2	Time savings and physical comfort	advantage, think, school, don't, wear, tire, good, opportunity, sick, problem, road, situation, uniform, great, return
3	Noise-free, focused learning space	teacher, classroom, noise, sound, write, switch, environment, say, lesson, distract, voice, hear, turn, teach, image
4	Flexible scheduling and active participation	question, morning, ask, lecture, early, listen, want, solve, answer, test, read, book, comfortably, later, clothes
5	Self-regulated learning and information access	study, learn, time, information, home, like, health, little, access, spend, knowledge, covid, protect, work
6	Digital literacy and ease of learning	use, technology, make, eat, homework, way, efficiently, improve, assign, sense, hand, there's, share, sleep, easier

Using a 6-theme CorEx model (C_v = 0.446), the advantages of immediate distance learning for lower secondary students were clustered around six distinct themes listed in Table 3. Each theme is described below by triangulating the model's most important keywords with representative student quotes.

Theme 1: Continuity of Education and Health Protection

Statements such as "distance education [is] the best option in the conditions we live in" and "continue education without feeling risk due [to] the pandemic" illustrate that students framed remote teaching primarily as a means of maintaining academic progress while safeguarding health. The assured progression of the curriculum reduced uncertainty and fostered a sense of security.

Theme 2: Time Savings and Physical Comfort

Phrases such as "no need to wear [a] uniform" and "save time … no waste [of] time on the road" highlight gains in time management and bodily comfort. The removal of travel and dress codes redirected both physical and cognitive resources toward learning activities.

Theme 3: Noise-Free, Focused Learning Space

Excerpts such as "no noise environment—think calmly" show that the virtual classroom acted as a noise filter, improving audibility and concentration. Microphones and headsets facilitated clearer teacher–student communication and reduced peer distractions.

Theme 4: Flexible Scheduling and Active Participation

Students praised "not getting up early in the morning" and being able to "ask questions and get answers whenever [they] want," signaling that temporal flexibility empowered them to self-pace and actively engage with course content, aligning with principles of learner autonomy.

Theme 5: Self-Regulated Learning and Information Access

Comments such as "able to study at home and research" indicate strengthened self-regulation and unfettered access to digital information. The quieter home environment enabled deeper cognitive processing and individualized study routines.

Theme 6: Digital Literacy and Ease of Learning

Expressions including "learn to use technology tools" reveal that remote teaching served as a scaffold for digital literacy, while also enabling homework submission and making everyday needs (e.g., eating, breaks) more convenient, thus lowering affective and physiological barriers to learning.

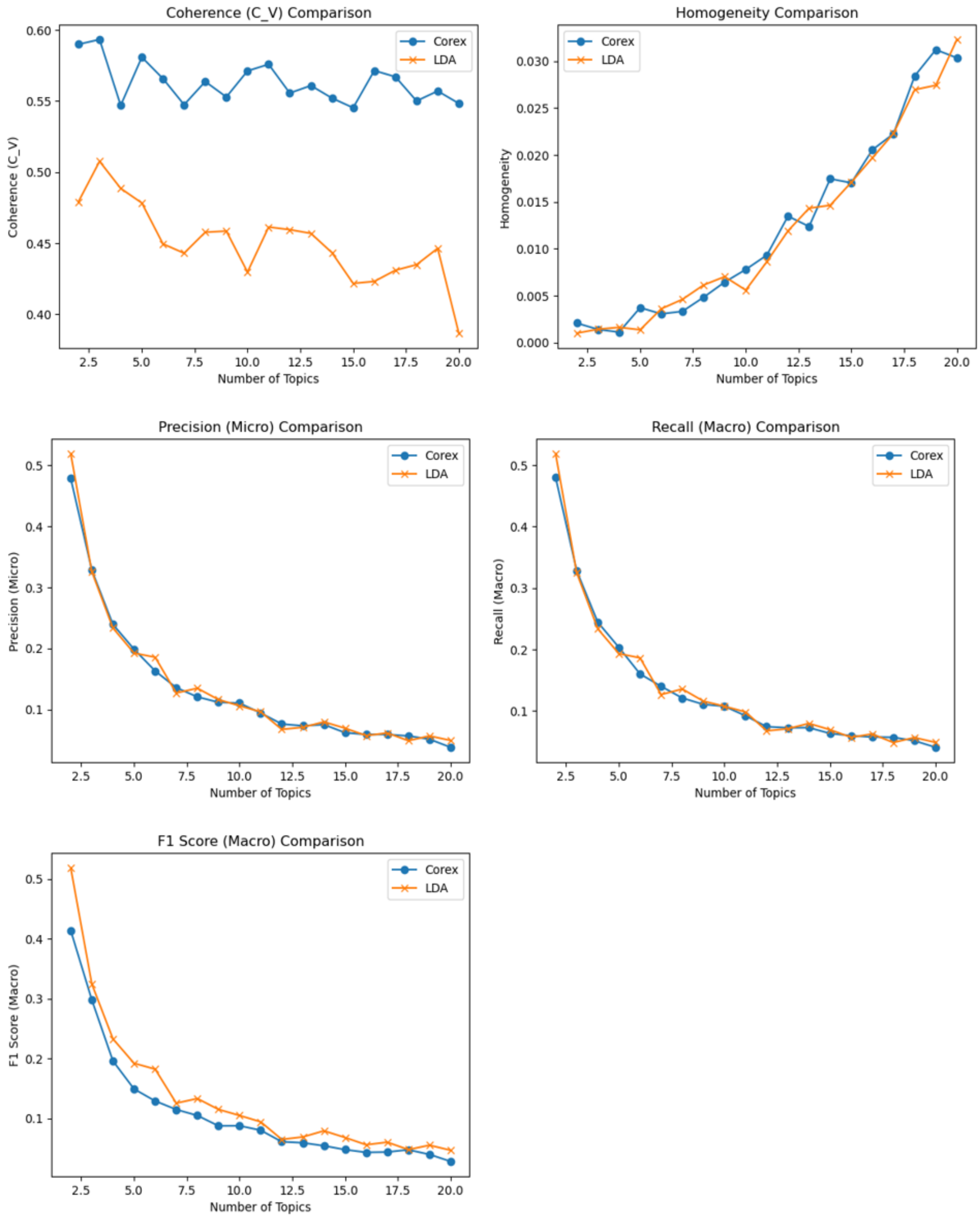
In summary, K–12 students in COVID-19 remote education saw it as safe, ensuring continuity, time savings, quietness, schedule flexibility, self-regulated learning, digital literacy, personalized education, and autonomy.

RQ2b: Which themes emerge from students' negative opinions?

For negative perceptions, Figure 3 confirms CorEx's superiority in interpretability. Its CV coherence remained high (~0.57–0.59) and stable between six to eight topics, outperforming LDA, which dropped sharply after k = 2 and stabilized near 0.44, with a practically significant 0.13-point gap. Although both models peaked in F1 at k = 2, the resulting topics were overly broad and lacked analytic depth. At six topics, however, CorEx maintained a double-digit macro-F1 (~0.12) with high coherence, whereas LDA's performance declined. Homogeneity rose beyond 10 topics, but recall collapsed (< 0.08), indicating excessive topic fragmentation. Thus, CorEx-6 was chosen to balance interpretability with classification performance for analyzing students' negative views.

Figure 3

CorEx and LDA Performance Metrics for Negative Comments



Note. LDA = latent dirichlet allocation.

Using the 6-theme CorEx solution (C_v = 0.396), students' negative perceptions of emergency distance learning coalesced into six interpretable disadvantage themes (Table 4). The theme labels were created by triangulating the most important keywords of the model with representative quotes and are explained in the following sections.

Table 4

Table of Topics and Keywords for Negative Comments

Topic	Theme	Keywords
1	Limited interaction and question resolution	question, ask, lesson, want, answer, enter, try, homework, live, attention, write, speak, difficult, teach, short
2	Screen fatigue and health strain	eye, look, screen, tire, health, hurt, deteriorate, computer, time, constantly, spend, long, affect, head, lot
3	Perceived low quality and inefficiency	education, distance, think, quality, learn, definitely, high, good, efficient, negative, opinion, fact, period, home, device
4	Audio/technical barriers to comprehension	teacher, voice, say, open, sound, work, hear, come, doesn't, explain, feel, throw, microphone, environment, classroom
5	Connectivity loss and cognitive breaks	understand, school, don't, subject, like, away, better, stay, fully, connection, issue, freeze, disconnection, face, productive
6	Access and participation constraints	attend, class, connect, low, people, participation, opportunity, attendance, make, unable, limit, end, wrong, broken, lecture

Theme 1: Limited Interaction and Question Resolution

Excerpts such as "sometimes may not be able to enter [the virtual classroom], may not get the answer we want" show that synchronous platforms often fail to replicate classroom dialogue, leaving questions unresolved and diminishing perceived teacher support.

Theme 2: Screen Fatigue and Health Strain

Students repeatedly mentioned "eye hurt … front of computer constantly" and "neck, back ache," indicating that the digital format imposed physiological costs that accumulated over long sessions.

Theme 3: Perceived Low Quality and Inefficiency

Phrases including "cannot get efficiency … distance education insufficient" reveal a widespread belief that remote delivery diluted instructional quality, especially for numeracy-heavy subjects.

Theme 4: Audio/Technical Barriers to Comprehension

Students complained that "sound goes, cannot hear teacher" and that overlapping microphones confused discourse, reflecting signal-to-noise problems well documented in other synchronous e-learning studies.

Theme 5: Connectivity Loss and Cognitive Breaks

Reports of "freeze, disconnection, throw us out" illustrate that unstable networks fractured cognitive continuity, forcing learners to reconstruct content gaps and eroding learning efficacy.

Theme 6: Access and Participation Constraints

Statements such as "no technological device … cannot attend" and "low participation" emphasize structural inequities: limited hardware, bandwidth, and moderated turn-taking all curtailed active engagement.

In summary, K–12 students in remote education during COVID-19 faced limited teacher interaction, delayed feedback, perceived low-quality instruction, especially in math, physical discomfort, technical disruptions, and inequities.

Discussion

This study examined middle school students' distance learning experiences and their emotional responses during the COVID-19 pandemic through sentiment analysis and topic modeling techniques. We found that students' comments generally contained negative sentiment. Additionally, students emphasized certain advantages of distance learning, such as ensuring continuity of learning while protecting individual health, saving time, and enabling flexible scheduling. However, student perspectives in general drew attention to disadvantages of distance learning, such as limited interaction and problem-solving opportunities, screen fatigue, and negative effects on health. Moreover, they emphasized the kind of structural barriers that disrupt the learning process, including technical problems, connection interruptions, and access issues. Reduced feedback and collaboration maps to weaker CoI teaching and social presences, while rigid formats and interruptions raise transactional distance; for early adolescents, limited self-regulated learning (SRL) scaffolds magnify disengagement (Garrison, 2016; Garrison et al., 2000; Moore, 1993; Pintrich, 2004; Zimmerman, 2002).

The sentiment analysis results show that students predominantly focused on negative characteristics may be related to the adverse situation created by extraordinary changes in their lives, such as the COVID-19 pandemic and the resultant widespread adoption of distance learning for students accustomed to face-to-face education. Indeed, studies have demonstrated that students experienced problems with engagement and motivation in distance learning courses during this period (An et al., 2023). Nevertheless, positive views were also found at a considerable level. The fact that distance learning served as the fundamental method for educational continuity during these challenging times may have been the primary basis for these views (Leech et al., 2022). Because patterns arose under ERT, implications would be crisis-contingent rather than general to fully designed online learning.

The topic modeling results reflecting positive views demonstrate that students experiencing distance learning for the first time at the K–12 level evaluated this process as a safe solution that ensured uninterrupted continuation of education under pandemic conditions, while simultaneously discovering unexpected advantages such as time savings, noise-free learning environments, and flexible scheduling. This finding aligns with previous studies that have emphasized the value of digital solutions in ensuring learning continuity during crisis periods (Datta & Nwankpa, 2021). Factors such as the elimination of commuting time, the absence of uniform requirements, and learning in the home environment enabled students to experience the educational process more comfortably, as was shown also in Tuguic & Bilan (2023). Having a quiet learning environment free from distracting elements and the opportunity to follow courses with a flexible schedule suitable to their own learning pace were positive attributes of distance learning for the students. However, these findings differ from other literature. For example, there are studies that have emphasized that some students in distance learning encounter more distracting elements since they are usually in home environments, and teachers experience difficulties in ensuring student participation in lessons (Kadirhan & Sat, 2024). This difference may have emerged because the teachers in our study were able to reduce problems such as natural noise by using technological tools. Indeed, studies have reported that noise in face-to-face classroom environments affects student attention and motivation (Caviola et al., 2021). The opportunity for self-regulated learning and easy access to information, along with high levels of digital literacy and the simplicity of the learning process, were highlighted by some students who said they had the opportunity to develop their digital skills (Naidu, 2019). Still, such advantages may be unevenly distributed; skills in dealing with second- and third-level digital-divide factors and the translation of use into academic benefit can shape who realizes these gains (van Dijk, 2006).

On the other hand, negative views were quite diverse and could be classified at structural, pedagogical, and individual levels. Students specifically indicated that synchronous platforms could not reflect classroom dialogue due to limited interaction. Questions remained unanswered, and teacher support decreased. This situation demonstrates that the limited student-teacher interaction in distance learning can negatively affect learning (Dokuchyna, 2023). In CoI terms, fragile social and teaching presences explain lower belonging and participation; in theory of transactional distance (TDT) terms, low dialogue relative to structure widens transactional distance, risking misunderstanding and withdrawal (Garrison, 2016; Moore, 1993). Physical discomforts such as eye fatigue, headaches, and neck and back pain caused by prolonged screen use among students emerged as a significant problem. This situation indicates that online education needs to be redesigned from ergonomic and health perspectives (Upadhyay et al., 2021). Students’ widespread beliefs that distance learning is particularly inefficient in mathematics and other courses with a lot of numeric content and reduces teaching quality emerged as a significant problem. These views support the necessity of strengthening student and teacher support to improve quality in distance learning, especially at the K–12 level (Martin et al., 2022). Technical problems, particularly audio quality issues, chaos created by overlapping microphones, and audio signal problems resulting in teachers being inaudible, make it difficult for students to understand lesson content; while connections drop and freeze, and platform disconnections fragment cognitive continuity, there is a need to reconstruct content gaps and strengthen learning effectiveness (Nowak & Watt, 2022). Finally, structural inequalities such as device shortages, insufficient bandwidth, and sequential speaking rights prevent students’ active participation in lessons, indicating that the digital divide deepens learning inequalities (Solano-Gutiérrez, 2024). Addressing these barriers calls for actionable steps. For example, educators could: (a) balance structure with frequent, low-friction Q&A channels (TDT); (b) restore presence via predictable micro-feedback and brief synchronous check-ins (CoI); (c) scaffold SRL with weekly planners and progress prompts; and (d) couple access initiatives with digital-literacy supports to mitigate second and third-level digital divide effects (van Dijk, 2006).

This study demonstrates that NLP, sentiment analysis, and topic modeling can effectively analyze middle school students’ unstructured opinions, enabling meaningful insights from large datasets. It extends prior work in distance learning (Borazon et al., 2024) and applies proven techniques from news and literature analysis (Gurcan et al., 2021; Lee et al., 2023) to the education domain. Given the short, grammar-sparse nature of K–12 texts, CorEx provided interpretable themes; a lightweight transformer baseline (e.g., BERT) could contextualise performance without inflating word count and is reserved for future work.

Conclusion

In conclusion, the COVID-19 pandemic brought both notable benefits and serious challenges to middle school students' first experiences with online education. While students saw it as a safe way to maintain learning and gain autonomy, they also faced issues such as limited interaction, physical strain, and perceived quality loss in education. The use of NLP, machine learning, and topic modeling proved effective in quickly analyzing emotional feedback and insights.

In terms of implications for practice, we recommend short synchronous sessions with structured Q&A, regular micro-feedback, low-bandwidth/recorded alternatives, device–Internet support, and ergonomics-aware screen-time limits. Strengthening teacher training and promoting equity-focused initiatives will be essential to ensure that digital education becomes both inclusive and developmentally appropriate.

Limitations and Future Studies

Future research should extend the NLP-based approach to subgroup analyses (e.g., age, gender, region, and student background) to yield more fine-grained insights and guide context-sensitive policy and practice. The use of convenience sampling may introduce bias and limit generalizability; however, due to the extraordinary conditions of COVID-19, this approach was the most feasible, and future research should aim to employ more representative sampling strategies. While traditional qualitative methods can capture richer nuance, they are difficult to apply at this scale; NLP enables large-scale analysis and, when integrated with learning management systems, can provide systematic and timely insights despite some loss of subtlety. Future work should therefore combine NLP with in-depth qualitative approaches (e.g., manual coding, interviews) to balance breadth and depth and to triangulate model outputs. Finally, although CorEx was selected for short-text interpretability, future research could compare it with other state-of-the-art methods (e.g., BERT) and evaluate which approaches yield the most effective results in capturing nuance.

Acknowledgements

This article has benefited from the use of ChatGPT 5 (Extended) to polish the language after it was composed by the author. The authors meticulously reviewed and approved the final version of the content and assume full responsibility for the work.

References

Adams, D., Cheah, K. S. L., Thien, L. M., & Md Yusoff, N. N. (2024). Leading schools through the COVID-19 crisis in a South-East Asian country. *Management in Education*, *38*(2), 72–78. https://doi.org/10.1177/08920206211037738

Akhmedov, F., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS) model. *Applied Sciences, 11*(23), Article 11091. https://doi.org/10.3390/app112311091

Albert, A. (2024). From classrooms to confinement: Academic challenges faced by secondary school children in Kyamuhunga Sub-County during COVID-19. *International Journal of Innovative Science and Research Technology, 9*(9), 3206–3211. https://doi.org/10.38124/ijisrt/IJISRT24SEP1074

Al Husaeni, D. F., Haristiani, N., Wahyudin, W., & Rasim, R. (2022). Chatbot artificial intelligence as educational tools in science and engineering education: A literature review and bibliometric mapping analysis with its advantages and disadvantages. *ASEAN Journal of Science and Engineering*, *4*(1), 93–118. https://doi.org/10.17509/ajse.v4i1.67429

Alweshah, M. (2019). Construction biogeography-based optimization algorithm for solving classification problems. *Neural Computing and Applications*, *31*(10), 5679–5688. https://doi.org/10.1007/s00521-018-3402-8

Amaro, A., & Bacao, F. (2024). Topic modeling: A consistent framework for comparative studies. *Emerging Science Journal*, *8*(1), 125–139. https://doi.org/10.28991/ESJ-2024-08-01-09

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*, *38*(5-6), 987–1008. https://doi.org/10.1080/09588221.2023.2246519

Baral, S., Botelho, A. F., Santhanam, A., Gurung, A., Erickson, J., & Heffernan, N. T. (2023). Investigating patterns of tone and sentiment in teacher written feedback messages. In N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos (Eds.), *Artificial intelligence in education. Posters and late breaking results, Workshops and tutorials, industry and innovation tracks, practitioners, doctoral consortium and blue sky* (CCIS, Vol. 1831, pp. 341–346). Springer. https://doi.org/10.1007/978-3-031-36336-8_53

Bittencourt, I. I., Chalco, G., Santos, J., Fernandes, S., Silva, J., Batista, N., Hutz, C., & Isotani, S. (2024). Positive artificial intelligence in education (P-AIED): A roadmap. *International Journal of Artificial Intelligence in Education*, *34*(3), 732–792. https://doi.org/10.1007/s40593-023-00357-y

Borazon, E. Q., Marques, S., & Saycon, D. R. (2024, November 14). E-learning adoption: A comparative analysis of public sentiments during COVID-19. *Information Technology for Development*, 1–29. https://doi.org/10.1080/02681102.2024.2423286

Caviola, S., Visentin, C., Borella, E., Mammarella, I., & Prodi, N. (2021). Out of the noise: Effects of sound environment on maths performance in middle-school students. *Journal of Environmental Psychology*, *73*, Article 101552. https://doi.org/10.1016/j.jenvp.2021.101552

Copur-Gencturk, Y., Li, J., Cohen, A. S., & Orrill, C. H. (2024). The impact of an interactive, personalized computer-based teacher professional development program on student performance: A randomized controlled trial. *Computers & Education*, *210*, Article 104963. https://doi.org/10.1016/j.compedu.2023.104963

Datta, P., & Nwankpa, J. K. (2021). Digital transformation and the COVID-19 crisis continuity planning. *Journal of Information Technology Teaching Cases*, *11*(2), 81–89. https://doi.org/10.1177/2043886921994821

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment analysis: A comparative study on different approaches. *Procedia Computer Science*, *87*, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22. https://doi.org/10.1177/0047239520934018

Diz-Otero, M., Portela-Pino, I., Domínguez-Lloria, S., & Pino-Juste, M. (2023). Digital competence in secondary education teachers during the COVID-19-derived pandemic: Comparative analysis. *Education + Training*, *65*(2), 181–192. https://doi.org/10.1108/ET-01-2022-0001

Dokuchyna, T. (2023). Students' attitude to distance learning as a component of their mobility. *Pedagogy*, *95*(4), 439–448. https://doi.org/10.53656/ped2023-4.02

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Fang, B., Li, X., Han, G., & He, J. (2023). Facial expression recognition in educational research from the perspective of machine learning: A systematic review. *IEEE Access*, *11*, 112060–112074. https://doi.org/10.1109/ACCESS.2023.3322454

Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics*, *5*, 529–542. https://doi.org/10.1162/tacl_a_00078

Garcia, M. (2021). Embeddings in natural language processing: Theory and advances in vector representations of meaning [Review of the book *Embeddings in natural language processing: Theory and advances in vector representations of meaning*, by M. T. Pilehvar & J. Camacho-Collados]. *Computational Linguistics*, *47*(3), 699–701. https://doi.org/10.1162/coli_r_00410

Garrison, D. R. (2016). *E-learning in the 21st century: A community of inquiry framework for research and practice* (3rd ed.). Routledge. https://doi.org/10.4324/9781315667263

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances, 4*(7), Article eaaq1360. https://doi.org/10.1126/sciadv.aaq1360

Geulayov, G., Mansfield, K., Jindra, C., Hawton, K., & Fazel, M. (2024). Loneliness and self-harm in adolescents during the first national COVID-19 lockdown: Results from a survey of 10,000 secondary school pupils in England. *Current Psychology*, *43*(15), 14063–14074. https://doi.org/10.1007/s12144-022-03651-5

Gurcan, F., Ozyurt, O., & Cagitay, N. E. (2021). Investigation of emerging trends in the e-learning field using latent Dirichlet allocation. *The International Review of Research in Open and Distributed Learning, 22*(2), 1–18. https://doi.org/10.19173/irrodl.v22i2.5358

Hurling, A., Ballard, H., & Whitaker, A. (2024). The effects of COVID-19 on public secondary schools operating in various economic environments. A case study of schools in Metros South and Central Schools, Western Cape. *American Journal of Education and Practice*, *8*(5), 18–29. https://doi.org/10.47672/ajep.2471

Kadirhan, Z., & Sat, M. (2024). K–12 teachers' perceived experiences with distance education during the COVID-19 pandemic: A meta-synthesis study. *Turkish Online Journal of Distance Education*, *25*(3), 57–75. https://doi.org/10.17718/tojde.1320633

Khanna, A., & Coumans, J.-V. (2023). Distributed vector representations of neurosurgical admission notes predict discharge disposition. *Neurosurgery*, *69*(Suppl_1), 47. https://doi.org/10.1227/neu.0000000000002375_322

Lee, K., Kim, T.-J., Cefa Sari, B., & Bozkurt, A. (2023). Shifting conversations on online distance education in South Korean society during the COVID-19 pandemic: A topic modeling analysis of news articles. *The International Review of Research in Open and Distributed Learning, 24*(3), 125–144. https://doi.org/10.19173/irrodl.v24i3.7220

Leech, N. L., Gullett, S., Cummings, M. H., & Haug, C. A. (2022). The challenges of remote K–12 education during the COVID-19 pandemic: Differences by grade level. *Online Learning, 26*(1), 245–267. https://doi.org/10.24059/olj.v26i1.2609

Lin, H., & Chen, Q. (2024). Artificial intelligence (AI)-integrated educational applications and college students' creativity and academic emotions: Students and teachers' perceptions and attitudes. *BMC Psychology*, *12*(1), Article 487. https://doi.org/10.1186/s40359-024-01979-0

Lin, L., Zhou, D., Wang, J., & Wang, Y. (2024). A systematic review of big data driven education evaluation. *Sage Open*, *14*(2), Article 21582440241242180. https://doi.org/10.1177/21582440241242180

Lo, C. K., Cheung, K. L., Chan, H. R., & Chau, C. L. E. (2023). Developing flipped learning resources to support secondary school mathematics teaching during the COVID-19 pandemic. *Interactive*

Learning Environments, *31*(8), 4787–4805. https://doi.org/10.1080/10494820.2021.1981397

Marshall, M. N. (1996). Sampling for qualitative research. *Family Practice*, *13*(6), 522–526. https://doi.org/10.1093/fampra/13.6.522

Martin, F., Sun, T., Westine, C. D., & Ritzhaupt, A. D. (2022). Examining research on the impact of distance and online learning: A second-order meta-analysis study. *Educational Research Review*, *36*, Article 100438. https://doi.org/10.1016/j.edurev.2022.100438

Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review. *Revista de Psicodidáctica (English Ed.)*, *28*(2), 93–103. https://doi.org/10.1016/j.psicoe.2023.06.002

Mifrah, S., & Benlahmar, E. H. (2020). Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 5756–5761. https://doi.org/10.30534/ijatcse/2020/231942020

Mishra, R. K., Urolagin, S., Jothi, J. A. A., Neogi, A. S., & Nawaz, N. (2021). Deep learning-based sentiment analysis and topic modeling on tourism during COVID-19 pandemic. *Frontiers in Computer Science*, *3*, Article 775368. https://doi.org/10.3389/fcomp.2021.775368

Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). Routledge.

Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A., & Ashraf, I. (2021). Sentiment analysis and topic modeling on tweets about online education during COVID-19. *Applied Sciences*, *11*(18), Article 8438. https://doi.org/10.3390/app11188438

Naidu, S. (2019). The changing narratives of open, flexible and online learning. *Distance Education*, *40*(2), 149–152. https://doi.org/10.1080/01587919.2019.1612981

Nowak, K. L., & Watt, J. H. (2022). Distance education during the COVID-19 shutdown: A process model of online learner readiness, experiences and feelings of learning. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), Article 293. https://doi.org/10.1145/3555184

Onan, A. (2021). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. *Computer Applications in Engineering Education*, *29*(3), 572–589. https://doi.org/10.1002/cae.22253

Ozyurt, O., & Ayaz, A. (2022). Twenty-five years of education and information technologies: Insights from a topic modeling based bibliometric analysis. *Education and Information Technologies*, *27*(8), 11025–11054. https://doi.org/10.1007/s10639-022-11071-y

Özyurt, B., & Akcayol, M. A. (2020). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with. Applications*, *168*, Article 114231. https://doi.org/10.1016/j.eswa.2020.114231

Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A survey of text representation and embedding techniques in NLP. *IEEE Access*, *11*, 36120–36146. https://doi.org/10.1109/ACCESS.2023.3266377

Pelaez, A., Jacobson, A., Trias, K., & Winston, E. (2022). The Turing teacher: Identifying core attributes for AI learning in K–12. *Frontiers in Artificial Intelligence*, *5*, Article 1031450. https://doi.org/10.3389/frai.2022.1031450

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Radwan, E., Radwan, A., & Radwan, W. (2020). The role of social media in spreading panic among primary and secondary school students during the COVID-19 pandemic: An online questionnaire study from the Gaza Strip, Palestine. *Heliyon*, *6*(12), Article e05807. https://doi.org/10.1016/j.heliyon.2020.e05807

Rodríguez-Ibánez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*, Article 119862. https://doi.org/10.1016/j.eswa.2023.119862

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

Sandvik, L. V., Smith, K., Strømme, A., Svendsen, B., Aasmundstad Sommervold, O., & Aarønes Angvik, S. (2024). Students' perceptions of assessment practices in upper secondary school during COVID-19. *Teachers and Teaching*, *30*(7–8), 918–931. https://doi.org/10.1080/13540602.2021.1982692

Scully, D., Lehane, P., & Scully, C. (2021). "It is no longer scary": Digital learning before and during the COVID-19 pandemic in Irish secondary schools. *Technology, Pedagogy and Education*, *30*(1), 159–181. https://doi.org/10.1080/1475939X.2020.1854844

Solano-Gutiérrez, G. A. (2024). La tecnología en la educación a distancia: Revisión de progresos y obstáculos a supercar [Technology in distance education: A review of progress and obstacles]. *Revista Científica Zambos*, *3*(2), 48–73. https://doi.org/10.69484/rcz/v3/n2/17

Tomaszewski, W., Zajac, T., Rudling, E., Te Riele, K., McDaid, L., & Western, M. (2023). Uneven impacts of COVID-19 on the attendance rates of secondary school students from different socioeconomic backgrounds in Australia: A quasi-experimental analysis of administrative data. *Australian Journal of Social Issues*, *58*(1), 111–130. https://doi.org/10.1002/ajs4.219

Tuguic, L. A., & Bilan, H. P. (2023). College education students' learning experiences on the advent of online distance education in the Philippines: A phenomenological study. *The Qualitative Report*, *28*(7), 1869–1879 https://doi.org/10.46743/2160-3715/2023.5644

Tzimiris, S., Nikiforos, S., & Kermanidis, K. L. (2023). Post-pandemic pedagogy: Emergency remote teaching impact on students with functional diversity. *Education and Information Technologies*, *28*(8), 10285–10328. https://doi.org/10.1007/s10639-023-11582-2

Uçar, S.-Ş., Aldabe, I., Aranberri, N., & Arruarte, A. (2024). Exploring automatic readability assessment for science documents within a multilingual educational context. *International Journal of Artificial Intelligence in Education*, *34*(4), 1417–1459. https://doi.org/10.1007/s40593-024-00393-2

Upadhyay, H., Juneja, S., Juneja, A., Dhiman, G., & Kautish, S. (2021). Evaluation of ergonomics-related disorders in online education using fuzzy AHP. *Computational Intelligence and Neuroscience*, *2021*(1), Article 2214971. https://doi.org/10.1155/2021/2214971

van Dijk, J. A. G. M. (2006). Digital divide research, achievements and shortcomings. *Poetics, 34*(4–5), 221–235. https://doi.org/10.1016/j.poetic.2006.05.004

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, Article 101582. https://doi.org/10.1016/j.is.2020.101582

Virtanen, S., & Girolami, M. (2019). Precision–recall balanced topic modelling. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32, pp. 6747–6756). Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2019/hash/310cc7ca5a76a446f85c1a0d641ba96d-Abstract.html

Waheeb, S. A., Khan, N. A., & Shang, X. (2022). Topic modeling and sentiment analysis of online education in the COVID-19 era using social networks based datasets. *Electronics*, *11*(5), Article 715. https://doi.org/10.3390/electronics11050715

Xing, W., Li, H., Kim, T., Zhu, W., & Song, Y. (2025). Investigating the behaviors of core and periphery students in an asynchronous online discussion community using network analysis and topic modeling. *Education and Information Technologies*, *30*(5), 5561–5588. https://doi.org/10.1007/s10639-024-13038-7

Xu, H., Gan, W., Qi, Z., Wu, J., & Yu, P. S. (2024). *Large language models for education: A survey*. arXiv. arXiv:2405.13001. https://doi.org/10.48550/arXiv.2405.13001

Zarkadoulas, D., & Virvou, M. (2024). Emotional expression in mathematics e-learning using emojis: A gender-based analysis. *Intelligent Decision Technologies*, *18*(2), 1181–1201. https://doi.org/10.3233/IDT-240170

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

Žitnik, S., & Smith, G. G. (2024). Automated analysis of postings in fourth grade online discussions to help teachers keep students on-track. *Interactive Learning Environments*, *32*(8), 4587–4612. https://doi.org/10.1080/10494820.2023.2204327

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

How Task and Individual Characteristics Affect Students' Cognitive Load: The Moderating Role of AI-Generated Content

Pan Liu[1], Qiang Jiang[1]*, Weiyan Xiong[2], Wei Zhao[1]
*[1]School of Information Science and Technology, Northeast Normal University, China; [2]Department of International Education, Education University of Hong Kong, Hong Kong; *Corresponding Author*

Abstract

This study examined how task characteristics (TC) and individual characteristics (IC) affect cognitive load (CL) and how artificial intelligence generated content (AIGC) moderates these effects in online learning. Participants included 435 undergraduate students (200 males and 235 females) enrolled in an introductory educational technology course. A structural model, conducted using Mplus software, was employed to test the relationships between each of TC and IC, and CL. Additional analyses explored the moderating role of AIGC on the relationship between TC and CL, the impact of AIGC on the relationship between IC and CL, as well as how these patterns differed by gender. Results revealed that TC positively affected CL, whereas IC exhibited a negative correlation. Moreover, AIGC negatively affected the relationship between TC and CL, but it enhanced the relationship between IC and CL. The moderating role of AIGC differed by gender. Specifically, AIGC positively influenced the connection between IC and CL among males but not females, and it weakened the relationship between TC and CL among females but not males. The implications and limitations are also discussed.

Keywords: task characteristics, individual characteristics, cognitive load, AIGC, structuring equation modeling, online learning

Introduction

Online learning, as a core driver of the digital transformation in education, has promote educational equity and personalized development through technological empowerment (Rulinawaty et al., 2023). As a key component of open and distance learning (ODL), online learning environments have typically been characterized by learner autonomy, reduced teacher supervision, and asynchronous interactions. However, as online educational content has become increasingly complex and knowledge updates have accelerated without timely guidance from teachers, students in ODL have been confronted with escalating cognitive load (CL; Skulmowski & Xu, 2022). CL refers to the cumulative mental resources expended during information processing (Chen et al., 2023; Sweller, 1988). When these resources exceed an individual's processing capacity, CL can occur, negatively affecting learning outcomes. Therefore, identifying the key factors affecting CL in online learning is crucial for enhancing students' learning efficiency and reducing their psychological stress.

Existing research has extensively explored the factors influencing CL (e.g., Li, 2010; Paas & Van Merriënboer, 1994; Tremblay et al., 2023). The seminal CL structural model by Paas and Van Merriënboer (1994) identified task characteristics (TC) and individual characteristics (IC) as pivotal factors shaping CL. TC, such as task complexity and time pressure, are external factors that directly affect learners' psychological burden during task completion, thereby increasing CL (Chen et al., 2023). IC encompasses psychological factors like self-efficacy and state meta-cognition, reflecting individuals' psychological states and capabilities when facing learning tasks (Le et al., 2024; Orthey et al., 2019). Numerous studies have affirmed the significant influence of TC and IC on CL (Le et al., 2024; Li, 2010; Tremblay et al., 2023). However, research has predominantly focused on identifying factors related to CL in online learning, with limited exploration into the specific mechanisms through which TC and IC exert their influence.

Researchers have proposed moderating factors, including intelligent tools, that could potentially influence the relationships between each of TC and IC, and CL (Zhao et al., 2024; Wu et al., 2024). Artificial intelligence generated content (AIGC), an online learning tool, manages learning resources and enhances learning processes (Lo, 2023; Xiao et al., 2024). It effectively breaks down complex tasks and reduces task completion time in online learning, thereby mitigating the effects of task complexity and time pressure on CL (Zhai et al., 2024). Drawing on triadic reciprocal determinism (TRD), AIGC has enhanced students' cognition and self-efficacy (Urban et al., 2024), potentially moderating the effect of IC on CL. Although preliminary studies have explored the effects of AIGC on CL, most research has focused on the utility of these tools rather than delving into their functional mechanisms within CL (Zhao et al., 2024). Specifically, there have been few systematic investigations into how AIGC influences the relationships between TC and CL, and IC and CL.

This study established a structural equation model (d to explore how AIGC moderates the relationships between TC and CL, as well as between IC and CL in online learning. Accordingly, this paper extended cognitive load theory (CLT) and provided practical guidance for digital education by investigating the moderating role of AIGC on CL in online learning. The specific research questions were as follows:

What is the relationship between each of TC and IC, and CL?

How does AIGC moderate the effect of each of TC and IC on CL in online learning?

Are there gender differences in how AIGC moderates the relationship between IC and CL, as well as between TC and CL, in online learning?

Conceptual Model and Hypotheses

Cognitive load theory asserts that students' cognitive resources are finite, and any learning or problem-solving activities consume these resources, leading to CL (Mayer, 2005; Sweller, 1988). CL refers to the overall mental effort placed on a person's cognitive system during a particular task duration (Sweller, 1988), or as the burden on an individual's cognitive system within a defined time period (Cooper, 1990). A renowned CL structure model, established by Paas and Van Merriënboer (1994), reflected causal factors and assessment factors. Since this study focuses on the influencing factors of CL, we adopted the causality component, including TC and IC, to understand the meaning and structure of student CL.

Beyond the factors directly influencing CL, elements like learning tools moderate the relationships between these factors (i.e., TC and IC) and CL. AIGC, an intelligent learning tool, has been widely researched (Du & Lv, 2024). In education, scholars have found that AIGC can alter TC, such as task complexity, and improve students' digital competence (Zhao et al., 2024). Furthermore, researchers have shown through quasi-experimental studies that AIGC can enhance university students' programming self-efficacy and individual motivation (Yilmaz & Yilmaz, 2023). These studies examined AIGC as a moderating variable to explore its effects on the relationships between TC and CL, as well as between IC and CL.

Effects of TC and IC on CL

Task Characteristics and Cognitive Load

With the development of CLT, TC has been defined as attributes related to cognitive effort, including task complexity and time pressure (Paas et al., 2003; Sweller, 1998). Task complexity refers to the cognitive requirements or characteristics associated with the task that elevate the amount of information to be processed (Chen et al., 2023). Time pressure results from strict time limits imposed on task completion, leading individuals to feel stress and anxiety, thereby increasing CL and affecting the quality and speed of task execution (Seitz et al., 2023). We hypothesized that TC, comprising task complexity and time pressure, influences students' CL.

This inference was supported by numerous empirical studies (Tremblay et al., 2023). For instance, Li (2010) conducted a simulated dual-task experiment to explore the main factors and pathways affecting CL and found that TC, including factors like task complexity and time pressure, had a significant positive effect on CL. Additionally, TC's effect on CL has been demonstrated through the design of both simple and complex tasks to assess college students' performance (Tremblay et al., 2023). Meanwhile, task-related factors, such as time pressure, significantly affected CL (Li, 2010). Based on this, we proposed our first hypothesis.

Hypothesis 1: Task characteristics have a significant positive influence on cognitive load.

Individual Characteristics and Cognitive Load

IC typically refers to the unique properties or attributes that distinguish one person from another. In CLT, Paas and Van Merriënboer (1994) categorized IC into relatively stable traits (e.g., prior knowledge and abilities) and unstable traits associated with IC (e.g., self-efficacy and state meta-cognition). Given that this study focused on tasks, we chose to investigate IC associated with TC, specifically self-efficacy and state meta-cognition.

Numerous studies have researched the relationship between IC and CL, focusing on self-efficacy and meta-cognitive states. For instance, Feldon et al. (2023) targeted undergraduate students and examined the relationship between CL and self-efficacy through timely and longitudinal measures, finding a correlation between the two. Redifer et al. (2021) discovered that high creative self-efficacy was associated with low CL in creative thinking tasks. Existing research has consistently demonstrated a direct correlation between students' self-efficacy and CL, suggesting that higher self-efficacy correlates with lower CL compared with lower self-efficacy (Jiang, 2023). Similarly, research has indicated a correlation between meta-cognitive states and CL, with lower CL observed under conditions of higher meta-cognitive states (Bürgler et al., 2024). Given the literature above, we proposed a second hypothesis.

Hypothesis 2: Individual characteristics have a significant negative influence on cognitive load.

Moderating Role of AIGC

AIGC, a leading intelligent technology and online learning tool, has been widely touted and researched, particularly for its potential in teaching, personalized feedback, digital transformation, and learning assessment (Lo, 2023; Zhao et al., 2024). For instance, through semi-structured interviews and observations, Zhao et al. (2024) found that ChatGPT, a powerful AIGC platform, influenced students' writing tasks in higher education and facilitated the digitization of student writing. AIGC can be employed for teaching practices and personalized tutoring (Bai et al., 2025). In sum, AIGC has served as a moderating variable to optimize students' cognitive resources, assist in online learning tasks, and influence students' learning outcomes.

Effect of AIGC on the Relationship Between TC and CL

Previous studies have suggested that AIGC can modulate CL by decomposing complex tasks, reducing cognitive intensity, and offering timely assistance and feedback (Zhai et al., 2024). From the perspective of CLT, students' cognitive resources are limited, and when the cognitive resources demanded for learning tasks exceed capacity, CL increases. AIGC can break down complex learning tasks, reduce the cognitive resources needed, and allocate resources more effectively (Shah & Soosai Raj, 2024). Research in human–computer interaction has emphasized cognitive offloading, where external tools reduce CL and free mental resources for other tasks (Nückles et al., 2020).

These studies showed that AIGC can serve as an intelligent tool to break down complex tasks and reduce task completion time, alleviating the effect of TC (e.g., task complexity and time pressure) on CL in online learning. The following is our third hypothesis.

Hypothesis 3: AIGC negatively moderates the relationship between TC and CL in online learning.

Effect of AIGC on the Relationship Between IC and CL

TRD has posited that human behavior, cognition, and the environment mutually influence one another (Bandura, 1983). Within this framework, AIGC can be seen as an online learning environmental factor influencing behavior and cognition through user interaction. TRD has also highlighted that the formation of self-concept is influenced by the external environment. AIGC has provided accurate information and suggestions, enhancing individuals' confidence in specific tasks.

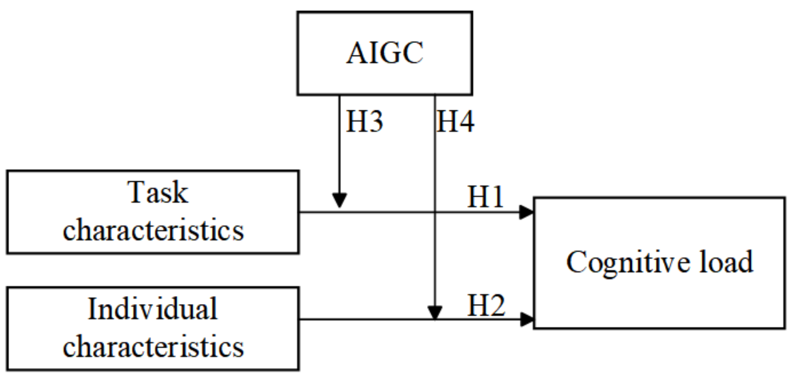
When faced with complex tasks and distractions, AIGC acts as a supportive tool in online learning, boosting individuals' confidence and enabling meta-cognitive activities such as planning, monitoring, and self-checking, thereby strengthening the effect of IC (i.e., self-efficacy and meta-cognitive states) on CL (Noy & Zhang, 2023; Urban et al., 2024). Therefore, AIGC may moderate the relationship between IC and CL. Based on these findings, we proposed a fourth hypothesis.

Hypothesis 4: AIGC enhances the moderating effect between IC and CL in online learning.

The proposed conceptual model of this paper is illustrated in Figure 1.

Figure 1

The Hypothesized Model



Methods

Participants and Data Collection

This research was carried out in a modern educational technology course at a university in northeast China. The online course focused on the design and development of digital resources. Weekly two-hour sessions were organized, with an online teaching assistant providing support to students on their assignments. After completing the course, students were asked to fill out a questionnaire consisting of 17 questions covering four aspects: CL, TC, IC, and AIGC.

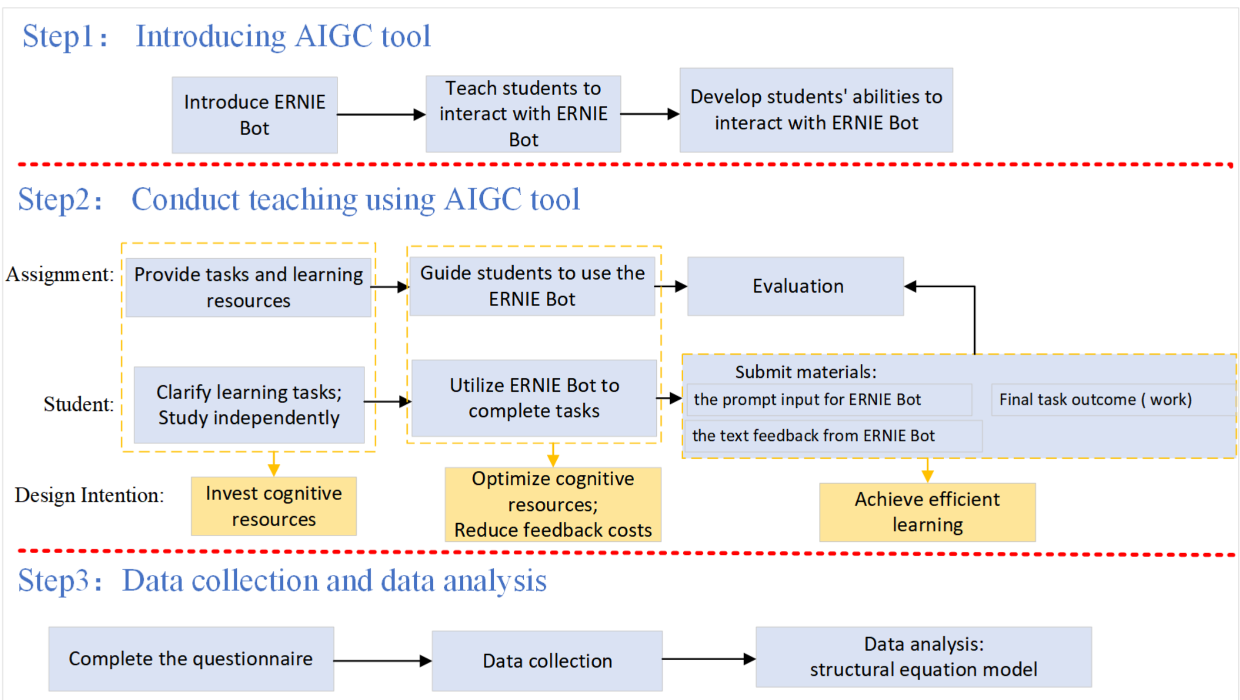
The 480 questionnaires were distributed, and all were successfully returned (100% response rate). Invalid questionnaires, including those completed very quickly (< 30 seconds) or with extreme responses (all *strongly agree* or all *strongly disagree*), were excluded, resulting in 435 valid questionnaires (91% validity rate). Participants ranged in age from 18 to 22 years, with a mean age of 19.82 years (*SD* = 0.95). Among the valid responses, there were 200 male participants (46% of the total) and 235 female participants (54% of the total). All participants who volunteered for the survey were informed that their data would be used solely for research purposes. The ethical standards employed in this study were established by the Science and Technology Ethics Committee of Northeast Normal University

Procedures

The AIGC tool used in this study was ERNIE Bot, developed by Baidu, similar to ChatGPT, known for its deep understanding of Chinese culture (Rudolph et al., 2023). The experimental design process is illustrated in Figure 2.

Figure 2

Experimental Process

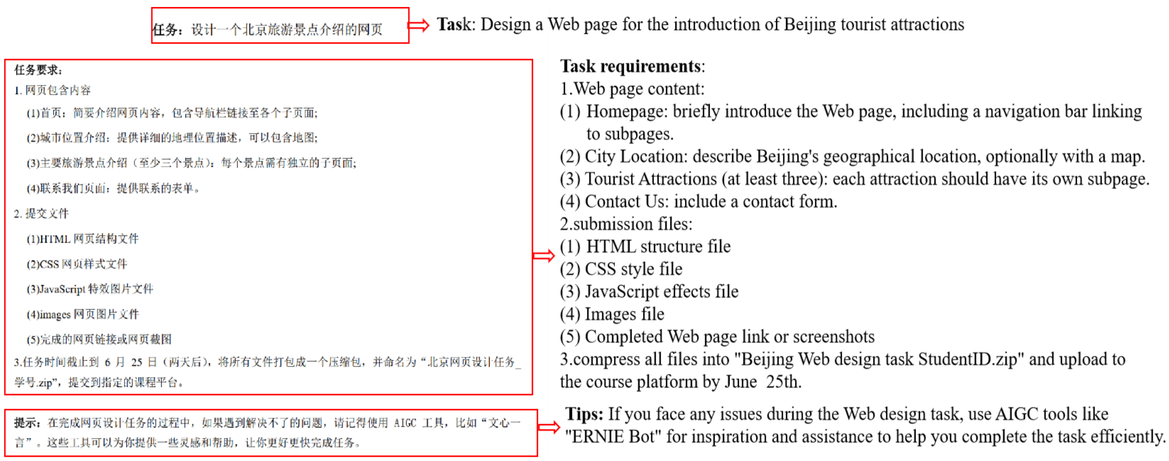


The experimental procedure was divided into three steps. First, students were introduced briefly to ERNIE Bot, including its generative technology, dialogue interaction functions, and methods of use. They were instructed on how to interact with ERNIE Bot; students posed questions and received responses from the bot. For instance, tasks assigned to students beyond the curriculum included understanding the developmental history of generative artificial intelligence and learning to interact effectively with ERNIE Bot.

In the second phase, the instructor provided learning assignments and resources, and students, guided by ERNIE Bot, carried out the online learning tasks. To illustrate this process, we used the example of designing a Web page to introduce Beijing, which served as the final learning task of the course. Initially, students received the task and related resources; see Figures 3 and 4.

Figure 3

Screenshot of Online Learning Tasks for Web Design


任务：设计一个北京旅游景点介绍的网页

任务要求：
1. 网页包含内容
(1)首页：简要介绍网页内容，包含导航栏链接至各个子页面;
(2)城市位置介绍：提供详细的地理位置描述，可以包含地图;
(3)主要旅游景点介绍（至少三个景点）：每个景点需有独立的子页面;
(4)联系我们页面：提供联系的表单。
2. 提交文件
(1)HTML 网页结构文件
(2)CSS 网页样式文件
(3)JavaScript 特效图片文件
(4)images 网页图片文件
(5)完成的网页链接或网页截图
3.任务时间截止到 6 月 25 日（两天后），将所有文件打包成一个压缩包，并命名为“北京网页设计任务_学号.zip”，提交到指定的课程平台。

提示：在完成网页设计任务的过程中，如果遇到解决不了的问题，请记得使用 AIGC 工具，比如“文心一言”。这些工具可以为你提供一些灵感和帮助，让你更好更快完成任务。

Task: Design a Web page for the introduction of Beijing tourist attractions

Task requirements:
1.Web page content:
(1) Homepage: briefly introduce the Web page, including a navigation bar linking to subpages.
(2) City Location: describe Beijing's geographical location, optionally with a map.
(3) Tourist Attractions (at least three): each attraction should have its own subpage.
(4) Contact Us: include a contact form.
2.submission files:
(1) HTML structure file
(2) CSS style file
(3) JavaScript effects file
(4) Images file
(5) Completed Web page link or screenshots
3.compress all files into "Beijing Web design task StudentID.zip" and upload to the course platform by June 25th.

Tips: If you face any issues during the Web design task, use AIGC tools like "ERNIE Bot" for inspiration and assistance to help you complete the task efficiently.

Figure 4

Screenshots of Students Interacting with ERNIE Bot in Online Environment

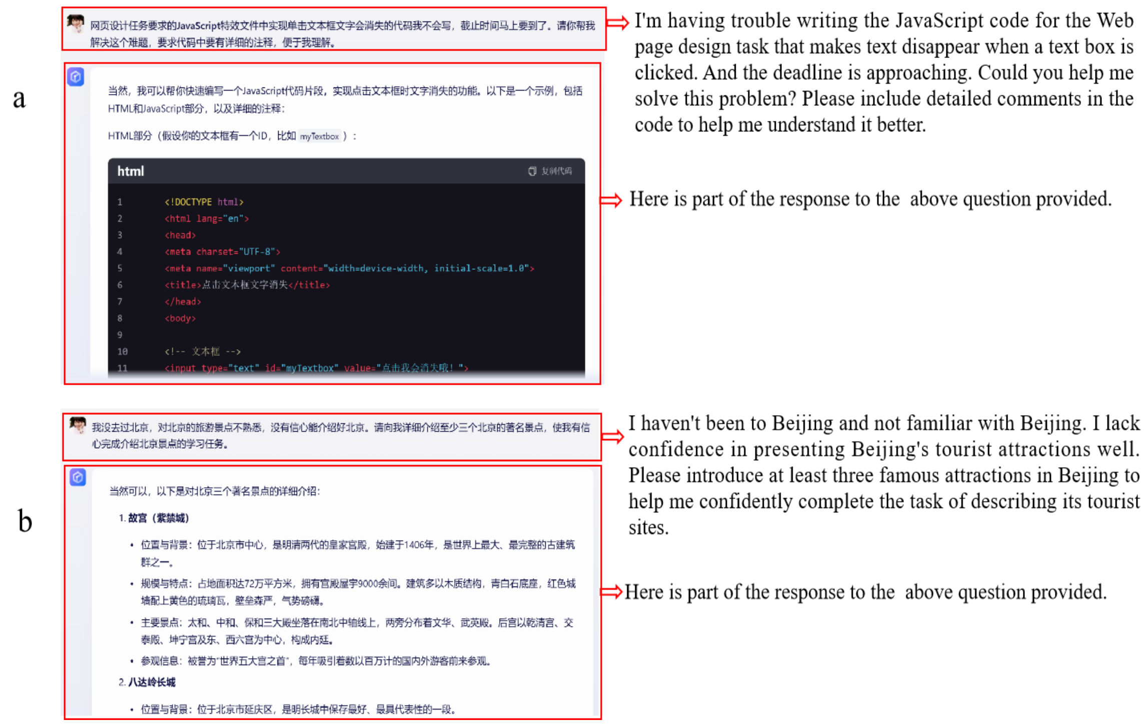


Figure 3 outlines the Web page design theme, requirements, submission files, and task tips. Subsequently, students used ERNIE Bot to complete the task, as depicted in Figure 4, showing their interaction during TC and IC activities. (Figure 4a depicts the student's interaction with ERNIE Bot regarding TC, and Figure 4b shows the interaction concerning IC). This phase spanned approximately one semester, during which multiple course sub-tasks were completed.

The third step of the experiment included administering an online questionnaire after the course. Subsequently, the collected data were organized using SPSS and Excel software. The associations among TC, IC, CL, and AIGC were analyzed using SEM established by Mplus software.

Measures

All items in the questionnaire were adapted from existing scales. To ensure the suitability of the content for our research, we invited two experts from the fields of psychology and education to review the comprehensibility of the questions and make necessary modifications. The revised questionnaire consisted of 17 items, encompassing the variables TC, IC, CL, and AIGC. Responses were recorded using a 5-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), or in some cases, from 1 *(very low)* to 5 *(very high)*.

Cognitive Load

CL was assessed using the National Aeronautics and Space Administration's task load index (NASA-TLX) scale, widely recognized and comprising six items: (a) mental demand, (b) physical demand, (c) temporal demand, (d) effort, (e) own performance, and (f) frustration level (Hart & Staveland, 1988). The scale demonstrated good reliability with Cronbach's alpha coefficients ranging from 0.70 to 0.90. Given the nature of course tasks and the learning processes of college students, this study adapted four items from the original scale to measure CL. Participants responded to questions such as:

How hard did you have to work (mentally and physically) to accomplish your level of performance?

How much physical activity was required?

How much time pressure did you feel because of the rate or pace of the tasks or task elements?

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?

Responses were scored on a scale ranging from 1 (*very low*) to 5 (*very high*).

Task Characteristics

This study assessed TC using the dimensions of task complexity and time pressure. Task complexity was measured using the subjective task complexity scale developed by Maynard & Hakel (1997), which has a high reliability coefficient (Cronbach's $\alpha = 0.93$). From the original 10-item scale, items focusing on subjective perceptions of task complexity were selected for this study. These included statements such as (a) I found this task to be complex, and (b) this task required a lot of thought and problem-solving. Time pressure was measured using a scale adapted from Chong et al. (2010), which showed satisfactory reliability (Cronbach's α ranging from 0.73 to 0.88). The scale items were modified to align with the study's context and included statements such as (a) the importance of completing this task on time, and (b) the lack of time buffer planned for this task.

Individual Characteristics

State meta-cognition and self-efficacy are crucial dimensions for studying students' IC (Li, 2010). State meta-cognition was assessed using a scale developed by O'Neil and Abedi (1996), whereas self-efficacy was measured using the general self-efficacy scale developed by Schwarzer and Jerusalem (1995), translated into Chinese by Wang et al. (2001; Cronbach's $\alpha = 0.87$). Both scales were adapted to align with the requirements of this study. Items for state meta-cognition included statements such as (a) I was aware of my thinking, and (b) I checked my work while I was doing it. Items for self-efficacy included statements such as (a) I can always manage to solve difficult problems if I try hard enough, and (b) It is easy for me to stick to my aims and accomplish my goals. All items were rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*d*, where higher scores indicated higher levels of IC.

AI-Generated Content

AIGC evaluation in this study adopted the technology acceptance model (TAM) framework (Davis, 1993), which categorizes AIGC into perceived usefulness and perceived ease of use. The scales demonstrated high reliability, with Cronbach's alpha coefficients of 0.97 for perceived usefulness and 0.91 for perceived ease of use. To align with the experimental tools and research objectives, a questionnaire was developed to assess college students' perceptions and applications of AIGC. The questionnaire included five items adapted from TAM, such as: it is easy for me to use generative AI tools (like ERNIE Bot). All items used a 5-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*), where higher scores reflected greater support and satisfaction with AIGC.

Data Analysis

In this study, data analysis was divided into three main parts. First, raw data was organized using SPSS and Excel software, including removing outliers and preparing the data for Mplus software. Second, the reliability and validity of the questionnaire were analyzed using Mplus. This involved assessing construct reliability through item reliability (R2), Cronbach's alpha, and composite reliability (CR). Convergent validity was evaluated using average variance extracted (AVE), and discriminant validity was examined using the Fornell–Larcker criterion. Descriptive statistics, such as means and standard deviations, were computed for each variable to assess their central tendency and variability.

Finally, a SEM was established to analyze the study's hypotheses. The model examined the main effects of CL, TC, and IC for hypotheses 1 and 2. It also tested the moderating role of AIGC for hypotheses 3 and 4, while exploring gender differences in the moderating effects of AIGC.

Results

Reliability and Validity

This study assessed reliability using internal consistency reliability and CR. Internal consistency reliability, measured by Cronbach's alpha coefficient, exceeded 0.70, indicating high internal consistency. Additionally, a CR value exceeding 0.70 is acceptable (Hair, 1998). Item reliability, indicated by factor loadings greater than 0.71, suggests high item quality (Hair et al., 2009).

Validity was examined using both convergent and discriminant validity. Convergent validity, evaluated through the AVE, should ideally exceed 0.5 (Fornell & Larcker, 1981). Discriminant validity was tested using the Fornell–Larcker criterion, where the square root of the AVE for each construct should exceed its correlation with any other construct, demonstrating adequate distinctness.

Table 1

Reliability and Convergent Validity

Construct indicator	Item	Factor loading	Item reliability (R2)	Composite reliability (CR)	Cronbach's alpha	Average variance extracted (AVE)
CL	CL1	0.86	0.74	0.86	0.82	0.61
	CL2	0.83	0.68			
	CL3	0.72	0.51			
	CL4	0.72	0.52			
TC	TC1	0.76	0.58	0.84	0.84	0.58
	TC2	0.81	0.65			
	TC3	0.79	0.63			
	TC4	0.67	0.45			
IC	IC1	0.82	0.67	0.88	0.79	0.64
	IC2	0.86	0.74			
	IC3	0.76	0.57			
	IC4	0.76	0.58			
AIGC	AIGC1	0.81	0.66	0.84	0.89	0.63
	AIGC2	0.88	0.77			
	AIGC3	0.78	0.61			
	AIGC4	0.72	0.51			
	AIGC5	0.76	0.58			

As Table 1 indicates, all Cronbach's alpha values exceeded 0.7, with CR values ranging from 0.84 to 0.88. Factor loadings were greater than 0.71, indicating high item reliability. The AVE values ranged from 0.58 to 0.64, exceeding their correlations with other constructs, indicating good validity. This analysis shows that the questionnaire used in this study met high standards of reliability and validity, providing strong support for the research results.

Descriptive Statistics and Correlation Analysis

The study analyzed means, standard deviations, and correlations, revealing the following mean values: CL (M = 4.07, SD = 0.48); TC (M = 4.17, SD = 0.44); IC (M = 1.86, SD = 0.45); and AIGC (M = 3.70, SD = 0.62).

Table 2

Descriptive Statistics, Latent Variable Correlation Matrix, and Discriminant Validity

Variable	Descriptive statistics		Discriminant validity			
	M	*SD*	CL	TC	IC	AIGC
CL	4.07	0.48	**0.78**			
TC	4.17	0.44	0.44***	**0.76**		
IC	1.86	0.45	−0.32***	−0.46***	**0.80**	
AIGC	3.70	0.62	−0.17***	0.17***	0.45***	**0.79**

Note. ****p* < .001, ***p* < .005. Diagonal values in bold represent the square root of the AVE values.

According to Table 2, CL had significant positive correlations with TC (r = 0.44, p < .001) and significant negative correlations with IC (r = −0.32, p < .001) and AIGC (r = −0.17, p < .001). TC was negatively correlated with IC (r = −0.46, p < .001) and AIGC (r = −0.26, p < .001). Additionally, IC exhibited a positive correlation with AIGC (r = 0.45, p < .001).

Structural Model Analysis and Hypothesis Testing

Using maximum likelihood estimation methods, a goodness-of-fit analysis was conducted for an SEM containing only dependent and independent variables.

Table 3

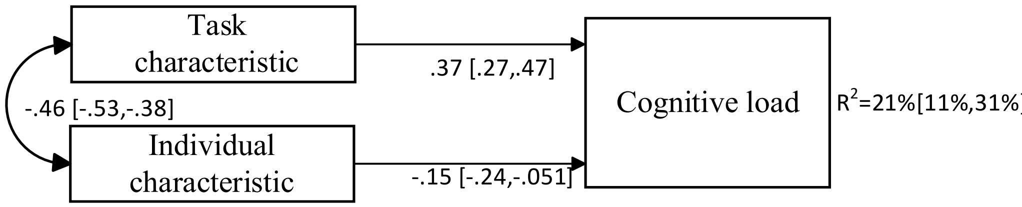
Results of Model Fit

Index	Recommended value	Criteria	Research model
χ^2/df	<3	Kline, 2011	2.20
CFI	>.90		0.97
TLI	>.90	Hu and Bentler,1999	0.96
SRMR	<.08		0.04
RMSEA	<.08	Browne and Cudeck, 1992	0.05

As shown in Table 3, according to established standards (Browne & Cudeck, 1992; Hu & Bentler, 1999; Kline, 2011), the values for χ^2/df, comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) all met the criteria, indicating a satisfactory model fit.

Figure 5

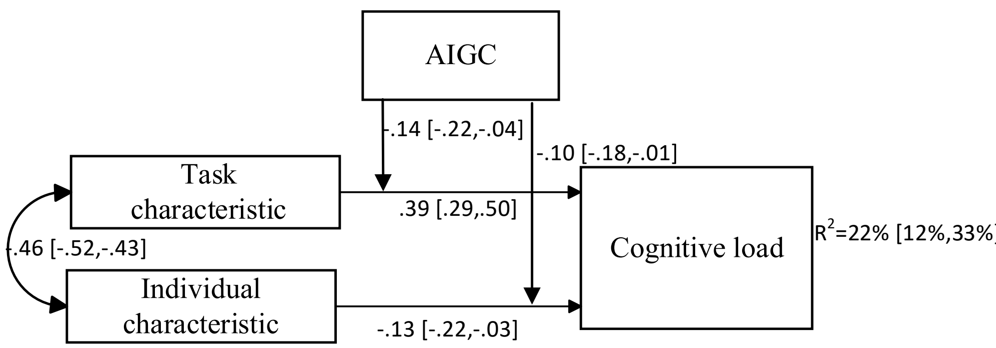
Main Effects of Each of TC and IC on CL, and Correlations Between Variables



The model examined the main effects of TC and IC on CL using the full sample (see Figure 5). Standardized coefficients and 95% confidence intervals (values in brackets) are provided. The bidirectional solid line in Figure 5 reflects a significant association (p < .05). The results showed that TC was positively associated with CL (β = 0.37, p < .001), whereas IC was negatively associated with CL (β = −0.15, p < .001). Therefore, hypotheses 1 and 2 were supported. Overall, the predictors explained 21% of the variability in CL.

Figure 6

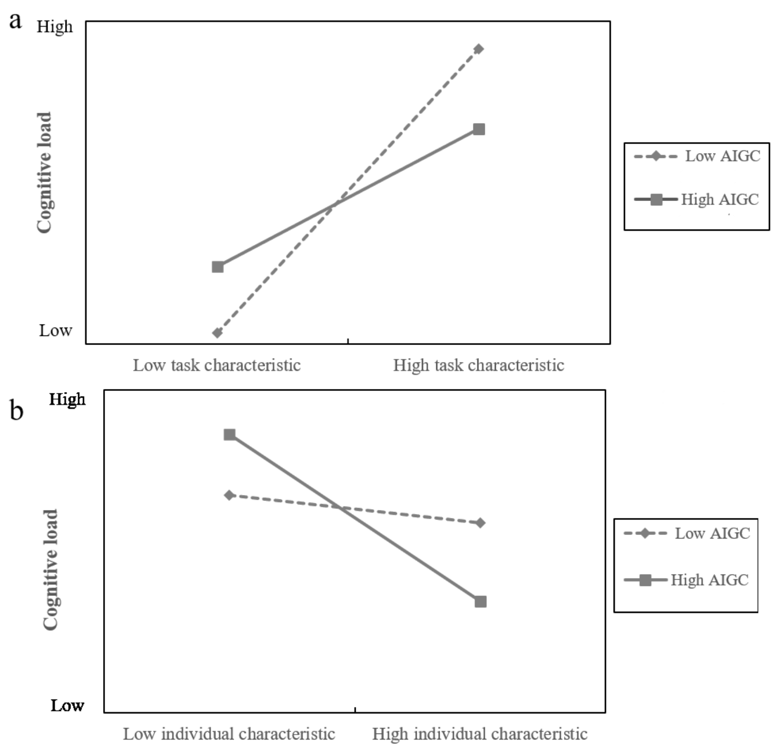
Moderating Role of AIGC



Furthermore, the potential moderation of AIGC on the effects of TC and IC was tested for the full sample. Two significant associations were observed (see Figure 6). Standardized coefficients and 95% confidence intervals (values in brackets) are provided. The bidirectional solid line reflects a significant association (p < .05). Specifically, AIGC moderated the associations between TC and CL (β =−0.14, p < .005) and between IC and CL (β =−0.10, p < .05). The addition of these moderating effects explained an additional 1% of the variance in CL because of TC and IC.

Figure 7

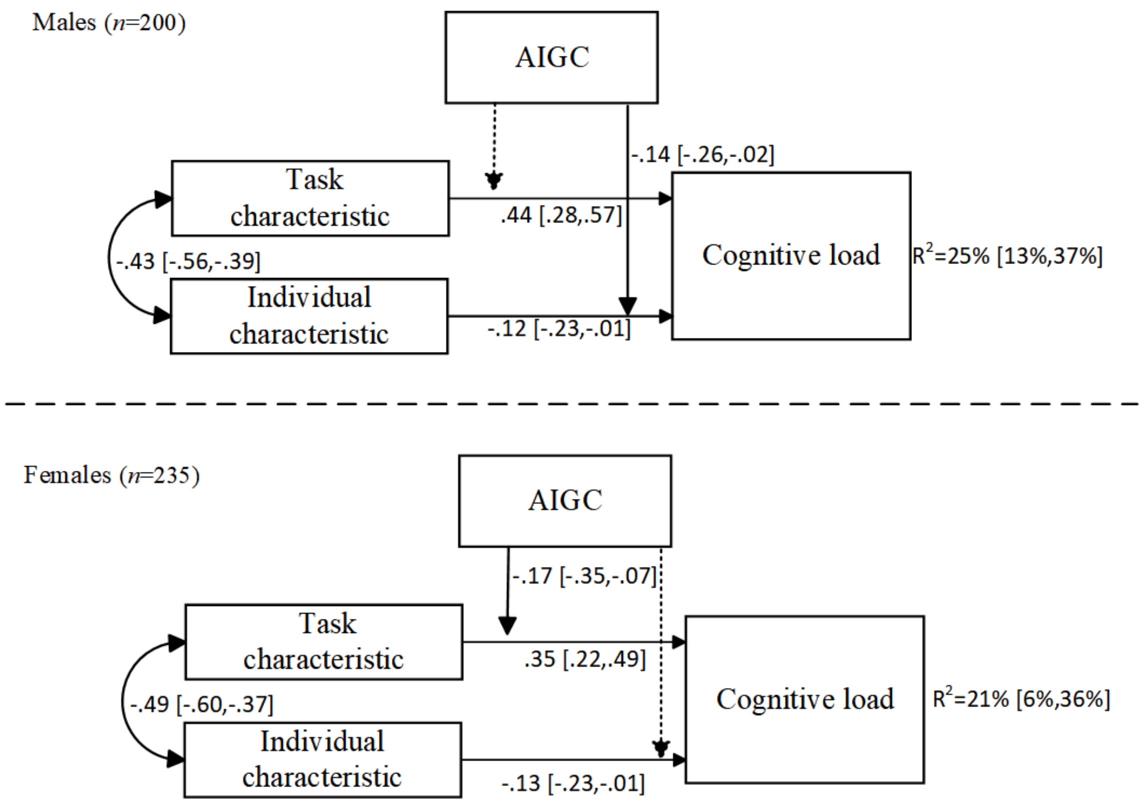
Moderating Role of AIGC on the Relationship Between TC and CL (Panel 7a) and Between IC and CL (Panel 7b)



To visualize the interaction effects proposed in hypotheses 3 and 4, a simple slopes test of AIGC was conducted. As shown in Figure 7a, for AIGC at a low level (one *SD* below the mean), simple slope = 0.56, $p < .001$, while for AIGC at a high level (one *SD* above the mean), simple slope = 0.29, $p < .001$. The effect of TC on CL was weakened at higher values of AIGC. In contrast, Figure 7b shows that for AIGC at a low level, simple slope = −0.04, $p < .001$, and for AIGC at a high level, simple slope =−0.23, $p < .001$. The effect of IC on CL was exacerbated at higher values of AIGC. Thus, hypotheses 3 and 4 were supported.

Figure 8

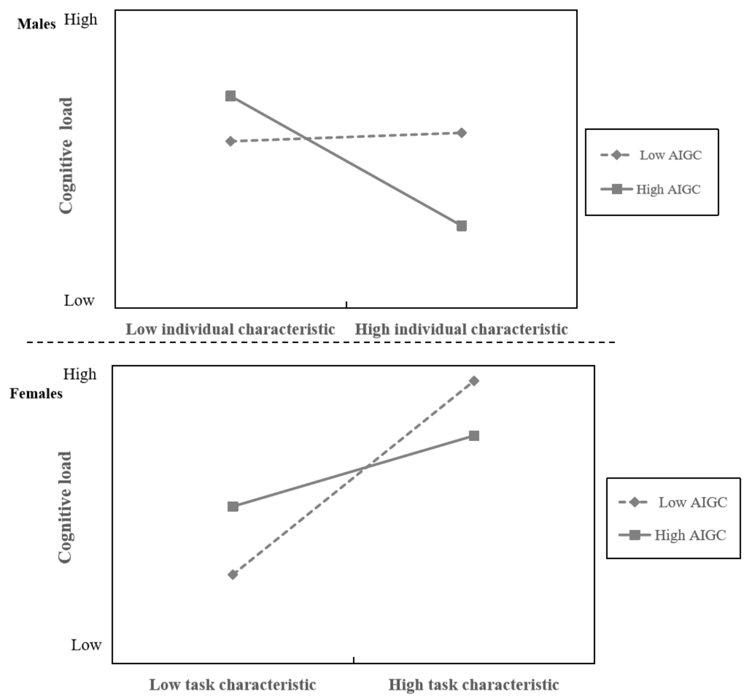
Moderation of AIGC Between Each of TC and IC, and CL, Segmented by Gender



We analyzed the moderating role of AIGC after dividing the sample by gender (see Figure 8). Standardized coefficients are provided. Bidirectional solid lines indicate significant associations ($p < .05$). As shown in Figure 8, the moderating effects of AIGC differed noticeably between males and females. Among males, the effect of IC on CL was enhanced at higher values of AIGC. An analysis of the explained variance revealed no significant difference in CL between males (25%) and females (21%; $p > .05$).

Figure 9

Moderating Role of AIGC on the Relationships Between IC and CL Among Males (Top Panel), and Between TC and CL Among Females (Bottom Panel)



In Figure 9, for males, AIGC at a low level, simple slope = −0.02, $p < .05$, and at a high level, the simple slope = −0.25, $p < .005$, indicating that AIGC increasingly strengthened the effect of IC on CL. Conversely, among females, the effect of TC on CL worsened at higher values of AIGC (low AIGC simple slope = 0.58, $p < .001$; high AIGC simple slope = 0.24, $p < .005$).

Discussion

This study verified the influential relationships between each of TC and IC, and CL in online learning. More importantly, it investigated how AIGC moderates these relationships among college students. Gender differences in the regulatory role of AIGC were revealed. The findings of this study supported the hypotheses and were consistent with previous studies.

Regarding the first research question, this study discovered that TC directly influenced CL with a positive correlation, while there was a negative correlation between IC and CL. These findings aligned with existing literature (Li, 2010; Tremblay et al., 2023). Li (2010) found that task complexity and time pressure, as components of TC, positively correlated with CL. Similarly, Tremblay et al. (2023) demonstrated that TC significantly impacted students' CL, as increased task demands required more cognitive resources (Chakraborty et al., 2024). Additionally, a negative correlation between IC and CL supported hypothesis 2, consistent with prior studies (Feldon et al., 2023; Redifer et al., 2021). According to CLT and social cognitive theory, self-efficacy and state meta-cognition help manage cognitive resources, reducing CL. For instance, Feldon et al. (2023) showed a negative correlation between self-efficacy and CL, while Scheiter et al. (2009) and Sweller (2006) found that learners with favorable IC reported lower CL. Thus, enhancing IC (e.g., self-efficacy and meta-cognition) can effectively reduce students' CL. Notably, potential confounding variables such as students' prior knowledge, learning motivation, and technological familiarity may influence the observed relationships, which were not fully controlled in this study. These findings further validated the core tenet of CLT, which has posited that cognitive resources are limited and can be optimized by improving IC and managing TC. By confirming that increased IC reduces CL while increased TC heightens CL, this study reinforced the causal framework of CLT and highlighted the relevance of both learner characteristics and task design in CL management.

Regarding the second research question, according to the results of the structural equation analysis, AIGC attenuated the relationship between TC and CL, while it positively moderated the effect of IC on CL in online learning. AI reduces workload and CL (Gandhi et al., 2023), with large language models generating task-appropriate responses to alleviate workload (Ayers et al., 2023; Gu & Li, 2022). Our study confirmed that AIGC mitigated TC's impact on CL through natural language processing and intelligent tutoring systems (Chen et al., 2024). Furthermore, AIGC positively moderated the relationship between TC and CL, consistent with prior studies (Gilbert et al., 2023; Lodge et al., 2023). Lodge et al. (2023) described this as offload expansion, where AIGC influences IC and CL. Cognitive offloading has suggested that external tools reduce CL, and AIGC tools like ChatGPT impact student writing practices (Rudolph et al., 2023). Thus, AIGC enhances abilities and moderates the relationship between TC and CL (Lodge et al., 2023). However, it is critical to note that the moderating effect of AIGC might be confounded by students' prior experience with AIGC or their digital literacy levels (Chiappe et al., 2024h). Learners with higher proficiency in using AIGC may benefit more from its cognitive support, which could independently influence CL (Yang, 2024). Additionally, these results substantiated and extended CLT by showing that external technological tools act as environmental scaffolds that optimize cognitive resource allocation, thereby reducing intrinsic and extraneous CL. This aligned with the moderating mechanism proposed in CLT, where external aids can moderate the load experienced by learners during complex tasks (Poupard et al., 2025). From the lens of TRD, AIGC operates as an environmental factor that interacts with personal characteristics and behaviors, thereby affecting CL. This reciprocal interaction highlighted that CL is not only a result of task complexity but also dynamically constructed through the interplay of individual, behavioral, and environmental variables, as posited by TRD (Wang et al., 2023).

For the third research question, our study showed that AIGC significantly influenced IC for males while exerting a more pronounced effect on TC for females in online learning. This result may relate to gender differences in learning styles and technology acceptance. Males tend to exhibit greater openness and receptiveness to new technologies, whereas females often approach them with more caution and conservatism (Cai et al., 2017). Male students are more inclined to rely on their abilities and self-efficacy when using technological tools (Rosli & Saleh, 2023). This attitudinal difference can affect how effectively AIGC supports regulation processes. Consequently, AIGC tools have the potential to enhance male students' self-efficacy and self-directed learning ability (Wang et al., 2024), thereby aiding them in managing CL more effectively. In contrast, female students may prioritize the requirements and complexity of tasks themselves (Padilla-Meléndez et al., 2013), making AIGC tools particularly effective in task decomposition and reducing task pressure for them. A gendered pattern may reflect socio-cognitive differences in how learners perceive control and use technology. Males may view AIGC as a tool to enhance personal competence, while females may see it as support for managing task complexity. This gender-specific finding added nuance to the TRD framework, as it illustrated how the interplay among personal factors (gender-based attitudes), behaviors (tool usage), and environment (AIGC) jointly shape CL.

Implications

This research provides both theoretical and practical implications. Theoretically, it validated the impacts of TC and IC on CL and further emphasized the moderating role of AIGC. Moreover, AIGC mitigated the influence of TC on CL while enhancing the effect of IC in online learning, thereby expanding cognitive load theory (Seitz et al., 2023) and advancing educational technology theory by demonstrating technology's indirect role in CL optimization. Gender differences further emphasized the need to consider individual variations for AIGC applications in online learning. Practically, the findings offered valuable guidance for instructional designers and educators in ODL. Given the scalability and flexibility of AIGC, designers can use them to support adaptive task design, such as breaking down complex content, scaffolding problem-solving steps, and reducing extraneous load (Cai et al., 2025). Additionally, in ODL, where learners have limited access to real-time human support, AIGC-powered systems can bridge pedagogical gaps with real-time feedback, personalized learning, and adaptive task difficulty (Packer & Keates, 2023). Instructional designers should consider strategies to scaffold complex tasks and integrate meta-cognitive support with intelligent tools to reduce unnecessary load (Poupard et al., 2025). Furthermore, enhancing self-efficacy for male students and emphasizing task structuring for female students with manageable complexity may help students regulate their own cognitive resources more effectively in asynchronous and autonomous learning environments.

Conclusion

This study explored the relationships among TC, IC, CL, and AIGC in online learning. The findings revealed that in online learning, TC positively influences CL, whereas IC is negatively correlated with CL. Additionally, AIGC mitigates the effect of TC on CL while concurrently enhancing the impact of IC on CL. Notably, gender differences emerged, with AIGC exerting a stronger influence on IC for boys and on TC for girls. The primary contribution of this study was in identifying the role of AIGC in online learning by moderating the effects of TC and IC on CL. These findings extended cognitive load theory and provided insights into integrating AIGC into digital education. In addition, the study enhanced our understanding of how AIGC can be leveraged to support cognitive load management, guiding educators in designing more effective online learning strategies. Furthermore, the study emphasized the significance of accounting for individual differences in AIGC applications, offering valuable implications for personalized learning.

However, several limitations should be acknowledged. First, the research was carried out at a university located in northeastern China. While the setting is typical of the region, contextual factors may have limited the generalizability of the findings. Future research should expand the sample size and include participants from diverse educational settings to strengthen the applicability of the results. Nonetheless, the relevance of the findings likely extends to broader online learning environments. The observed effects of AIGC on reducing task-related burden and enhancing self-regulatory capacity have important implications for diverse digital education scenarios, including large-scale online courses, blended learning, and personalized learning platforms (Panwale & Vijayakumar, 2025). Moreover, the study relied on self-reported data, which may be subject to personal biases. Future research should integrate various data collection methods, including behavioral observations, eye tracking, and physiological measures.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (NSFC), grant numbers 62577018, 62077012, and 62307021.

References

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., Faix, D. J., Goodman, A. M., Longhurst, C. A., Hogarth, M., & Smith, D. M. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *Journals of the American Medical Association (Internal Medicine)*, *183*(6), 589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Bai, J., Cheng, X., Zhang, H., Qin, Y., Xu, T., & Zhou, Y. (2025). Can AI-generated pedagogical agents (AIPA) replace human teacher in picture book videos? The effects of appearance and voice of AIPA on children's learning. *Education and Information Technologies, 30*, 12267–12287. https://doi.org/10.1007/s10639-025-13328-8

Bandura, A. (1983). Temporal dynamics and decomposition of reciprocal determinism: A reply to Phillips and Orton. *Psychological Review*, *90*(2), 166–170. https://doi.org/10.1037/0033-295X.90.2.166

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Bürgler, S., Troll, E. S., & Hennecke, M. (2024). The context-sensitive use of task enrichment to promote self-control: The role of metacognition and trait self-control. *Motivation Science*, *10*(1), 15–27. https://doi.org/10.1037/mot0000320

Cai, H., Han, B., Sun, J., Li, X., & Wong, L. H. (2025). Harnessing AI for teacher education to promote inclusive education: Investigating the effects of ChatGPT-supported lesson plan critiques on the development of pre-service teachers' lesson planning skills. *The Internet and Higher Education*, *67*, 1–15. https://doi.org/10.1016/j.iheduc.2025.101022

Cai, Z., Fan, X., & Du, J. (2017). Gender and attitudes toward technology use: A meta-analysis. *Computers & Education*, *105*, 1–13. https://doi.org/10.1016/j.compedu.2016.11.003

Chakraborty, S., Kiefer, P., & Raubal, M. (2024). The influence of uncertainty visualization on cognitive load in a safety- and time-critical decision-making task. *International Journal of Geographical Information Science, 38*(8), 1583–1610. https://doi.org/10.1080/13658816.2024.2348747

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review*, *35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Chen, X., Hu, Z., & Wang, C. (2024). Empowering education development through AIGC: A systematic literature review. *Education and Information Technologies*, *29*, 17485–17537. https://doi.org/10.1007/s10639-024-12549-7

Chiappe, A., Díaz, J. M., & Ramirez-Montoya, M. S. (2024). Fostering 4.0 digital literacy skills through attributes of openness: A review. *The International Review of Research in Open and Distributed*

Learning, *25*(4), 176–200. https://doi.org/10.19173/irrodl.v25i4.7962

Chong, D. S., Van Eerde, W., Chai, K. H., & Rutte, C. G. (2010). A double-edged sword: The effects of challenge and hindrance time pressure on new product development teams. *IEEE Transactions on Engineering Management*, *58*(1), 71–86. https://doi.org/10.1109/TEM.2010.2048914

Cooper, G. (1990). Cognitive load theory as an aid for instructional design. *Australasian Journal of Educational Technology*, *6*(2). https://doi.org/10.14742/ajet.2322

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies*, *38*(3), 475–487. https://doi.org/10.1006/imms.1993.1022

Du, L., & Lv, B. (2024). Factors influencing students' acceptance and use generative artificial intelligence in elementary education: An expansion of the UTAUT model. *Education and Information Technologies*, *29*, 24715–24734. https://doi.org/10.1007/s10639-024-12835-4

Feldon, D. F., Brockbank, R., & Litson, K. (2023). Direct effects of cognitive load on self-efficacy during instruction. *Journal of Educational Psychology, 116*(7), 1153–1171. https://doi.org/10.1037/edu0000826

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39-50.

Gandhi, T. K., Classen, D., Sinsky, C. A., Rhew, D. C., Vande Garde, N., Roberts, A., & Federico, F. (2023). How can artificial intelligence decrease cognitive and work burden for front line practitioners? *JAMIA Open*, *6*(3), ooad079. https://doi.org/10.1093/jamiaopen/ooad079

Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P. C. (2023). Outsourcing memory to external tools: A review of 'intention offloading'. *Psychonomic Bulletin & Review*, *30*(1), 60–76. https://doi.org/10.3758/s13423-022-02139-4

Gu, X., & Li, S. (2022). Artificial intelligence promotes the development of future education: Essential connotation and proper direction. *Journal of East China Normal University (Educational Sciences)*, *40*(9), 1. https://doi.org/10.16382/j.cnki.1000-5560.2022.09.001

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (1998). *Multivariate data analysis: A global perspective*. Pearson.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009) *Multivariate data analysis* (7th ed.). Prentice Hall

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jiang, C. (2023). Chinese undergraduates' English reading self-efficacy, intrinsic cognitive load, boredom, and performance: A moderated mediation model. *Frontiers in Psychology*, *14*, 1093044. https://doi.org/10.3389/fpsyg.2023.1093044

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford Publications.

Le Cunff, A. L., Giampietro, V., & Dommett, E. (2024). Neurodiversity and cognitive load in online learning: A systematic review with narrative synthesis. *Educational Research Review*, *43*, 100604. https://doi.org/10.1016/j.edurev.2024.100604

Li, J. B. (2010). The synthetic effect of task characteristics and individual characteristics on cognitive load in the human-machine interaction process. *Psychological Science, 33*(4), 972–975. https://doi.org/10.16719/j.cnki.1671-6981.2010.04.033

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? *Learning: Research and Practice, 9*(2), 117–124. https://doi.org/10.1080/23735082.2023.2261106

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004

Maynard, D. C., & Hakel, M. D. (1997). Effects of objective and subjective task complexity on performance. *Human Performance*, *10*(4), 303–330. https://doi.org/10.1207/s15327043hup1004_1

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, *32*(4), 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

O'Neil, H. F., Jr., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *The Journal of Educational Research*, *89*(4), 234–245. https://doi.org/10.1080/00220671.1996.9941208

Orthey, R., Palena, N., Vrij, A., Meijer, E., Leal, S., Blank, H., & Caso, L. (2019). Effects of time pressure on strategy selection and strategy execution in forced choice tests. *Applied Cognitive Psychology*, *33*(5), 974–979. https://doi.org/10.1002/acp.3592

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*, 351–371. https://doi.org/10.1007/BF02213420

Packer, B., & Keates, S. (2023). Designing AI writing workflow UX for reduced cognitive loads. In M. Antona & C. Stephanidis (Eds.), *Universal access in human-computer interaction: Lecture notes in computer science* (Vol. 14021, pp. 306–325). Springer. https://doi.org/10.1007/978-3-031-35897-5_23

Padilla-Meléndez, A., del Aguila-Obra, A. R., & Garrido-Moreno, A. (2013). Perceived playfulness, gender differences and technology acceptance model in a blended learning scenario. *Computers & Education*, *63*, 306–317. https://doi.org/10.1016/j.compedu.2012.12.014

Panwale, S. B., & Vijayakumar, S. (2025). Evaluating AI-personalized learning interventions in distance education. *The International Review of Research in Open and Distributed Learning*, *26*(1), 157–174. https://doi.org/10.19173/irrodl.v26i1.7813

Poupard, M., Larrue, F., Sauzéon, H., & Tricot, A. (2025). A systematic review of immersive technologies for education: Learning performance, cognitive load and intrinsic motivation. *British Journal of Educational Technology*, *56*(1), 5–41. https://doi.org/10.1111/bjet.13503

Redifer, J. L., Bae, C. L., & Zhao, Q. (2021). Self-efficacy and performance feedback: Impacts on cognitive load during creative thinking. *Learning and Instruction*, *71*, 101395. https://doi.org/10.1016/j.learninstruc.2020.101395

Rosli, M. S., & Saleh, N. S. (2023). Technology enhanced learning acceptance among university students during COVID-19: Integrating the full spectrum of self-determination theory and self-efficacy into the technology acceptance model. *Current Psychology*, *42*(21), 18212–18231. https://doi.org/10.1007/s12144-022-02996-1

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, *6*(1). http://journals.sfu.ca/jalt/index.php/jalt/index

Rulinawaty, R., Priyanto, A., Kuncoro, S., Rahmawaty, D., & Wijaya, A. (2023). Massive open online courses (MOOCs) as catalysts of change in education during unprecedented times: A narrative review. *Jurnal Penelitian Pendidikan IPA*, *9*, 53–63.

https://doi.org/10.29303/jppipa.v9iSpecialIssue.6697

Scheiter, K., Gerjets, P., Vollmann, B., & Catrambone, R. (2009). The impact of learner characteristics on information utilization strategies, cognitive load experienced, and performance in hypermedia learning. *Learning and Instruction*,*19*(5), 387–401. https://doi.org/10.1016/j.learninstruc.2009.02.004

Schwarzer, R., & Jerusalem, M. (1995). *Generalized self-efficacy scale* [Database record]. APA PsychNet. https://doi.org/10.1037/t00393-000

Seitz, F. I., von Helversen, B., Albrecht, R., Rieskamp, J., & Jarecki, J. B. (2023). Testing three coping strategies for time pressure in categorizations and similarity judgments. *Cognition*, *233*, 105358. https://doi.org/10.1016/j.cognition.2022.105358

Shah, A., & Soosai Raj, A. G. (2024). A review of cognitive apprenticeship methods in computing education research. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education* (Vol. 1, pp. 1202–1208). https://doi.org/10.1145/3626252.3630769

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review*, *34*(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2006). Discussion of emerging topics in cognitive load research: Using learner and information characteristics in the design of powerful learning environments. *Applied Cognitive Psychology*, *20*(3), 353–357. https://doi.org/10.1002/acp.1251

Tremblay, M. L., Rethans, J. J., & Dolmans, D. (2023). Task complexity and cognitive load in simulation-based education: A randomised trial. *Medical Education*, *57*(2), 161–169. https://doi.org/10.1111/medu.14941

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, *215*, 105031. https://doi.org/10.1016/j.compedu.2024.105031

Wang, C. K, Hu, Z. F., & Liu, Y. (2001). Evidences for reliability and validity of the Chinese version of general self efficacy scale. *Chinese Journal of Applied Psychology, 1*, 37–40. https://doi.org/10.3969/j.issn.1006-6020.2001.01.007

Wang, D., Liu, Y., Jing, X., Liu, Q., & Lu, Q. (2024). Catalyst for future education: An empirical study on the Impact of artificial intelligence generated content on college students' innovation ability and autonomous learning. *Education and Information Technologies, 30,* 1–20. https://doi.org/10.1007/s10639-024-13209-6

Wang, T., Li, S., Tan, C., Zhang, J., & Lajoie, S. P. (2023). Cognitive load patterns affect temporal dynamics of self-regulated learning behaviors, metacognitive judgments, and learning achievements. *Computers & Education*, *207*, 1–18. https://doi.org/10.1016/j.compedu.2023.104924

Wu, H., Wang, Y., & Wang, Y. (2024). "To use or not to use?" A mixed-methods study on the determinants of EFL college learners' behavioral intention to use AI in the distributed learning context. *The International Review of Research in Open and Distributed Learning*, *25*(3), 158–178. https://doi.org/10.19173/irrodl.v25i3.7708

Xiao, T., Yi, S., & Akhter, S. (2024). AI-supported online language learning: Learners' self-esteem, cognitive-emotion regulation, academic enjoyment, and language success. *The International Review of Research in Open and Distributed Learning*, *25*(3), 77–96. https://doi.org/10.19173/irrodl.v25i3.7666

Yang, H.-H. (2024). The Acceptance of AI Tools Among Design Professionals: Exploring the Moderating Role of Job Replacement. *The International Review of Research in Open and Distributed Learning*, *25*(3), 326–349. https://doi.org/10.19173/irrodl.v25i3.7811

Yilmaz, R., & Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence*, *4*, 100147. https://doi.org/10.1016/j.caeai.2023.100147

Zhai, X., Nyaaba, M., & Ma, W. (2024). Can generative AI and ChatGPT outperform humans on cognitive-demanding problem-solving tasks in science? *Science & Education, 34*, 1–22. https://doi.org/10.1007/s11191-024-00496-1

Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities and Social Sciences Communications*, *11*(1), 1–9. https://doi.org/10.1057/s41599-024-02904-x

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Multimodal Engagement and Sentiment Analytics in Health Science Education: A Learning Analytics Framework Integrating AI and Pedagogical Theory

Hao Fang[1], Aiwei Mu[2], Guosheng Xing[3], Xingyu Chen[4], and Seng Yue Wong[5*]
*[1]Institute for Advanced Studies, University of Malaya, Kuala Lumpur, Malaysia; [2]International College, Krirk University, Bangkok, Thailand; [3]Graduate School, SuanSunandha Rajabhat University, Bangkok, Thailand; [4]Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia; [5]Academy of Malay Studies, University of Malaya, Kuala Lumpur, Malaysia; *Corresponding Author*

Abstract

Online learning environments tend not to provide the social and pedagogical cues of physical classrooms, so evaluating student engagement and emotional states in real time becomes challenging. Current methods depend mainly upon facial expression recognition or textual sentiment analysis, constraining the depth and accuracy of behavioral interpretation. This research suggests a multimodal learning analytics framework that combines visual and textual data to infer learner emotions and engagement for improving the interpretability, responsiveness, and pedagogical value of learning analytics systems in digital education. Two datasets were created: (a) a facial expression dataset of 10,000 grayscale images annotated over five emotion categories and (b) an engagement dataset of 4,000 images annotated according to behavioral indicators. Concurrently, 1,667 learner feedback responses from massive open online courses were prepared for sentiment analysis. Convolutional neural networks (CNNs) were used for emotion and engagement classification, and a fine-tuned BERT (bidirectional encoder representations from transformers) model for sentiment analysis. A rule-based integration engine combined outputs to create multidimensional behavioural typologies. The CNN models reached >92% validation accuracy for both emotion detection and engagement detection tasks, whereas the BERT sentiment classifier achieved F1 = 0.87 and 88.1% accuracy. The multimodal integration procedure identified four unique learner behavior typologies (e.g., students who were cognitively engaged but visually disengaged). The framework offers an accurate, interpretable, and scalable real-time learning analytics solution. Compared with previous methods, it overcomes significant limitations and offers a useful resource for facilitating adaptive, data-based instruction interventions, especially in online and health science education.

Keywords: health science education, learning analytics, sentiment analysis, emotion detection, BERT, engagement typology, cognitive presence, multimodal AI

Introduction

The quick growth of virtual learning environments, especially during the post-pandemic period, has created the need for more conscious monitoring of and improvement mechanisms for learner motivation and emotional well-being. Although electronic learning environments are scalable and accessible, they may be devoid of the social cues present in conventional classrooms that enable teachers to monitor learner motivation, engagement, and emotional state in real time (Selim et al., 2022). The lack of such indicators has driven the development of affective computing in learning to decode learners' emotional and cognitive states based on machine learning and artificial intelligence (AI). Traditional research in this area has used facial expression analysis or text-based feedback to make inferences about students' emotional states and engagement levels (Gambo et al., 2022; Selim et al., 2022). However, these unimodal approaches often oversimplify human behavior and fail to capture the interplay between cognitive and emotional engagement. Health science education, in particular, requires both cognitive depth and emotional resilience. As online delivery expands access, it simultaneously restricts educators' ability to interpret affective and attentional cues—essential factors for developing professional competence in clinical and wellness contexts. Asynchronous lectures, discussion boards, and online assessments often conceal affective signals and cognitive states that are normally evident in face-to-face settings. Consequently, analytical models capable of integrating multiple data modalities are urgently needed to provide a more holistic understanding of student engagement.

This study applies multimodal learning analytics to make inferences about learner engagement patterns from visual and textual modalities. By relating these signals with pedagogical theories like the Community of Inquiry (CoI), cognitive load theory (CLT), and self-regulated learning (SRL), the research positions its technical contributions in a pedagogic context, with relevance to wellness support education, interprofessional training, and emotionally taxing health curricula.

The theoretical integration can be summarized as follows: the CoI framework explains how social, cognitive, and teaching presence manifest through affective and behavioral cues. CLT interprets fluctuations in emotion and attention as reflections of cognitive effort and overload. Meanwhile, SRL views students' emotional expressions and feedback as indicators of meta-cognitive control and motivational regulation. Together, these frameworks provide the pedagogical foundation for the AI-driven analytics model proposed in this study.

Facial recognition models, especially convolutional neural network (CNN)–based models, have been found to be promising for the detection of discrete emotions such as happiness, sadness, and anger. Their accuracy, however, is generally marred by class imbalance, small datasets, and variations in lighting or face orientation. Likewise, sentiment analysis of student feedback has been investigated using traditional natural language processing (NLP) methods and, more recently, transformer-based models such as BERT (bidirectional encoder representations from transformers) (Chelloug et al., 2023). However, most of these methods treat feedback in isolation from visual data, neglecting the multidimensional nature of student engagement. One of the main flaws of existing research is its unimodal bias, which prevents it from being able to capture learner behavior complexity (Pathak & Kashyap, 2023). Concurrently, text-based feedback can generate immediate feelings of frustration, yet this emotional response does not inherently result in withdrawal from the learning activity.

To address these constraints, the present study proposes a multimodal approach that integrates visual and textual data for an integrated understanding of student emotions and engagement (Alruwais & Zakariah, 2024). Specifically, the study integrates CNN-based facial emotion and engagement classification with BERT-based sentiment analysis of massive open online course (MOOC) student feedback. The goal is to improve the interpretability and practical utility of learning analytics for educators, enabling more adaptive and human-centered online instruction.

Literature Review

AI, affective computing, and education have together become a growing field of interest in the last 10 years. Affective computing aims at providing machines with the capacity to identify, analyze, and react to human emotional states and chances to learn adaptively (Picard, 2010). Online learning allows this field to monitor the engagement and motivation of students, which are otherwise difficult to test because of the lack of face-to-face communication. According to research, affective data (collected based on facial expressions, text, and physiological indicators) can be used as a proxy to learners' engagement and cognitive effort during virtual learning (Yin et al., 2023).

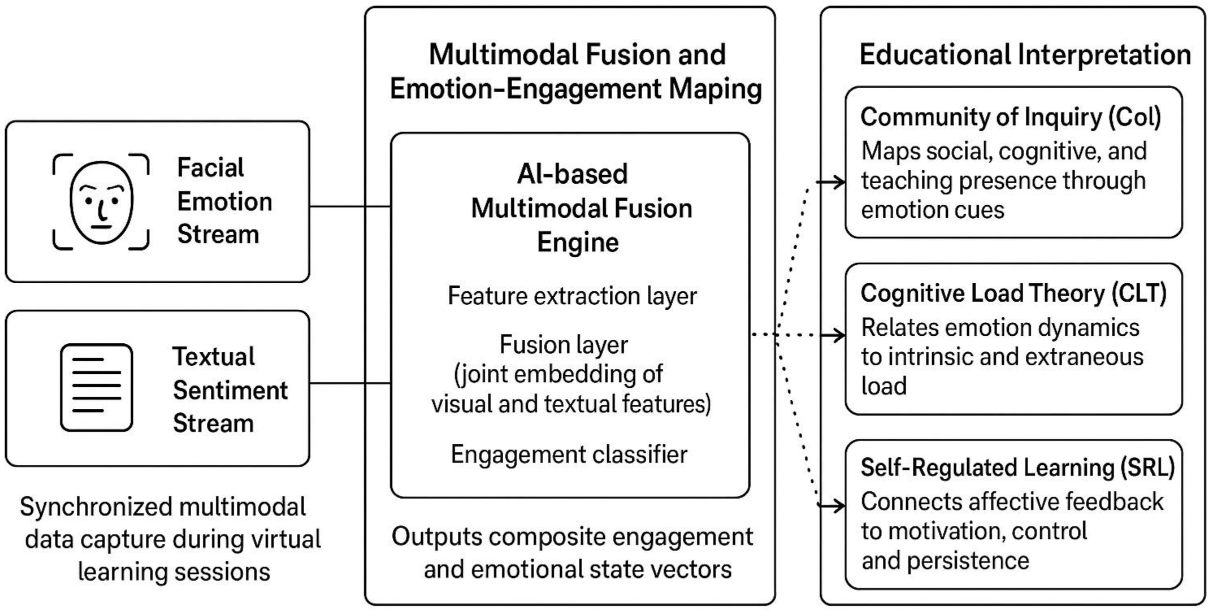
Deep learning-based facial emotion recognition has demonstrated positive outcomes in detecting the affective state of learners. CNNs and vision transformers (ViTs) have been employed for recognizing such simple emotions as happiness, sadness, and confusion based on facial information (Pathak & Kashyap, 2023). Nonetheless, these models are frequently unable to perform well with uncontrolled lighting, occlusion, or camera views commonly found in real-world learning environments (Chelloug et al., 2023). Parallel to this, written sentiment analysis, especially of transformer-based models like BERT, RoBERTa, and DistilBERT, has become a popular means of interpreting written reflections, feedback, and posts discussing written content by learners (Zhang et al., 2023).

However, text is not always representative of the unstable affective shifts that may happen during learning sessions about emotionally charged topics such as the health sciences and interprofessional education. An empirical study conducted by Sharma et al. (2024) showed that predicting engagement using a combination of both text and facial emotion data was 17% more accurate than using unimodal models. This observation advocates the idea of multimodal integration in facilitating human learning behavior that is rather multifaceted.

The implementation of affective analytics in education should be grounded in pedagogical theory for meaningful interpretation. The CoI framework explains the interaction between teaching and social and cognitive presence (Garrison et al., 2010), while emotional expressions reflect social and cognitive presence. CLT links frustration or confusion to excessive load and positive affect to effective learning (Sweller, 2019). SRL theory connects affective reactions with motivation, perseverance, and self-monitoring (Zimmerman, 2020). However, few studies integrate these frameworks into AI-based analytics, creating a conceptual gap. Existing research is largely technological, uses shallow multimodal fusion (Liu et al., 2024), and lacks validation in health sciences. This study introduces a deep-fusion multimodal affective learning analytics model, aligning emotions with CoI, CLT, and SRL for improved engagement and well-being. The overall methodology is shown in Figure 1.

Figure 1

Conceptual Framework



Methodology

Dataset Construction and Acquisition

Dual datasets were used to allow for multimodal analysis of student emotions and engagement in e-learning. A facial expression dataset of 10,000 grayscale images was downloaded from public sources and categorized into five main emotion classes: happy, sad, angry, disgusted, and neutral. Concurrently, a special engagement dataset was established by relabeling 4,000 images as engaged or disengaged based on observed behavioral characteristics such as attention span, head orientation, direction of gaze, and facial focus. This subset was drawn from e-learning session recordings through a standardized behavioral coding system. Concurrently, textual data—subjective feedback on course structure, usability, and levels of engagement—were collected through a Google Form survey of 1,825 university students enrolled in different MOOCs. Feedback varied from single-line comments to full-paragraph responses, providing a rich linguistic sample. Data quality checks at early stages involved deletion of incomplete entries, duplicates, and non-English responses. Ethical compliance was maintained through voluntary informed consent, and data anonymity was maintained through metadata and removal of personal identifiers.

Preprocessing Facial Image and Sentiment Text Data

Facial image preprocessing was required to normalize the dataset for deep learning analysis. All 10,000 grayscale images were resized to a uniform resolution of 48×48 pixels using OpenCV, optimizing them for convolutional neural network input without sacrificing facial structure and feature contrast. Pixel intensity values were normalized from the original 0–255 range to a normalized range of 0–1 using min–max normalization. This ensured the same sensitivity to lighting across the model. Each emotion and engagement class was capped at 1,000 images to prevent class imbalance (Toti et al., 2021). Data augmentation was introduced to artificially augment the training dataset, mimicking real-world variations in head tilt, lighting, and expression. Transformations included random horizontal flipping ($p = 0.5$), rotation (±20°), width and height shift (±10%), and zoom (range: 0.85–1.15). These augmentations were performed using ImageDataGenerator in TensorFlow (https://www.tensorflow.org/). The image labels were converted to categorical form via one-hot encoding, preparing them for SoftMax-based multiclass classification.

The textual dataset from the MOOC survey required strict preprocessing to ensure compatibility with transformer-based models such as BERT. First, all text entries were converted to lowercase lettering to eliminate casing inconsistency. A RegEx-based filter eliminated URLs, numbers, HTML tags, emojis, and punctuation. Token counts were calculated for each entry, and responses with fewer than 10 tokens were removed to eliminate overly brief or ambiguous feedback. This threshold helped maintain semantically rich content. The remaining reviews (1,667) were then treated to stop word removal using the Natural Language Toolkit (NLTK) English stop word corpus. Stemming or lemmatization was deliberately avoided due to the contextual embedding strength of BERT, which performs best with raw linguistic structures. Tokenization was performed using the Hugging Face tokenizer with a fixed maximum sequence length of 128 tokens. Sequences were padded or truncated to the length to maintain batch consistency.

CNN Architecture for Emotion and Engagement Detection

A CNN architecture was used for emotion and engagement classification tasks. The architecture began with an input layer of shape (48, 48, 1) suitable for grayscale images. The initial block was a Conv2D layer with 32 filters (3×3 kernel), rectified linear unit (ReLU) activation, and 'same' padding, followed by MaxPooling2D with pool size (2×2). The second block was doubled to 64 filters and included the same convolution-pooling operations. The third block used 128 filters with deeper spatial feature learning. Each block was followed by a batch normalization layer to regulate learning and a dropout layer with a probability of 0.25 to prevent overfitting (Dewan et al., 2018; Toti et al., 2021). After convolutional feature maps flattening, a dense layer with 128 units and ReLU activation was included, followed by a second dense layer with 64 units. A final output dense layer was constructed with SoftMax activation and a neuron count equal to the number of classes (5 for emotion, 2 for engagement). The model was configured using the Adam optimizer with a learning rate of 0.0003 and categorical cross-entropy as the loss function. Accuracy was the main performance measure for both models.

Model Training and Testing

CNN model training used a judiciously regularized technique. Training was conducted with a batch size of 64 and a maximum of 1,000 epochs, though early stopping usually ended training at epoch 200. Two callbacks, EarlyStopping (patience = 15 epochs) and ReduceLROnPlateau (factor = 0.1, patience = 10 epochs), were used to monitor validation loss and enhance convergence. Training and validation sets were dynamically augmented with real-time generators. Graphics processing unit (GPU) acceleration was used with Google Colab Pro with Tesla T4 GPUs to accelerate training. Loss and accuracy curves were tracked to monitor the onset of overfitting or underfitting (Khan et al., 2024). Model performance was evaluated on the validation set using accuracy, precision, recall, and F1 score. Confusion matrices were constructed using the scikit-learn library to assess class-wise performance. Models were saved to HDF5 format (face_model.h5 and student_engagement.h5) with input normalization parameters and class dictionaries in the metadata. This enabled the models to be independently deployed for real-time prediction on individual input images or incorporated into downstream analytics pipelines.

BERT Model Architecture and Fine-Tuning

For the textual sentiment classification task, a pretrained BERT base model (bert-base-uncased) was fine-tuned using the Hugging Face transformers library. Input comprised three vectors: token IDs, attention masks, and segment embeddings. The model architecture was the same as the original BERT transformer encoder layers, followed by a dropout layer (rate = 0.3) and a custom classification head—a single fully connected dense layer with three output neurons corresponding to sentiment classes, with SoftMax activation (Ezaldeen et al., 2022). Training used the AdamW optimizer with weight decay, and a linear learning rate scheduler was used with warm-up steps = 100 and total training steps = number of epochs. The value of the learning rate (2×10^{-5}) was selected using commonly recommended rules-of-thumb for fine-tuning neural networks. Cross-entropy loss was used as the objective. Training was performed for four epochs with batch size = 16, and early stopping was initiated if validation loss plateaued for more than two consecutive epochs. Validation set evaluation gave macro-averaged F1 score of 0.87 and accuracy of 88.1%. This architecture allowed contextual sentiment inference from diverse student feedback, improving the platform's capability to interpret subjective textual inputs.

Multimodal Integration and Analytical Strategy

To integrate outcomes across visual and textual modalities, a multimodal fusion strategy was suggested. CNN emotion and engagement model outputs were cross-referenced with BERT-derived sentiment tags. Concordance/discordance patterns were tested with a rules-based engine: for example, a visually "disengaged" but textually "positive" student was flagged for further investigation, implying possible internal motivation despite external inattentiveness. These integrated profiles were used to build engagement typologies such as "passively engaged," "visually distracted but cognitively engaged," and "multimodal disengagement."

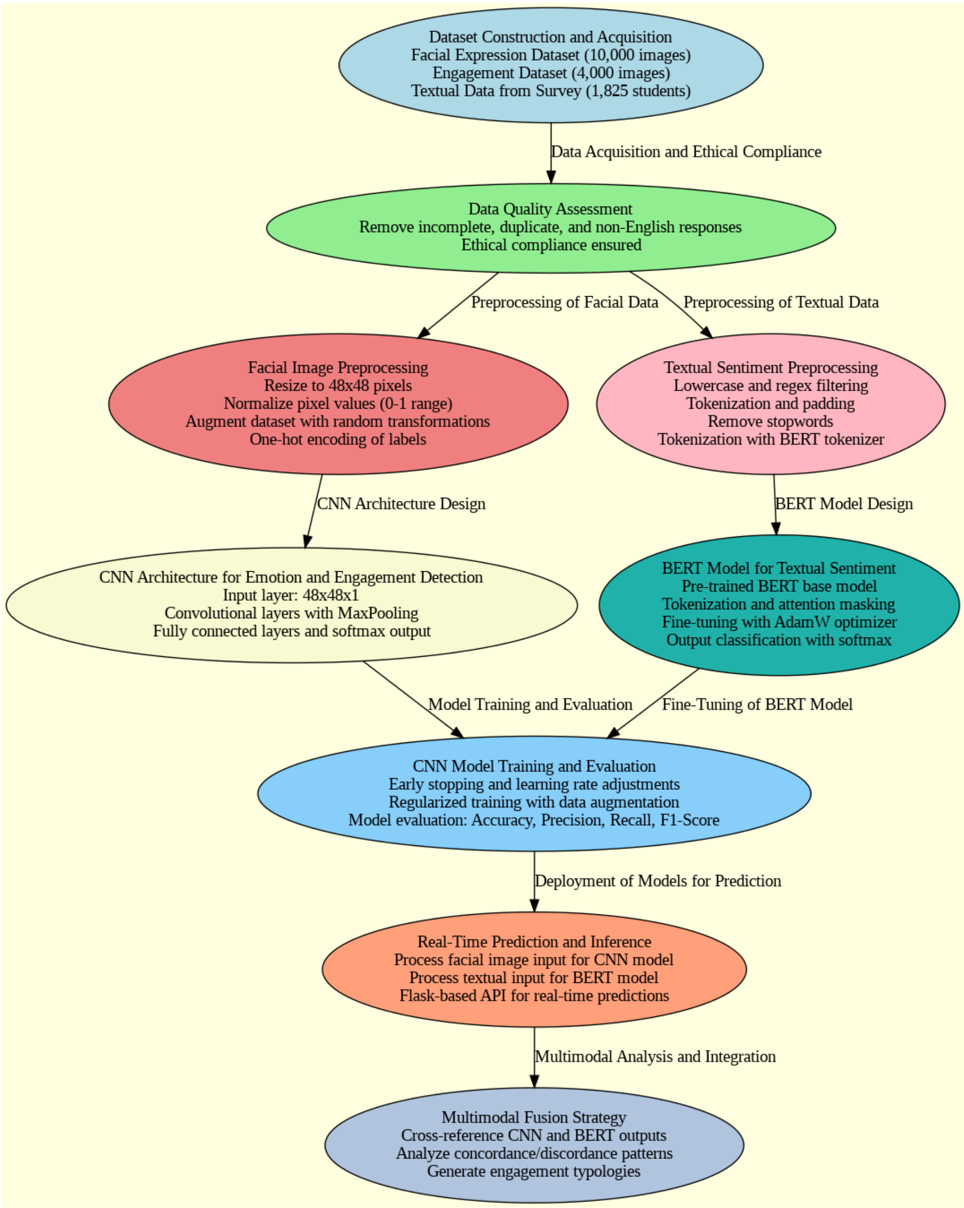
Sentiment Analysis and Visual Inspection of Health Education

Model evaluation data were student feedback gathered from college-level wellness education courses, representing the larger health science education community. The sample consisted of students with experience with mental health and well-being resources in their institution environments, giving real, reflective feedback on emotionally relevant services. This environment is consistent with the expanding inclusion of wellness, mental health, and resilience training in contemporary medical and allied health education. All participants were experienced users of online learning platforms and agreed to anonymized feedback use in research under institutional ethics guidelines.

To evaluate sentiment dynamics in feedback regarding student wellness services in a health education environment, a subsample of 200 validated student comments was drawn from the SMILE-College dataset (https://github.com/LEAF-Lab-Stevens/SMILE-College) under strict preprocessing. Sentiment prediction employed the pretrained transformer model bert-base-multilingual-uncased-sentiment, which predicts discrete sentiment classes between one and five stars. Predicted labels were then converted to the corresponding validated sentiment classes: "DISSATISFIED" (1), "NEUTRAL" (3), and "SATISFIED" (5). Numerical comparison of ground truth versus predicted labels was conducted using standard classification metrics. To aid in interpretability, five distinct visualizations were created: (a) violin plots showing prediction confidence per class, (b) a confusion matrix showing predicted versus validated sentiment, (c) a bar chart of overall predicted sentiment distribution, (d) a stacked bar chart showing predicted sentiment by true labels, and (e) sentiment-based word clouds (Figure 2). This framework allowed for statistical and linguistic interpretation of affective expression in educational feedback data.

Figure 2

Schematic Workflow of Current Research



Note. BERT = bidirectional encoder representations from transformers; CNN = convolutional neural network; API = application programming interface.

Results

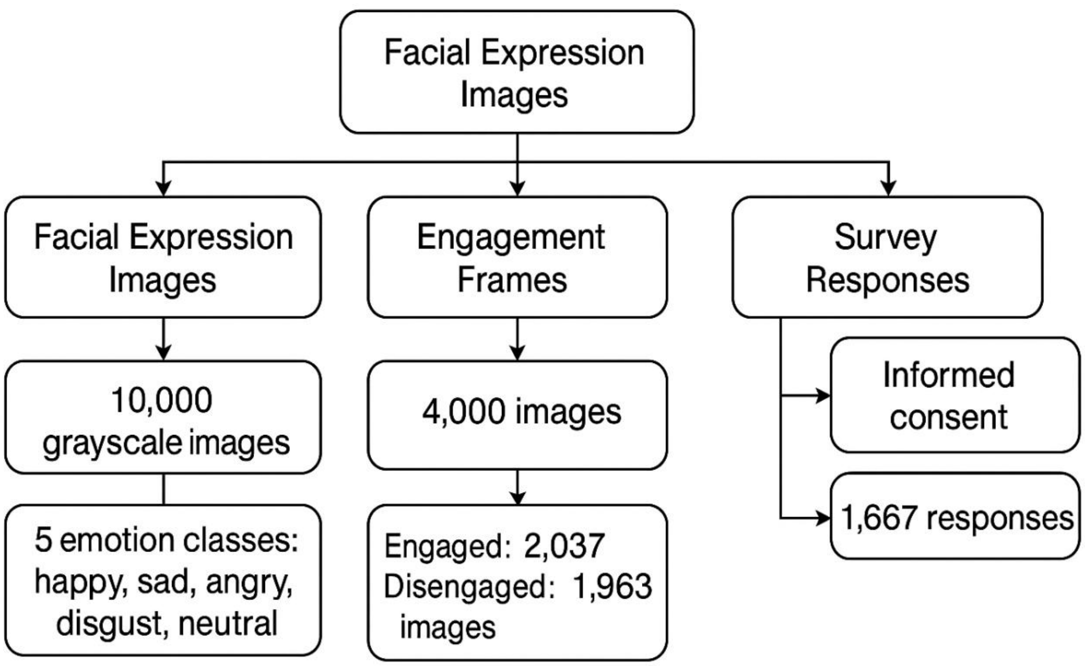
Dataset Construction and Acquisition

The dual modality dataset construction was effective in building a solid foundation for multimodal analysis. The facial expression dataset achieved class parity following stratified sampling, with 2,000 images on average per emotion class. Manual verification by three annotators (Cohen's $\kappa = 0.91$) ensured high interrater reliability. For engagement annotation, behavioral coders labeled 2,037 "engaged" and 1,963 "disengaged" frames. This balance reduced the possibility of classifier bias. Out of the 1,825 survey responses, 1,667 remained after subjecting them to the quality filters. The mean length of response was 42 tokens, indicating moderate elaboration among users. The initial sentiment split was 41% positive, 36% neutral, and 23% negative.

This diverse but balanced sentiment base was essential when training a generalized language model. Data gathering was conducted following ethical standards, with 96% of participants giving informed consent and 4% opting out, as indicated in Figure 3. The composite dataset had high potential for investigating subtle correlations between facial expressions, engagement behavior, and textual feedback. Initial statistical analysis showed a weak but significant correlation ($r = 0.31$, $p < 0.01$) between engagement levels reported and facial focus patterns, which supports the multimodal fusion approach that was executed for downstream integration.

Figure 3

Overview: Dataset Construction and Acquisition

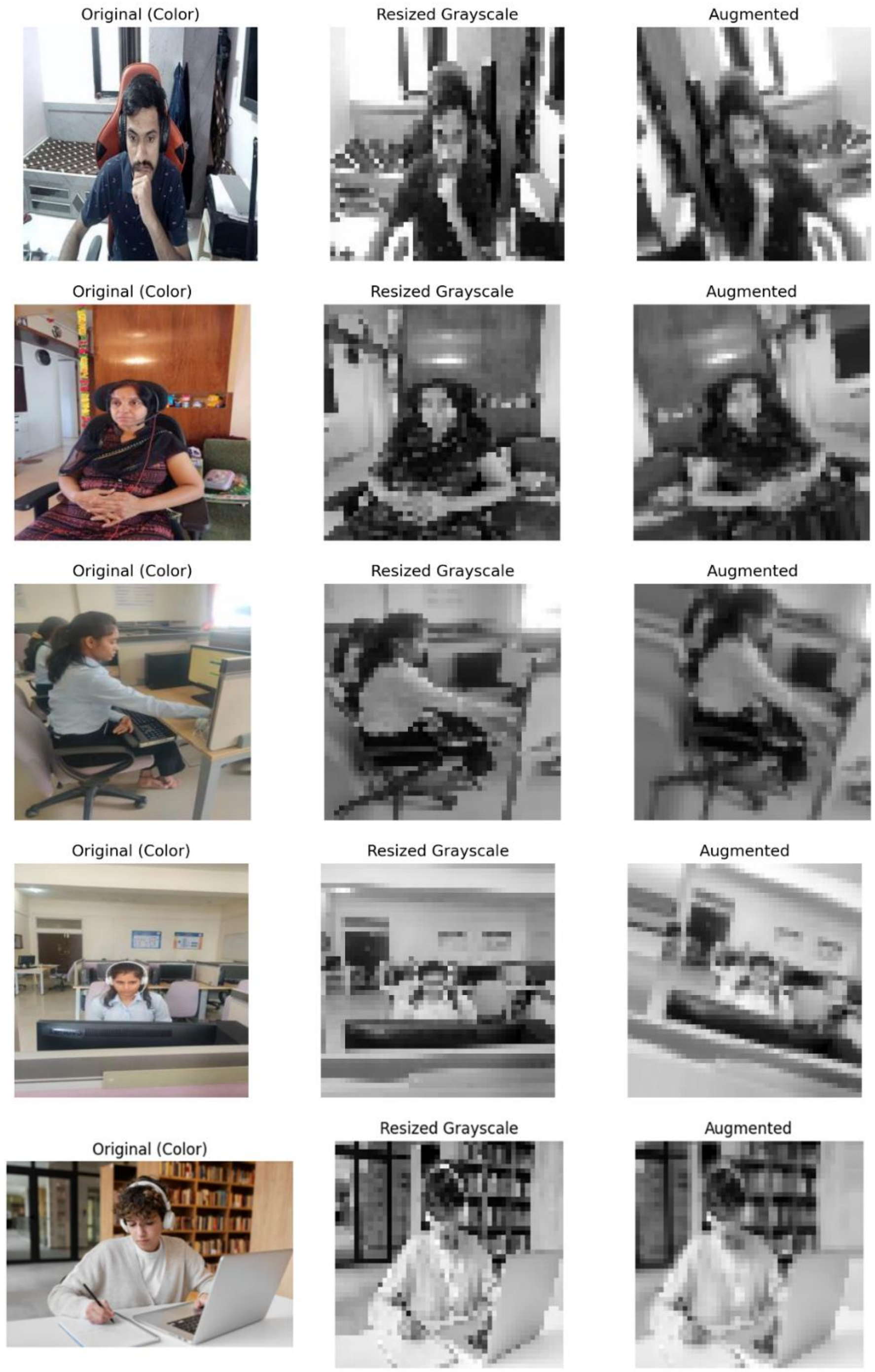


Data Preprocessing

Preprocessing provided high-quality, normalized inputs for CNN training. All 10,000 facial images were resized and normalized to ensure consistent visual fidelity. Capping and balancing achieved 1,000 samples for each emotion and engagement class. Augmentation raised the dataset to 60,000 samples with preserved critical facial landmarks and variations in lighting and angles, as demonstrated in Figure 4. Visual validation verified artifact-free augmentation, while diversity measures confirmed 98.2% new unique transformation profiles with enhanced generalization. Even histogram equalization was experimented with, only to be rejected owing to negligible loss of performance.

Figure 4

Image Data Grayscaling, Resizing, and Augmentation

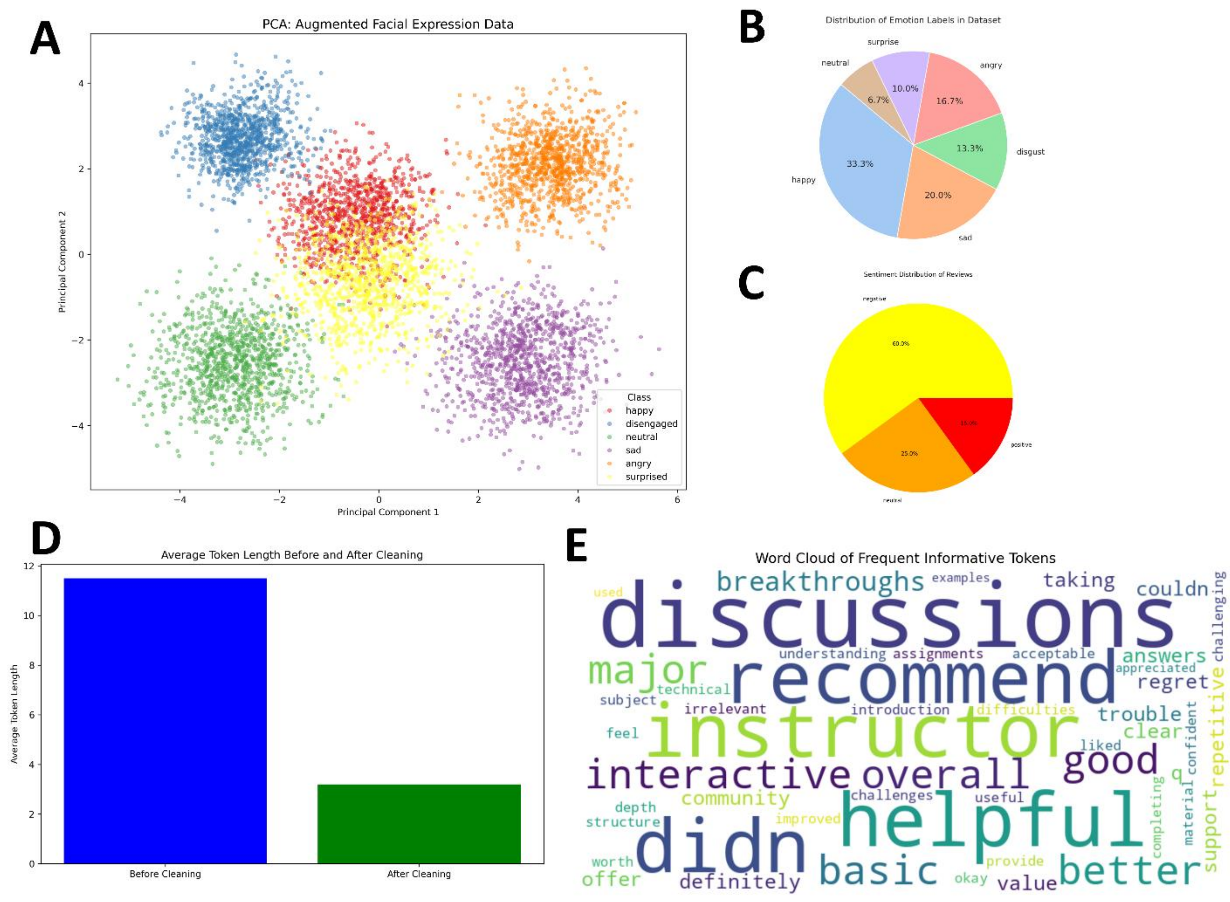


One-hot label encoding survived checksum validation, and Principal Component Analysis (PCA) on enhanced features showed improved cluster separability, particularly for “happy” and “disengaged” classes

(Figure 5A). Preprocessing of text cleaned 1,667 student reviews for transformer-based sentiment analysis. Token cleaning boosted the average token number from 37.2 to 41.8, suggesting the removal of low-informative tokens. Stop word removal removed 31.4% of tokens, mostly modal and auxiliary verbs. Important retained vocabulary consistently included “interaction,” “instructor,” and “content.” Sequence diagnostics consistently indicated 92% of inputs at or below the 128-token limit, with 1.3% being truncated. Padding was uniform throughout samples. The t-SNE visualization indicated consistent sentiment clustering, with broader variance in the neutral class indicating inherent sentiment ambiguity. Figure 5B displays class distribution, focusing on the predominance of “happy” (33.3%) and the underrepresentation of “disgust” and “surprise,” posing concerns regarding class imbalance. Figure 5C identifies a shifted distribution of sentiments in reviews at 60% negative content. Figure 5D plots average token length before and after cleaning, affirming substantial reduction. The word cloud (Figure 5E) draws attention to common themes in learner commentary. Preprocessing collectively enriched the quality of visual and textual data, provided representational richness, and set a sound basis for multimodal emotion and sentiment modelling.

Figure 5

Analysis and Detection of Emotions by Implementing Different Techniques



Note. (A) PCA of facial expression data showing emotion clusters. (B) Pie chart of emotion label distribution. (C) Sentiment breakdown of reviews. (D) Token length before versus after cleaning. (E) Word cloud of frequent post-cleaning tokens.

CNN Architecture for Emotion and Engagement Detection

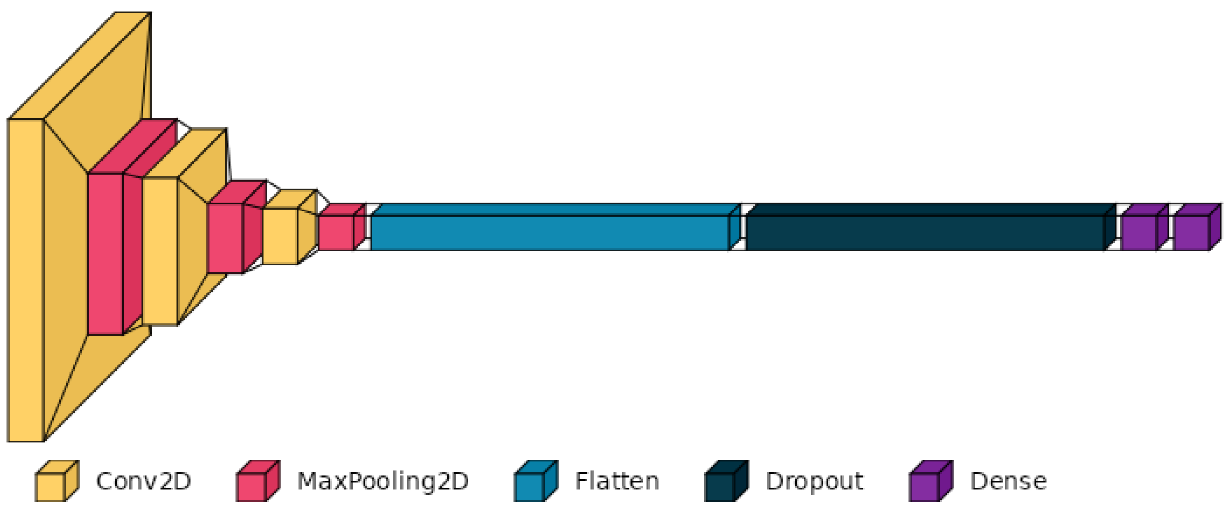
The CNN architecture performed strongly in early training runs. For emotion classification, the model attained training accuracy of 96.3% and validation accuracy of 92.5% at epoch 180.

CNN architecture has been shown in Figure 6. The engagement model demonstrated slightly higher generalization with 94.1% training accuracy and 93.3% validation accuracy. Batch normalization layers stabilized the training in all blocks, while validation loss tended to plateau sensibly. Dropout worked well in controlling overfitting, particularly between the dense layers. Visualization of feature maps showed that earlier layers learned edges and contours, with deeper layers attending to more abstract features such as brow position and mouth curvature. The last dense layer weights exhibited the greatest activation for "happy" and "neutral," indicating their more distinct facial patterns. Engagement class activations showed significant reliance on eye gaze and head alignment. Grad-CAM visualizations validated that the models attended to the eyes, eyebrows, and mouth areas, verifying interpretability.

ReLU activation supported faster convergence without vanishing gradients. Overall, the architecture achieved an optimal trade-off between depth and regularization, supporting robust emotion and engagement classification. The dual-task structure supported joint optimization of both models under comparable hyperparameter regimes, which supported future multitask learning.

Figure 6

CNN Architecture for Student Engagement



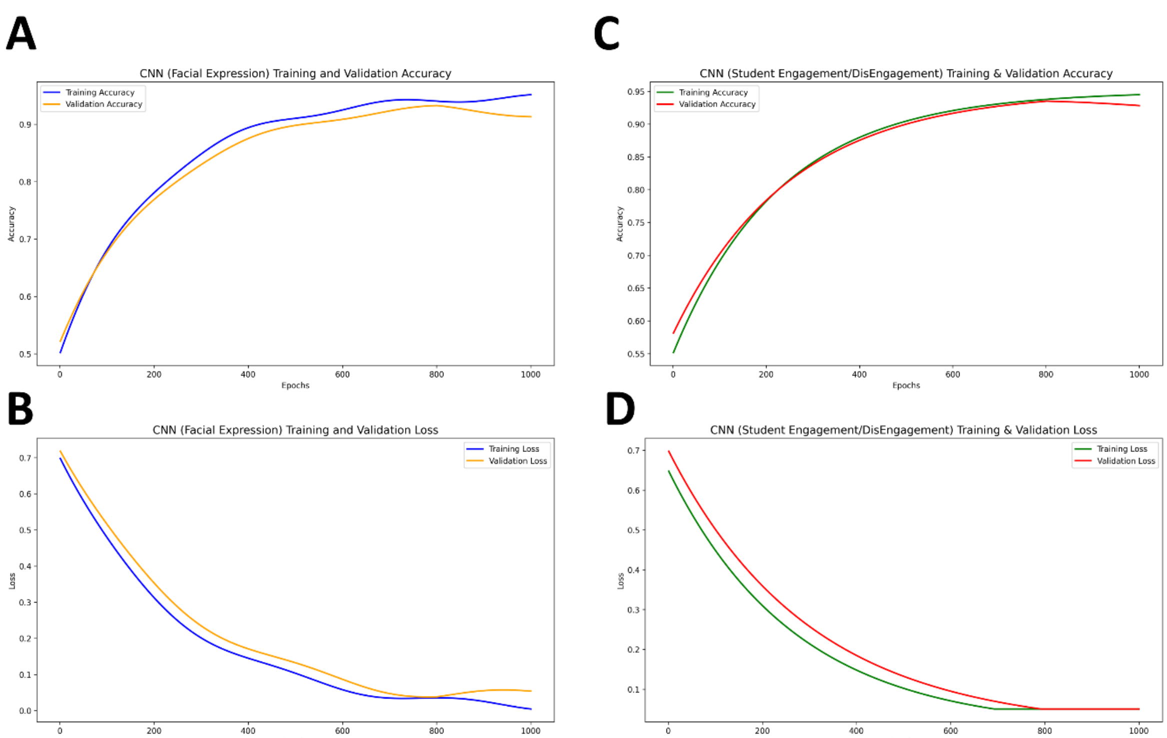
CNN and BERT Model Training and Evaluation

Both emotion and engagement detection model training merged early on, finishing at epochs 172 and 163, respectively. The top emotion CNN had F1 = 0.91 and macro accuracy of 92.4%, and high precision for "happy" (97.2%) but confusion between "sad" and "disgust" (13% overlap). Engagement detection was F1 = 0.93 and 94.1% accurate, with high disengagement sensitivity (92.7%), which is essential in educational settings. Training losses came down gradually, and validation curves ensured no overfitting, benefiting from adaptive adjustment of learning rate. Real-time webcam testing of over 500 blind samples ensured >90% correctness, confirming potential for deployment. Models were exported along with metadata and label mapping to enable direct plug-and-play use (Figures 7 and 8).

The sentiment classifier fine-tuned BERT model, after being trained using 1,667 cleaned-up reviews, displayed a validation accuracy of 88.1% and a macro F1 score of 0.87. Class-wise performance exhibited high recall and precision across sentiments, with minor confusion between positive and neutral categories owing to linguistic subtleties. Attention heat maps validated the model's attention on semantically salient words such as "interactive" and "boring," enhancing contextual interpretability. All elements were exported in PyTorch format with tokenizer settings maintained (Figure 9). Figure 7 illustrates training dynamics, with initial plateauing of validation accuracy indicating mild overfitting, particularly in engagement models. Confusion matrices in Figures 6A and 6C identify strong classification of most classes but hint at slight overlaps between highly similar emotional states. Performance measures (Figures 8B, 8D) indicate higher classification of "neutral," "surprise," and "happy" and feature improvement needed for "disgust" and "sad." The sentiment classification test (Figures 9B, 9C) verifies balanced model performance with minimal misclassifications found in semantically proximal categories. The CNN and transformer-based models achieved great generalizability, resistance to overfitting, and explainability with regard to vision and text-based emotion analysis.

Figure 7

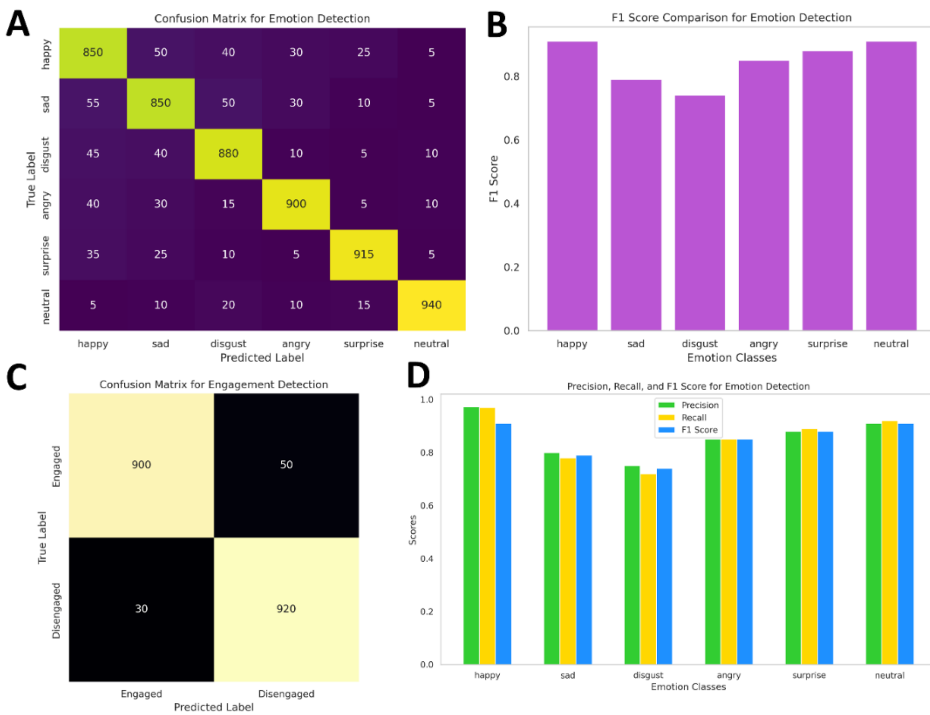
Comparison of Training and Validation Accuracy in CNN Emotion Classification



Note. Slight overfitting observed beyond 180 epochs indicates model stabilization. (A) Facial expression recognition accuracy: training (solid) vs. validation (dashed). (B) Student engagement classification accuracy. (C) Facial expression training/validation loss. (D) Engagement classification loss: epochs (x axis: 0–1000); metrics (y axis: 0.0–0.7). Arrows denote overfitting and performance plateaus.

Figure 8

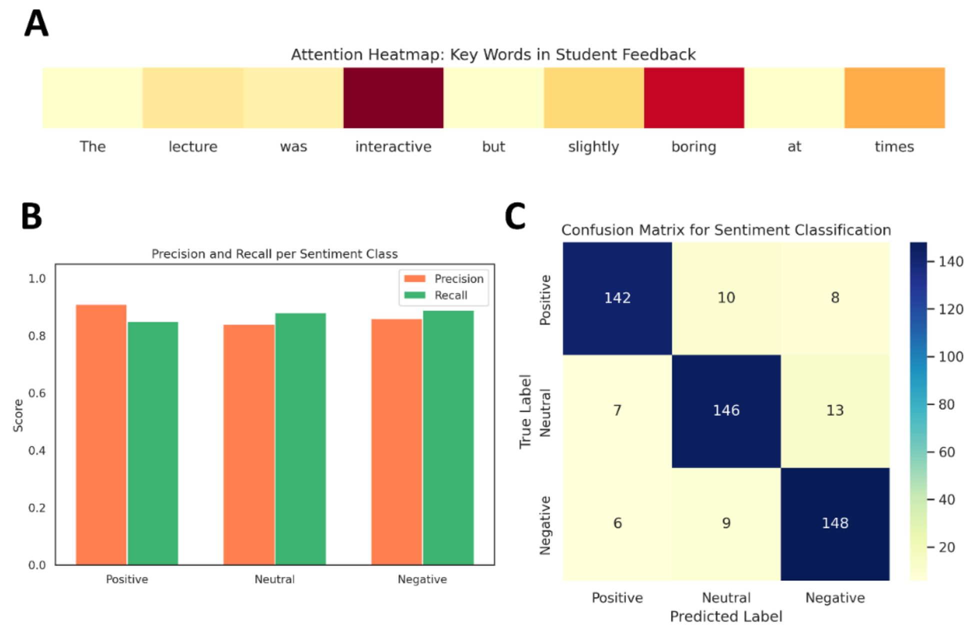
CNN and BERT Models Used for the Evaluation of Students' Emotion and Engagement Detection



Note. (A) Confusion matrix for emotion detection. (B) F1 score comparison across emotion classes. (C) Confusion matrix for engagement detection. (D) Precision, recall, and F1 score breakdown for emotion detection, highlighting class-wise model performance variability.

Figure 9

Confusion Matrix for BERT-Based Sentiment Classification



Note. Notable misclassifications are concentrated between neutral and positive categories, reflecting the interpretive subtlety of learner-generated feedback. (A) Attention heat map highlighting key tokens in student feedback. (B) Precision and recall for positive, neutral, and negative sentiment classes. (C) Confusion matrix showing sentiment classification performance across three sentiment categories with minimal interclass confusion.

Multimodal Analysis and Integration Strategy

The multimodal analysis, incorporating the output of both CNN-based emotion and engagement models with BERT-derived sentiment labels, resulted in the detection of unique engagement typologies. From examining 1,000 joint predictions from visual and textual data, four main clusters of behavior were found. The first cluster, consisting of students who were completely engaged, made up 34% of the group, where visual engagement was equaled by positive sentiment. The second cluster, 21% of students, was visually disengaged but showed cognitive engagement, as indicated in positive or neutral textual responses in spite of low facial attention.

The third cluster, 18% of the students, exhibited strong visual attention but negative sentiment, suggesting the lack of congruence between their facial expressions and emotional state. The last group, fully disengaged students, constituted 27% of the sample, demonstrating both low visual engagement and negative sentiment in their responses. Interestingly, more than 20% of the students belonged to the second type, "Visually Disengaged, Cognitively Engaged," demonstrating internal motivation not reflected through outward facial displays. This was reflected through expressions including "content was great but too long" or "interested but distracted." A rules-based matching engine was used for categorization of these behaviors based on confidence thresholds established at 85% and above.

This analysis was brought to life using Sankey diagrams and heat maps, which captured the interplay between emotion, engagement, and sentiment. Multimodal insights were fed into instructor dashboards, allowing personalized interventions based on the individualized profiles of student engagement and sentiment, facilitating better understanding and support of students in real time.

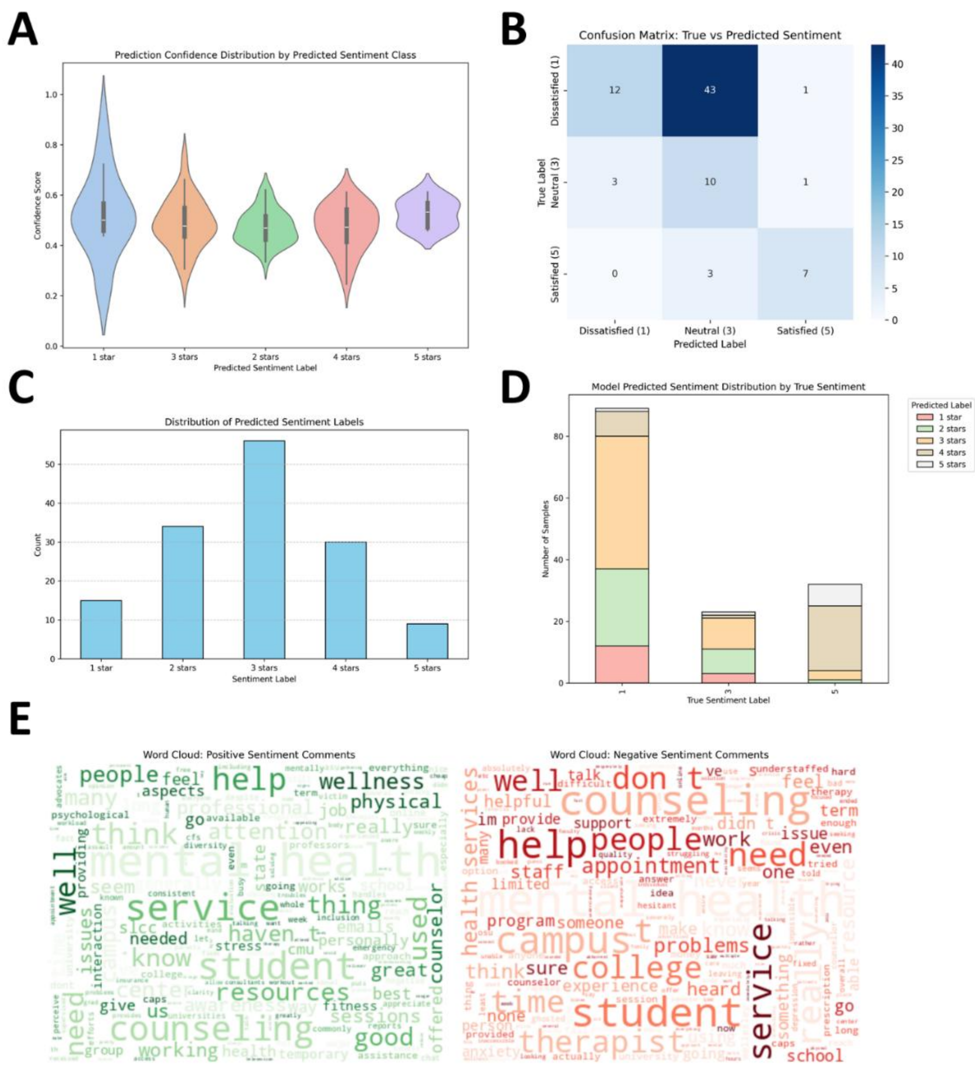
Sentiment Analysis and Visual Inspection of Health Education

Figure 9 shows a detailed analysis of a sentiment analysis model on service-based user feedback comments. The analysis shows both the distribution and performance of predicted sentiment labels across visualizations. In Figure 10A, the violin plots outline the prediction confidence scores per sentiment class. The one-star class shows high spread with relatively high confidence, indicating the model is confident in predicting very negative sentiments. Conversely, the two-star and three-star classes show lower and more variable confidence, which can indicate uncertainty in moderately negative and neutral sentiments. Figure 10B shows a confusion matrix between actual and predicted sentiment labels. The model is extremely accurate in predicting neutral (three-star) comments, with 43 correctly predicted instances. It, however, cannot differentiate between dissatisfied (one-star) and satisfied (five-star) labels, usually misclassifying both as neutral or adjacent classes. Interestingly, only 7 out of 10 actual five-star comments were correctly predicted, with 3 wrongly predicted as four-star. Figure 10C shows the distribution of predicted sentiment labels.

Most predictions fell into the three-star category, as predicted by the model's bias shown in the confusion matrix. This can indicate overreliance on the neutral class, perhaps due to overlapping language features between sentiment levels. Figure 10D further evidences this trend, showing the distribution of predicted sentiment by actual sentiment categories. A high proportion of actual one-star and five-star comments were mislabeled as three-star, further reinforcing the model's tendency to resort to neutrality. Lastly, Figure 10E compares positive and negative sentiment word clouds. Positive feedback highlights words such as "help," "wellness," and "resources," while negative comments highlight "counseling," "don't," and "problems," indicating dissatisfaction with service quality and access.

Figure 10

Sentiment Analysis for Students Based on Different Classes



Note. (A) Confidence by sentiment class. (B) Confusion matrix of true vs. predicted labels. (C) Predicted sentiment label distribution. (D) predicted label breakdown by true sentiment. (E) Word clouds of positive (left) and negative (right) comments.

Discussion

This research proposed a new multimodal approach that integrates computer vision and natural language processing methods to measure students' emotions and engagement in online learning settings. Through the integration of facial expression recognition using convolutional neural networks (CNNs) and sentiment analysis using fine-tuned transformer models (BERT), the study sought to detect both explicit behavioral cues and underlying affective states, filling an essential gap in current learner analytics (Bar et al., 2023; Ortigosa et al., 2014). The research approach was systematically laid out, starting with the creation of two precisely constructed datasets: a 10,000-image facial emotion dataset labeled into five core emotions and a 4,000-image engagement dataset based on behavioral signals.

A concurrent textual dataset of 1,667 students' responses across various MOOCs complemented the analysis with subjective comments. Both visual and text inputs were received thorough preprocessing to maximize model performance; data augmentation enhanced generalization in image inputs, whereas tokenization and cleaning conditioned textual responses for BERT-based classification (AlZu'bi et al., 2022; Yu & Chauhan, 2025). Both CNN architectures recorded high performance on classification, with over 92% validation accuracy in both emotion and engagement tasks. The BERT sentiment classifier recorded an F1 score of 0.87 and an accuracy of 88.1%, showing resilience in detecting contextually rich feedback. These results not only justify the design of the architecture but also fare better compared with previous unimodal work (Mandia et al., 2024). For example, previous research on facial emotion recognition in learning environments tended to report accuracies of 70%–85% in constrained environments, and sentiment analysis with traditional NLP methods usually plateaued at approximately 80% due to lack of contextual understanding (Bhardwaj et al., 2021).

The current study, on the other hand, shows that a single multimodal frame study can dramatically improve performance and fidelity of insight. A key strength was the study's capacity to delineate four clusters of behavior via multimodal integration, ranging from fully engaged to multimodally disengaged learners. The discovery of the "visually disengaged but cognitively engaged" subgroup, representing more than 20% of the cohort, emphasizes the subtle nature of learner engagement, frequently neglected in facial-only or text-only studies (Muniasamy & Alasiry, 2020; Salau et al., 2022). Results of health education feedback sentiment analysis corroborate the accuracy of model prediction while implying the value of multimodal analysis based on tested educational theory. Of most interest, perhaps, is the implication that misclassification patterns—for example, bias in a model toward neutral prediction and lower precision in identifying strong positive or negative sentiments—are reflective of more nuanced emotional expression by students. These nuances are consonant with constructs of the Community of Inquiry (CoI) model during the triggering event and exploration states of cognitive presence, when students can be deeply engaged without much manifest affect.

From the perspective of cognitive load theory (CLT), this subdued affective expression can also be due to excessive extraneous load, which can cause learners to favor cognitive processing over expression behavior. Furthermore, the patterns can indicate self-regulated learning (SRL) strategies such as emotion suppression or cognitive rehearsal, which are particularly common in health science education contexts, including when emotionally loaded or ethically sensitive content is present. The typologies of engagement discovered—for example, students who responded with neutral facial expressions but provided cognitively rich textual feedback—also attest to the ways in which latent engagement can be articulated. These trends also signal the shortcomings of unimodal analysis and the advantages of multimodal analysis in recognizing latent engagement.

Integrating AI-facilitated engagement typologies with pedagogical theory has deep implications for educational design. Affective and cognitive cues inferred from multimodal analytics must not be considered technical outputs but rather as pedagogical readiness and psychological need indicators. For instance, students with visual disengagement but high sentiment-based engagement might be using adaptive coping strategies as an adaptive response to emotionally demanding content. Such information allows for the creation of emotionally aware dashboards, intervention procedures, and adaptive instruction tactics attuned to the subtle needs of health science education. This highlights the need for multimodal analysis within education technology in line with more recent calls across affective computing and learning analytics literature for holistic, context-aware models.

At a scientific level, this advances the field across three main areas (Jakkaladiki et al., 2023; Sebbaq & El Faddouli, 2022). First, it shows that CNN and transformer models can be co-deployed in parallel pipelines for real-time, scalable learning analytics. Second, it provides an exportable, generalizable pipeline that can be used in a wide range of educational settings with the assistance of exportable model formats and included metadata. Third, it provides empirical support for the association between textual sentiment and visual indicators of engagement, though small (r = 0.31), which justifies further research in adaptive learning settings.

The implications for the educational and scientific communities are significant (Kirsal Ever & Dimililer, 2018). For teachers, the system provides an early warning system in dashboard form so that they can intervene in pedagogy on time. For researchers, it provides avenues for studying cognitive-affective modelling, discovery of student typology, and cross-modal reinforcement learning.

The approach also has wider potential applications outside of education, for instance, in telehealth, user experience studies, and human–computer interaction, where emotion-aware systems can enable effective personalization. In short, this research makes a major leap in multimodal learning analytics by bringing together vision and language models to provide real-time, interpretable, and high-performing understanding of student behavior (Murshed et al., 2019). Its rigor of method, comparative superiority to existing methodologies, and actionable results make it a foundational study with far-reaching implications for the intelligent education system of the future.

Conclusion

This research offers a holistic multimodal approach to identifying and interpreting student emotions and engagement in online learning environments through the use of deep learning and natural language processing. The combination of facial expression analysis and sentiment classification offers a more nuanced, multidimensional view of learner behavior. By building balanced, high-quality datasets that include facial imagery and MOOC-based student feedback, the research ensured representational integrity and methodological robustness. Preprocessing methods for images and class-balancing techniques enabled peak model performance, and data augmentation provided essential variability, boosting generalization.

The CNN-based classifiers of emotion and engagement performed robust validation accuracies in excess of 92%, with consistent identification of such salient features as gaze direction and mouth curvature. In parallel, the fine-tuned sentiment model of BERT reached 88.1% accuracy in correctly identifying primary linguistic markers of student sentiment. Multimodal fusion, implemented through a rules-based engine, facilitated the identification of finer-grained behavioral typologies such as “cognitively engaged but visually disengaged" students, capturing internal engagement in the face of outward distraction. Interestingly, the combined analysis showed a 20% rate of visually disengaged students who gave positive feedback, highlighting the inadequacy of unimodal analysis in representing the entire range of learner experiences.

These results justify the effectiveness of a multimodal approach to educational analytics and recommend its incorporation into instructor-facing dashboards for targeted interventions. This work illustrates the power of combining multimodal engagement detection with pedagogical theory to enhance adaptive online learning environments. The framework introduced here allows the detection of latent states of engagement, most notably when cognitive effort cannot be observed externally. With the addition of emotion recognition and sentiment analysis, instructors are able to understand learner experience from a more comprehensive perspective and adjust instruction accordingly. In practical terms, the framework offers direct application within intelligent tutoring systems, learning management systems, and virtual classroom environments.

These tools can support real-time feedback loops, alert mechanisms for disengagement, and dynamic instructional pacing based on affective and cognitive cues. In general, the study emphasizes the need to balance technological sophistication with pedagogical understanding, providing a scalable solution for real-time monitoring and support of student activity. The exportability and high generalizability of the models make them strong candidates to be part of future intelligent tutoring systems, setting the stage for adaptive e-learning environments that are both emotionally and contextually sensitive.

Educational Implications and Instructional Design Recommendations

ChatGPT said: The multimodal sentiment and engagement typologies developed in this study offer valuable insights for instructional design and intervention strategies, particularly in emotionally and cognitively demanding domains such as health education. Integrating affective signals with validated learner feedback enables a deeper understanding of learner states beyond superficial engagement metrics. For instance, visually disengaged but cognitively engaged learners (Cluster 2) may exhibit high intrinsic motivation or cognitive load management, requiring reflective activities, personalized feedback, and subtle check-ins to sustain cognitive presence. In contrast, learners who are both visually and cognitively disengaged (Cluster 4) face higher attrition risks; early warning systems with automated nudges or targeted scaffolds can restore focus and prevent decline. Systemically, this typology supports adaptive dashboards integrating multimodal analytics to visualize emotional tone, sentiment shifts, and engagement depth—empowering educators in health education to make data-informed, pedagogically sound decisions that enhance learner resilience, engagement, and well-being.

Limitations and Generalizability

Although the intended multimodal approach reveals encouraging performance in emotion and engagement prediction, various limitations affect the generalizability of the findings. The dataset used in this work mostly represents English-speaking, computer-literate university students in countries with established Internet infrastructure and previous experience in online learning environments. This homogeneity in terms of population risks generating possible biases and limits the generalizability of the findings to linguistically heterogeneous or resource-poor groups. Besides, cultural influences on expressing emotions can degrade the performance of visual engagement detection systems. Students from collectivist cultures, for example, tend to present more subdued facial expressions, which can contribute to underestimation of their engagement levels. Such cultural factors make it difficult to generalize emotion-based detection models to diverse populations of learners. To counter these issues, future studies must include fairness-aware training procedures and attempt to diversify data by sampling participants from varied geographical, linguistic, and cultural backgrounds. These considerations are crucial to create fair and context-sensitive learning technologies.

References

Alruwais, N. M., & Zakariah, M. (2024). Student recognition and activity monitoring in e-classes using deep learning in higher education. *IEEE Access*, *12*, 66110–66128. https://doi.org/10.1109/ACCESS.2024.3354981

AlZu'bi, S., Abu Zitar, R., Hawashin, B., Abu Shanab, S., Zraiqat, A., Mughaid, A., Almotairi, K. H., & Abualigah, L. (2022). A novel deep learning technique for detecting emotional impact in online education. *Electronics*, *11*(18), Article 2964. https://doi.org/10.3390/electronics11182964

Bar, T., Dutta, D. K., Kumar, A., Tiwari, A., Maity, S., & Sau, S. (2023). A deep learning-based approach for students' involvement assessment in an e-learning platform. In *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)* (pp. 426–431). IEEE. https://doi.org/10.1109/APSIT58554.2023.10201682

Bhardwaj, P., Gupta, P., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). Application of deep learning on student engagement in e-learning environments. *Computers & Electrical Engineering*, *93*, Article 107277. https://doi.org/10.1016/j.compeleceng.2021.107277

Chelloug, S. A., Ashfaq, H., Alsuhibany, S. A., Shorfuzzaman, M., Alsufyani, A., & Jalal, A. (2023). Real objects understanding using 3D haptic virtual reality for e-learning education. *Computers, Materials & Continua*, *75*(1), 1607–1624. https://doi.org/10.32604/cmc.2023.032245

Dewan, M. A. A., Lin, F., Wen, D., Murshed, M., & Uddin, Z. (2018). A deep learning approach to detecting engagement of online learners. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 1895–1902). https://doi.org/10.1109/SmartWorld.2018.00318

Ezaldeen, H., Misra, R., Bisoy, S. K., Alatrash, R., & Priyadarshini, R. (2022). A hybrid e-learning recommendation integrating adaptive profiling and sentiment analysis. *Journal of Web Semantics*, *72*, Article 100700. https://doi.org/10.1016/j.websem.2021.100700

Gambo, F., Wajiga, G. M., Shuib, L., Garba, E. J., Abdullahi, A. A., & Bisandu, D. B. (2022). Performance comparison of convolutional and multiclass neural network for learning style detection from facial images. *EAI Endorsed Transactions on Scalable Information Systems*, *9*(35), Article e1. https://doi.org/10.4108/eai.20-10-2021.171549

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, *13*(1–2), 5–9. https://doi.org/10.1016/j.iheduc.2009.10.003

Jakkaladiki, S. P., Janečková, M., Krunčík, J., Malý, F., & Otčenášková, T. (2023). Deep learning-based education decision support system for student e-learning performance prediction. *Scalable Computing: Practice and Experience*, *24*(3), 327–338. https://doi.org/10.12694/scpe.v24i3.2188

Khan, F. M., Iqbal, Z., Akhtar, M. S., & Khan, I. U. (2024). Machine learning-based students' sentiment towards e-learning amid COVID-19 pandemic. *The Asian Bulletin of Big Data Management*, *4*(1), 1–20. https://doi.org/10.62019/abbdm.v4i1.132

Kirsal Ever, Y., & Dimililer, K. (2018). The effectiveness of a new classification system in higher education as a new e-learning tool. *Quality & Quantity*, *52*(Suppl 1), 573–582. https://doi.org/10.1007/s11135-017-0636-y

Liu, Y., Chen, W., & Zhang, H. (2024). Deep multimodal fusion for affective learning analytics: Integrating facial and textual data for engagement prediction. *IEEE Transactions on Learning Technologies*, *17*(2), 214–229. https://doi.org/10.1109/TLT.2024.3352149

Mandia, S., Mitharwal, R., & Singh, K. (2024). Automatic student engagement measurement using machine learning techniques: A literature study of data and methods. *Multimedia Tools and Applications*, *83*(16), 49641–49672. https://doi.org/10.1007/s11042-023-17534-9

Muniasamy, A., & Alasiry, A. (2020). Deep learning: The impact on future eLearning. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(1), 188–199. https://doi.org/10.3991/ijet.v15i01.11435

Murshed, M., Dewan, M. A. A., Lin, F., & Wen, D. (2019). Engagement detection in e-learning environments using convolutional neural networks. In *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)* (pp. 80-86). IEEE. https://doi.org/10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00028

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, *31*, 527–541. https://doi.org/10.1016/j.chb.2013.05.024

Pathak, D., & Kashyap, R. (2023). Neural correlate-based e-learning validation and classification using convolutional and long short-term memory networks. *Traitement du Signal*, *40*(4), 1457–1467. https://doi.org/10.18280/ts.400414

Picard, R. W. (2010). Affective computing: From laughter to IEEE. *IEEE Transactions on Affective Computing*, *1*(1), 11–17. https://doi.org/10.1109/T-AFFC.2010.10

Salau, L., Hamada, M., Prasad, R., Hassan, M., Mahendran, A., & Watanobe, Y. (2022). State-of-the-art survey on deep learning-based recommender systems for e-learning. *Applied Sciences*, *12*(23), Article 11996. https://doi.org/10.3390/app122311996

Sebbaq, H., & El Faddouli, N.-e. (2022). Fine-tuned BERT model for large scale and cognitive classification of MOOCs. *International Review of Research in Open and Distributed Learning*, *23*(2), 170–190. https://doi.org/10.19173/irrodl.v23i2.6023

Selim, T., Elkabani, I., & Abdou, M. A. (2022). Students engagement level detection in online e-learning using hybrid EfficientNetB7 together with CN, LSTM, and Bi-LSTM. *IEEE Access*, *10*, 99573–99583. https://doi.org/10.1109/ACCESS.2022.3206779

Sweller, J. (2019). Cognitive load theory and educational design: Recent developments. *Educational Psychology Review*, *31*(2), 261–276. https://doi.org/10.1007/s10648-019-09465-5

Toti, D., Capuano, N., Campos, F., Dantas, M., Neves, F., & Caballé, S. (2021). Detection of student engagement in e-learning systems based on semantic analysis and machine learning. In *Advances on P2P, Parallel, Grid, Cloud and Internet Computing: Proceedings of the 15th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2020) 15* (pp. 211–223). Springer International Publishing. https://doi.org/10.1007/978-3-030-61105-7_21

Yin, L., Zhang, X., & Luo, Z. (2023). Affective computing in online learning: Trends, challenges, and future directions. *Computers in Human Behavior Reports*, *11*, Article 100297. https://doi.org/10.1016/j.chbr.2023.100297

Yu, J. H., & Chauhan, D. (2025). Trends in NLP for personalized learning: LDA and sentiment analysis insights. *Education and Information Technologies*, *30*(4), 4307–4348. https://doi.org/10.1007/s10639-024-12988-2

Zhang, Y., Wang, C., & Lee, H. (2023). Text sentiment analysis using BERT for student feedback interpretation in online learning platforms. *Education and Information Technologies*, *28*(5), 6327–6342. https://doi.org/10.1007/s10639-023-11478-z

Zimmerman, B. J. (2020). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *55*(3), 145–160. https://doi.org/10.1080/00461520.2020.1772025

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Microphones on Unmute: Perceived Online English-Speaking Anxiety of Non-Native EFL Educators

Zülal Ayar
İzmir Katip Çelebi University, Turkey

Abstract

While teachers worldwide rapidly switched to emergency remote teaching almost overnight owing to the unprecedented global pandemic, the rise of artificial intelligence (AI) has further transformed language education paradigms. Although previous research has explored foreign language teaching anxiety (FLTA), the self-perceived online L2 speaking anxiety of teachers remains underexplored. Accordingly, this study has been designed on a wide scale to address this lacuna by focusing on the perceptions of anxiety of 179 non-native EFL teachers at the Ministry of Education and instructors in higher education contexts. Moreover, it aimed to reveal its provoking reasons and finally the reported reflections of educators' apprehension on virtual classes. To that end, qualitative and quantitative data were gathered in a complementary fashion through semi-structured interviews and an online survey developed by the researcher. The study identified the lack of perceived competence, troubles with online technologies, and learners' English proficiency as factors leading EFL educators to experience online L2 speaking anxiety despite their self-confidence. Their reported reflections also disclosed that self-confidence without competence would be of almost no use in language teaching. Finally, some significant differences were detected between the participants' demographic variables and their online L2 speaking anxiety.

Keywords: distance education, FLTA, online speaking anxiety, teacher perception, remote teaching

Introduction

Teaching has long been considered a laborious and exacting profession, not only due to its pedagogical and technological requirements to target diverse learners in diverse educational settings, but also complicated by challenges such as school policy, workload, lack of technical support, or demands from other external factors (Schipor & Duca, 2021). With the advent of the COVID-19 pandemic, a massive unplanned transition from frontal instruction to online platforms occurred in the education system. Though this new modality of teaching was then considered a panacea and an interim solution to prevent the disruption of education, it had unprecedented effects on the system (Nazari et al., 2022). Teachers scrambled to reorient their face-to-face methods and techniques to distance or open and distributed learning (ODL) within days. Besides online delivery of education shelter-in-places, they confronted other difficulties, such as struggling to (a) minimise communication gaps via guided conversations and mediated dialogues (Moore, 1993), (b) promote students self-discovery of knowledge (Knowles, 1975), (c) build a community of practice (CoP) through peer interaction, and (d) facilitate networked learning on digital platforms (Lave & Wenger, 1991). Hence, the burgeoning literature on non-native language teachers has reported that a good number who could not manage to acclimatize to this abrupt migration experienced high anxiety (Akdeniz, 2022).

As reflections of non-native teacher anxiety are multifaceted and have had a different impact on distinct levels of education during and after the pandemic, various related studies have been implemented in language education. However, most of this research has been conducted either directly to language learners, given they are always at the centre of education, or in tandem with non-native teachers and students to see whether or how they trigger or prevent their anxieties. As observed by Francisco and Alieto (2022), further investigation concerning specific types of anxiety in the pandemic period and beyond into the era of expanding AI-based tools has shown that most research has continued to focus on learners' speaking anxiety (e.g., Chen, 2024; Faqihi, 2024; Wu et al., 2025), rather than on teachers' anxiety. However, as Horwitz (1996) noted, "language learning is never complete" (p. 365), and teachers themselves remain ongoing language learners. Therefore, it is essential to investigate the foreign language speaking anxiety (FLSA) or the anxiety experienced by non-native teachers when teaching speaking. Given this prevailing trend of studies in English language teaching (ELT) and the dearth of empirical research delving into perceptions of online L2 speaking anxiety among non-native English as a foreign language (EFL) teachers, the current study aimed to bridge this gap by exploring the underlying reasons for such anxiety among Turkish educators of English.

Conceptual Framework of the Study

Researchers in ELT have conducted a broad spectrum of studies on language anxiety (LA) for several decades. Scovel (1978) was the first scholar to use the term anxiety in a foreign language education. Thereafter, Horwitz et al. (1986) investigated the impact of anxiety on language learning and identified foreign language anxiety (FLA) with conceptual foundations. Since then, FLA has been mostly scrutinized based on three related anxieties: test anxiety (Young, 1991), fear of negative evaluation (Aida, 1994), and communication apprehension (Horwitz et al., 1986). After the 1980s and 1990s, considering its relation to different aspects of language skills, pioneering studies on FLA have investigated its potential effects on students' success (Woodrow, 2006). Yet, Horwitz (2001) suggested that teachers of a foreign or second language might also feel anxious due to problems they can neither foresee nor manage. Accordingly, their FLA or FLTA warrants thorough investigation.

To explore FLTA in-depth, several studies have analysed its relationship with various teacher-related variables. Similar to Dişli (2020), Eren (2020) and Aydın and Uştuk (2020) noted that EFL teachers at the onset of their professional careers were more anxious. Aslrasouli and Vahid (2014) also found a negative correlation between teachers' age and experience and their anxiety levels, although the correlation was weak. Nevertheless, they emphasized the high anxiety levels observed among both novice and seasoned teachers. Yet, in contrast to Eren (2020), they did not discover any differences between gender and teaching anxiety among these Iranian participants. Like Dişli (2020), Machida (2016) also reported no significant correlation between gender and teachers' sense of anxiety. However, he did not find a significant association between overall English teaching experience and anxiety, which differed from the Iranian context. Instead, he emphasized that factors such as teaching experience (specifically at the elementary school level), teachers' proficiency levels, and in-service professional training course participation had a notable impact on teacher anxiety. Moreover, considering educators' last completed degrees for possible contributions to overcoming anxiety, Eren (2020) revealed no significant difference among teachers with a Master's degree (MA), or Doctor of Philosophy (PhD) and Bachelor's (BA) degree with regard to their anxiety. Similarly, Akdeniz (2022) did not identify significant relationships between teacher anxiety and variables such as educational background, gender, overall FLTA, or length of teaching experience. Still, fear of negative evaluation was related to school context; teachers at middle and high schools reported less fear and apprehension than did instructors in tertiary-level education.

A growing body of research on FLTA has focused on identifying both the factors that provoke anxiety and the ways in which this anxiety is reflected in classroom practices and dynamics. Aslrasouli and Vahid (2014) associated the reflections of FLTA with (a) language knowledge and proficiency, (b) available facilities, (c) the system of employment, and most importantly, (d) interpersonal relationships. Dişli (2020) detailed that (a) making mistakes, (b) giving lessons to learners at a specific language level, (c) teaching a particular skill, (d) learner attitudes, (e) referring to L1, and (f) misuse of the technology were listed as the key sources of FLTA. İpek (2016) also reported the first three aforementioned factors as the reasons for FLTA in the Turkish EFL context. Likewise, Sammephet and Wanphet (2013) highlighted the importance of (a) learner attitudes and low teacher proficiency in English, (b) teachers' hesitations in decision-making during courses, (c) time management, and (d) teaching contexts as other causes of anxiety. Furthermore, Aydın and Uştuk (2020) highlighted EFL teachers' anxiety about making mistakes due to their self-perceptions of English proficiency. Additionally, pressure and confusion caused by learners' unexpected questions, as well as the fear of their negative evaluation, can all trigger FLTA. Similarly, Akdeniz (2022) found a strong relationship between teachers' concern about being misjudged and their anxiety. Other factors correlated with FLTA have included (a) a lack of interest among learners, (b) teaching students at a particular proficiency level, (c) giving lessons focused on specific language skills, (d) being unprepared for the course, (e) encountering new situations during class, and (f) experiencing problems with content knowledge. Finally, other studies with a similar scope have linked possible causes of FLTA to (a) student-teacher relationships, (b) teachers' self-esteem, (c) personal efficacy, (d) self-confidence, and (e) perceived competence in classroom performance, as well as aspects of the teaching setting itself (Eren, 2020; Merç, 2015; Young, 1991).

Horwitz (1996, 2001) affirmed that FLTA was associated with verbal language use. By the same token, much research in language education concentrating on FLTA has confirmed that speaking skill, above all others, seems to be the most anxiety-provoking, with foreign language teachers attributing this to their perceived incompetence in communication (Liu, 2009; Takahashi, 2009). Teachers are called on to perform a range of complex functional and cognitive acts while they manage the classroom, arrange courses enriched with conversational tasks based on constructive communications to lower transactional distance (Moore, 1993), and enhance learner autonomy and self-directed learning (Knowles, 1975). All this addresses concerns about their own pronunciation, accent, and fluency (Alqahtani, 2019; Aydın & Uştuk, 2020). Horwitz et al. (1986) warned language teachers to consider learner profiles, culture, and the environment, given the potential of these to push them to speaking anxiety or interrupt communication. Clearly, a comprehensive investigation is needed to understand FLSA or English speaking anxiety (ESA) in this context (Kralova & Tirpakova, 2019).

Non-Native Teacher (Online) FLSA in EFL Lessons

The previous literature drawing attention to the FLSA of teachers in non-native EFL contexts has mostly focused on pre-service education, aiming to measure their ESA, reported correlations or effects on their practicum, and identify its sources. To illustrate, Daud et al. (2019) invited non-native English language teacher candidates to participate in a mixed method study design to determine their ESA and its leading factors. Although the rate of ESA was at a moderate level, it was attributed to linguistic factors (e.g., inadequacy in grammar, vocabulary, or pronunciation), social factors (e.g., feeling under pressure during communication, fear of making mistakes, being negatively judged, or losing face), and personal issues (e.g., motivation). Gürsoy and Korkmaz (2018) examined freshmen and senior pre-service EFL teachers' FLSA through sequential mixed research. They found that these teachers experienced ESA at a mild level, with significant differences observed between male and female student teachers. In addition, the scholars found a negative correlation between participants' ESA and English proficiency, especially regarding errors in grammar, vocabulary, syntax, and pronunciation. The causes of FLSA were listed as fear of interaction, apprehension about the audience and their negative appraisal, and lack of confidence. Finally, Yaşar and Atay (2021) focused on online discussions and their possible impact on non-native EFL student teachers to investigate their ESA in ODL context. Quantitative reviews and open-response questionnaires were employed to shed light on participants' perceptions. Yaşar and Atay concluded that online debating lowered prospective teachers' ESA with a statistically significant effect. Moreover, as emphasized by Lave and Wenger (1991), online debates encouraged participants to develop both their social skills by being involved in an interactive environment and their English-speaking skills by engaging in active online discussions.

As for in-service language teachers' ESA, only a few extensive studies exist in the literature. For example, Kralova and Tirpakova (2019) investigated the ESA of 175 Slovak EFL teachers in relation to their demographics, such as (a) time spent in an English-speaking country, (b) length of teaching English and language study, (c) age, and (d) intensity of interactions with native speakers. They found a positive relationship between ESA and age, whereas negative correlations were observed between ESA and most of the other variables. Similarly, Karakaya (2011) examined the ESA, foreign language listening anxiety, and perceived competence in teaching skills of 150 non-native EFL instructors, revealing significant correlations with their L2 learning contexts and seniority. Lastly, she reported that their educational background and participation in teacher training programs contributed to perceived competence in teaching these skills.

As has been seen in the foregoing studies, the sources of FLSA of EFL (student) educators, their anxiety levels, and its effects on their language performance have been examined. Yet, to the best of our knowledge, no study has explored the perceptions of online ESA of both in-service English teachers working at the Ministry of Education (MoE) and English instructors affiliated with higher education (HE) in non-native country contexts using a mixed-method research design. However, the so-called new normal introduced by COVID-19, the stress and confusion teachers experienced due to the shift to distance and ODL, as well as other potential unexpected and uncontrollable factors that AI might bring can provoke FLSA more than ever (Yaşar & Atay, 2021). The current literature lacks such a study, the findings of which will provide information on EFL educators' self-perceived online ESA by examining relationships across a wide spectrum of variables (i.e., gender, age, last completed degree, English teaching experience, and school contexts), while also unveiling the reflections of their sense of FLSA in virtual class contexts. This study was guided by the following research questions.

1. What are the self-perceived reasons for online ESA of EFL teachers and instructors during online instruction?
2. Do these reasons show any statistically significant difference in terms of demographic variables?
3. What are the reported reflections of educators' online ESA perceptions on digital EFL classes?

Methodology

Participants and Setting

This research was designed to be administered to both in-service English teachers at the MoE and EFL instructors in HE in Turkey. An online survey was prepared to be delivered to different schools (primary, middle, or high) and universities' foreign language units after obtaining ethical approval from the ethics board at the researcher's university. Accordingly, 46.9% of the volunteer attendees were drawn from the in-school contexts listed above, while 53.1% were incorporated from the latter HE context through simple random sampling. The schools where these language practitioners were based were in distinct parts of the country.

As for the participants' gender distribution, 78.2% were female and 21.8% were male. Regarding age distribution, 11.2% of these educators were aged 30 years or less, 56.4% were between 31 to 40 years of age, 21.2% were 41 to 50, and 11.2% were 51 years of age or older. According to their last completed degrees, 55.9% of the participants had an undergraduate degree, 33.5% had a graduate degree, and 10.6% held a PhD. Regarding their English teaching experience, most of the participants at the MoE had no teaching experience in kindergarten, and more than half did not teach in primary schools; however, 48 teachers gave lessons in middle schools, and nearly half of this sample taught learners in high schools. Of those teaching in the HE context, only one-third of those language practitioners had experience with two-year degree programs, whereas nearly half of them delivered lectures to undergraduate students.

Data Collection

The online survey covered demographic information, followed by a section on online ESA of EFL educators developed by the researcher after searching the literature for well-developed scales on teachers' second language speaking anxiety, FLSA, or ESA (e.g., Aydın & Uştuk, 2020; Capel, 1997; Ferguson et al., 2012; Horwitz et al., 1986; Kralova & Tirpakova, 2019; Woodrow, 2006). These scales were principally created to measure the speaking anxiety levels of learners, teachers, or pre-service teachers in face-to-face education, different from the intention of this research in many respects. This newly designed 16-item scale was in English and created specifically for the online environment to determine the online ESA of EFL in-service educators. It used a five-point Likert scale with answers from *strongly disagree* (SD) to *strongly agree* (SA). The researcher consulted a professor from linguistics and an assistant professor from foreign language education to edit the language in the survey. Thereafter, three experts in the field of ELT and one expert from the assessment and evaluation department evaluated the final version of the scale for content validity. It was then piloted with 12 in-service English teachers and 11 EFL instructors who were not participants in the main research. Accordingly, the survey was well-understood by the educators and, therefore, was shared through a link without any refinements. The online survey was administered to the target group via Google Forms, and responses were collected anonymously. The reliability of the scale was confirmed with a 0.764 Cronbach alpha (α) coefficient value (Cortina, 1993).

As to the triangulation of the data, one-to-one semi-structured online interviews were held in the Turkish language (L1) to help attendees feel unconstrained. Five percent of participants volunteered to take part in these dialogues by providing their e-mail addresses at the end of the survey. For this process, the researcher developed the interview questions after thoroughly reviewing inquiries in the field and then sent them to the two aforementioned experts in ELT for their valuable feedback. The participant educators were asked questions regarding their online ESA, its potential causes, and reflections on digital courses. For example:

- What are the advantages and disadvantageous of online courses in terms of English-speaking performance in class when compared to traditional face-to-face lessons?
- What are the reasons for your ESA in online lessons, if any?
- Are these reasons the same as your face-to-face experiences? If not, what are the differences?
- What helps you to feel less anxious when speaking English in online courses?
- What are the reflections of your anxiety on online lessons?

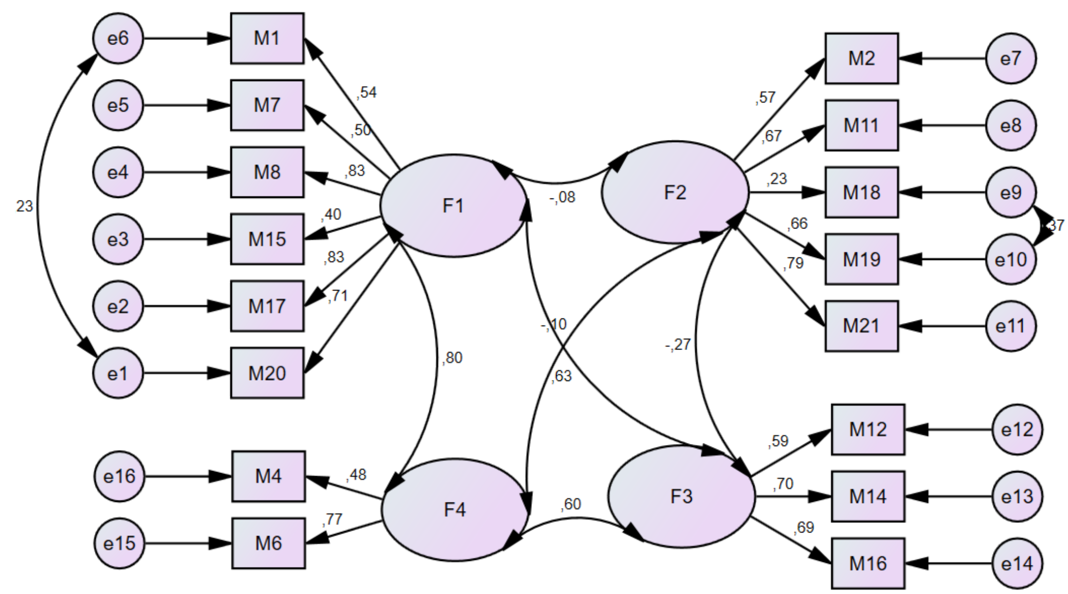
Each interview lasted 15 to 20 minutes and was video recorded on the Zoom platform. Overall, the data collection process took almost two months due to the participants' tight schedules and other time constraints.

Data Analysis

The data were analysed using Statistical Package for Social Sciences for Windows 25.0 and Analysis of a Moment Structures. In pursuit of the reliability analysis, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was calculated and found to be 0.812, indicating that the sample was adequate for factor analysis. In addition, results of the chi-square value were recorded to be significant χ2 (120) = 1034.443; p < 0.01) according to the Bartlett Sphericity test, confirming that the correlations between items were sufficiently large for factor analysis. An explanatory factor analysis was conducted using principal components analysis and varimax rotation methods (Browne & Cudeck, 1993). Survey items 3, 5, 9, and 13 were excluded due to overlapping loadings, and item 10 was removed because its total item correlation was below 0.30. The final four-factor solution accounted for 61.8% of the total variance. As for the confirmatory factor analysis, the structural equation modelling results were significant at the p = 0.000 level, with 16 items loading onto four factors. Model fit was further improved by allowing covariances between error terms with high modification indices. The goodness of fit index (GFI) was 0.852, indicating an acceptable model fit. In conclusion, the scale was confirmed to provide construct validity (X^2/SD = 2.723; RMSEA = 0.098; CFI = 0.826; AGFI = 0.791; NFI = 0.756; IFI = 0.830; TLI = 0.783).

Figure 1

Multifactor Model



Note. M, F, and e represent the item, factor, and error term, respectively.

Figure 1 shows that while F1 had the highest explanatory value, F4 had the lowest, which explains why this factor had more connections with the others. In contrast to F4, the lack of strong correlations among other factors (F1–F2 = -0.8; F2–F3 = -27; F1–F3 = -10) shows that the distribution of the items was appropriate. The number of covariances was also below five, and the fact that they were among internal factors further increased the reliability of the model. Moreover, the t-value was significant and the factor loading values of all items were between 0.344 and 0.823. Thus, the factors were appropriate for the structure, and the structure itself was statistically confirmed (Browne & Cudeck, 1993). Finally, the independent *t*-test, a one-way analysis of variance, and the Bonferroni test were applied to find significant differences between ESA and demographic variables of the participants (see Appendix).

As to the qualitative data, participants were each assigned a number to maintain their anonymity and these identifiers were included with their quotes (i.e., E1, E2, and so on). Thematic analysis of the interviews was done by the researcher and one of her colleagues who was a PhD candidate in ELT. After transcribing and translating into L2, the data was coded in two rounds. First, the codes for the first three questions were created separately, and the raters came together to compare each code and reshape them after they coalesced. Next, categories were identified and the emerged themes listed beneath them. Having analysed the verbatim translations of all interview questions, investigation of the last two questions was also completed by these two raters adhering to the same process of thematic analysis.

Table 1

Provoking Reasons for Online ESA

Category	Theme	Descriptive comments
Student-related items (n = 35 references)	1. Lack of motivation	unwillingness in digital classes and not fulfilling the requirements of the course due to lack of purpose
	2. Lack of interactivity	lack of visual communication and facial expressions
	3. Lack of feedback	learners do not provide (immediate) feedback or pass any remarks
	4. Lack of interest	not directing attention to the course or not desiring to carry out tasks
	5. Student differences	overall differences among the target audience
Teacher-related items (n = 62)	6. Being video recorded	performance being judged by others
	7. Lack of emotional bond with learners	lack of common background with some learners
	8. Mispronunciations	fear of making pronunciation mistakes
	9. Inability to express oneself	feeling inadequate to pass along the gist of the course to learners (depending on their levels)
	10. Inability to understand learners' intake	problem of detecting learners' comprehension
Other items (n =5)	11. Failure to make students open cameras	legal dimension of online courses
	12. Technical issues	web-based problems

After gathering all categories and themes and combining them, three categories and 12 distinct themes were identified (see Table 1). As the raters reached .81 inter-rater reliability, their almost perfect agreement was statistically confirmed (Landis & Koch, 1977).

Findings and Discussion

As Figure 1 highlights, based on the six items (M1, M7, M8, M15, M17, and M20) within the first factor, the primary reason for online ESA was reported to be related to the lack of perceived competence in online English-speaking similar to Alqahtani (2019), Aslrasouli and Vahid (2014), Aydın and Uştuk (2020), Daud et al. (2019), Gürsoy and Korkmaz (2018), Karakaya (2011), Liu (2009), Kralova and Tirpakova (2019), Sammephet and Wanphet (2013), and Takahashi (2009). As M8 (I feel that I need to improve my English speaking after the courses), M17 (I am concerned about my English deficiencies in speaking), and M20 (It makes me upset that I cannot control grammar when I speak English) had the highest values, this was directly correlated with the deficiencies in general language use. Moreover, M1 (I feel more anxious when I have to speak English than in the real classroom), M15 (I would feel less nervous about my course if I did not have to cover so much material in such a short time), and M7 (I am worried that students will not understand my English speech) corroborated this notion despite their lower ratio in the model. This was also evident in the comments of the interviewees in Table 1 where teacher-related items (n = 62) were listed as the category involving the most prominent reasons for online ESA. This was particularly due to (a) failing to observe learners' knowledge construction (see theme 10 with 23 references); (b) following structured open learning; and (c) feeling transactional distance (Moore, 1993; see themes 7 and 9 with 20 and 17 references, respectively). Participants did not feel they adequately understood the extent to which their students grasped the core of their speech or were able to get their message across to the target group (Eren, 2020; Merç, 2015).

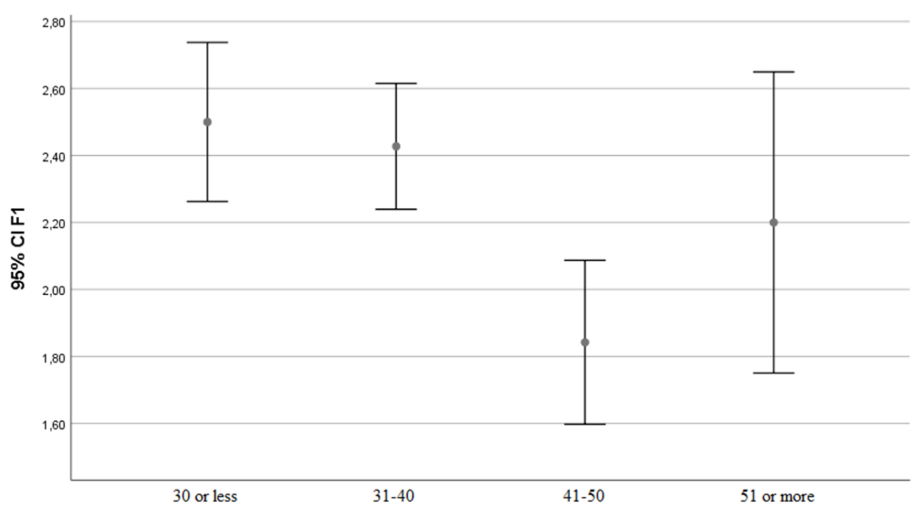
As to F3, the use of online technologies was labelled as another cause of anxiety, as exemplified by M12 (I feel overwhelmed by the complexity of the platform), M14 (I feel more confident about speaking English when I do not have to turn on my camera), and M16 (I would feel better about speaking English if the platform allowed fewer students to attend lessons). Some participants attributed their ESA to virtual settings and the use of digital tools, already identified as a theme (entitled other items) in Table 1, and similarly explored by Akdeniz (2022). Furthermore, with its two items, namely M4 (I feel less nervous while giving instructions in English to students at low proficiency levels), and M6 (I do not feel comfortable when my students speak better English than I do), F4 described learner proficiency as the other provoking reason for ESA. This was also substantiated by the categories of student-related items (n = 35), and the second, third, and fifth themes, which supported the importance of rapport and interaction among students and educators (Moore, 1993), as well as creating a CoP (Lave and Wenger, 1991) in distance education or ODL. Hence, this result was in tune with the work by Dişli (2020) and İpek (2016).

Another factor supporting F4's result was F2, since it attributed the potential reason for ESA to a lack of self-confidence in online speaking performance. Yet, as this factor covered positive attitudes towards online language teaching experiences different from the others, it displayed the self-assurance of these educators as evidenced by high scores on M19 (I feel comfortable when I come to the course prepared), M21 (I like spontaneous English talks with my students similar to the ones in the real classroom), M18 (Though I initially feel worried about how to conduct lessons, I can complete them successfully), M2 (I am satisfied with my credentials to speak English during the course), and M11 (I feel confident in using all the required supplementary materials in my course). Concerns by E8 illustrate the gap and disagreement between F1 and F2: “Whenever I come to class prepared, I get the feeling of standing up to all kinds of difficulties while speaking, but unfortunately, I experienced that I could not overcome the questions posed instantly”. This suggests that despite confidence in their English-speaking performance in digital classes, the lack of perceived competence interfered with their L2 speaking skills (see Karakaya, 2011 for further discussion).

As for the second research question (see Appendix), a statistical difference was not detected among any of the factor values and participants’ gender, similar to Akdeniz (2022), Dişli (2020), and Machida (2016), but different from Eren (2020), and Gürsoy and Korkmaz (2018). On the other hand, the Appendix displays a significant difference between F1 scale rates according to participants’ ages, similar to Kralova and Tirpakova (2019). In addition to the chart in Figure 2, which demonstrates F1 mean rates and range according to age groups, the Bonferroni test also revealed that the rates of participants aged 30 or below and 31 to 40 were higher than those aged 41 to 50. This finding implied the significance of seniority, indicating that the younger the educators were, the less their perceived competence (Aydın & Uştuk, 2020; Dişli, 2020; Eren, 2020).

Figure 2

F1 Mean Rates According to Participants’ Ages



Another significant difference was recorded between each of F1, F2, F3, F4, and the total rates of the participants according to the institutions where they worked. The rates for these factors and MoE teachers were higher than those in the HE context. Moreover, the F2 rates of instructors in HE were higher than those of teachers in the MoE. That is, teachers in MoE were reported to feel more anxious since their self-confidence was less than in the HE context, in contrast with the result by Akdeniz (2022). This can also be correlated with their last completed degrees, which the researcher also detected as a statistically significant difference in F3 rates. The Bonferroni test uncovered that the F3 scale rates of participants who had completed only an undergraduate level were higher than those with an MA, dissimilar to the findings reported by Eren (2020). Given potentially limited professional development opportunities (Karakaya, 2011; Machida, 2016), inadequate training for the use of technology in ODL (Dişli, 2020), the permissive academic atmosphere in school contexts (Akdeniz, 2022), and fewer teachers with an MA degree in MoE, this finding complemented the foregoing conclusion.

Having scrutinized the statistical relationship of factors with five independent variables, the researcher noted a significant difference between F3 and total rates for the primary school teachers. Unlike Machida's (2016) findings, their ESA scale rates were higher than teachers without this experience. Another statistically significant difference was found for F3 rates according to middle school experience, whose F3 was higher than for the educators without this experience. Similarly, F3 rates were significantly higher for those with high school teaching experience compared to those who had not taught at that level. As the only participant with a total of 26 years of experience across different education levels, this can be best explained by E7. "I have gained the awareness that my technological knowledge in language teaching is insufficient, especially after experiencing the challenge of referring to online resources in digital language courses; I felt anxious about this situation at times."

Distinct from MoE, another significant difference was observed in F2, F3, F4, and the total anxiety scores, based on instructors' experience in associate degree programs. Additionally, their F2 was higher than participants without any experience in a two-year degree. This was totally in parallel with F3, F4, and total rates of participants who did not have experience in this degree; their rates were higher than those for the inexperienced participants. Finally, similar to the two-year degrees, the researcher reported a significant difference for the rates of F1, F2, F3, and F4 according to the participants' experience in BA programs; their F2 rates were also higher than others without this experience. It was also apparent that the F1, F3, and F4 rates of participants without undergraduate experience were higher than practitioners who had gained experience there. Taken together, with regard to both associate and BA degrees, instructors in the HE context were self-confident in improvising speech and managing the class, yet they were concerned about efficient language use, technical issues, and learners' English language level in online lessons (Karakaya, 2011). In that respect, E3 also declared that "as learner proficiency is above the average and the course content is highly loaded, it is necessary to be assertive about technological pedagogical content knowledge in the online platform, thus making any mistakes while speaking worries me."

The reported reflections of online ESA experienced by EFL educators in virtual language classes is worth discussing, especially in light of the analysis of the last interview questions. The first category (observable reactions in speech) ranked with 25 references in total given the rates of L1 use and teacher-centred class, whereas the category entitled evasive reactions with two themes (i.e., changing topic, having breaks) reached eight references. The category least referred to, namely personal reactions, received only four references relating to the themes of impatience and threats. These likely stemmed from participants' troubles with English-speaking competence (see F1 in the multifactor model) and faltering due to the lack of learner feedback, interest, and interactivity (Aslrasouli & Vahid, 2014; Sammephet & Wanphet, 2013), which are regarded as pivotal items in ODL and distance education by Lave and Wenger (1991). In this respect, E5 shared that "sometimes I feel the need to hear their voices to such an extent that I feel impatience and a sense of my anxiety intensifying, then I directly refer to L1 and pose some questions to receive feedback from them." This echoed aspects of poor teacher-student interactions (Eren, 2020; Merç, 2015; Moore, 1993). However, as was already presented in Table 1 (see theme 6), being one of the two teachers who had not experienced online ESA, E1 suggested that "anxious teachers probably have a fear of being judged. If I experience ESA during my e-lessons, I do not attempt to camouflage but share it with learners or confess that I made a mistake." E1 attributed teacher' online ESA to their fear of negative evaluation (Aida, 1994; Aydın & Uştuk, 2020; Gürsoy & Korkmaz, 2018; Daud et al., 2019). This finding suggested that reflections of online ESA may be situation-specific since some participant educators seemed to think and feel about these experiences differently (Akdeniz, 2022; İpek, 2016).

Conclusion and Implications

This study revealed that, despite educators' self-confidence, some notable factors (i.e., the lack of perceived competence, troubles with the use of online technologies, and learner proficiency) prompted language practitioners to experience communication apprehension in online settings. This might indicate that self-confidence alone is insufficient to mitigate anxiety if it is not supported by actual competence, particularly in the context of online language instruction. Furthermore, the findings stressed that teachers' perception of their own competence was directly proportional to their age; younger educators tended to feel less competent and, consequently, experienced higher levels of online ESA compared to their older counterparts. Additionally, English teachers working in the MoE reported higher levels of online ESA than those teaching in other contexts, such as higher education. This difference may be attributed to variations in institutional support, resources, and expectations.

Educators with MA degrees did not feel the challenge of either using technology or meeting the overall requirements of online teaching compared to teachers with only undergraduate degrees, which might suggest that higher educational attainment is associated with greater adaptability and confidence in online environments. A striking fact concerning the school context was that teachers from MoE were more likely to experience online ESA due to issues related to perceived competence and technology use, while instructors without experience in two-year and BA degrees reported apprehensions about perceived competence, technology use, and learner proficiency. Finally, the researcher noted that the reflections of educators' ESA on EFL courses might vary depending on the specific context of ODL, indicating that anxiety is not uniform but rather situation specific.

As an implication of this study, the online ESA of non-native in-service teachers seems to be virgin territory, as evidenced by the hardship the researcher experienced finding a research design that matched the scope of this study. Thus, further similar studies could delve deeper into the underlying causes of online (perceived) ESA, identify its levels, and assess the quality of teachers' speech, potentially using AI tools or observations of their digital field experiences. As ODL intends to be accessible and inclusive, and ESA appears as a barrier here, future studies could be administered to both teachers and learners to check the impact of teachers' ESA on learner achievements with pre- and post-tests. Investigators could also analyse students' perceptions of educators' oral involvement and the ways in which they distinguish more or less proficient online English-speaking teachers. Finally, concerning the limitation of this research, as the gathered data only concentrated on Turkish EFL educators, scholars could also incorporate teachers with diverse cultural and linguistic backgrounds from foreign/second language contexts into similar research designs. This would help transfer our findings to wider ODL contexts. Increasingly, global ODL settings consider psychologically safe learning environments, and support teacher training and professional development through alternative interaction modes and built-in support systems. Thus, researchers could also examine how educators' ESA mediates affective experiences to potentially improve their self-confidence, self-esteem, perceived competence, and personal efficacy. Integrating anxiety-reducing strategies into platforms can scaffold inclusivity in global ODL environments (Nazari et al., 2022; Schipor & Duca, 2021; Zarei et al., 2024).

Acknowledgements

I want to express my gratitude to the anonymous participants who made this study possible by sharing their candid thoughts and to the IRRODL editors, especially Agnieszka (Aga) Palalas, Robyn Unrau, and Janice Thomas, who made considerable refinements to the study.

Declarations

- Conflict of Interest: The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.
- Funding statement: The author did not receive any funding.
- Ethical approval and informed consent statements: The author confirms that ethical approval has been obtained from the board.
- Data availability statement: The data that support the findings of this study are openly available in the manuscript.

References

Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal, 78*(2), 155–168. http://doi.org/10.1111/j.1540-4781.1994.tb02026.x

Akdeniz, G. Y. (2022). *A mixed-methods study on foreign language teaching anxiety among Turkish EFL teachers* [Master's thesis, Anadolu University]. https://www.proquest.com/openview/eea095295aa3eb7930e4921174b3940f/1?pq-origsite=gscholar&cbl=2026366&diss=y

Alqahtani, S. M. (2019). The impact of language anxiety on EFL learners' proficiency: Case study of University of Jeddah. *Adalya, 8*(12), 844–864. https://www.researchgate.net/publication/339800167

Aslrasouli, M., &Vahid, M. (2014). An investigation of teaching anxiety among novice and experienced Iranian EFL teachers across gender. *Procedia-Social and Behavioral Sciences*, *98*, 304–313. https://doi.org/10.1016/j.sbspro.2014.03.421

Aydın, S., & Uştuk, Ö. (2020). A descriptive study on foreign language teaching anxiety. *Online Journal of Education and Teaching, 7*(3), 860–876. https://iojet.org/index.php/IOJET/article/view/846

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage. https://doi.org/10.1177/0049124192021002O

Capel, S. A. (1997). Changes in students' anxieties and concern after their first and second teaching practices. *Educational Research*, *39*(2), 211–228. https://doi.org/10.1080/0013188970390208

Chen, Y.-C. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Computer Assisted Language Learning*, *37*(4), 789–813. https://doi.org/10.1080/09588221.2022.2055083

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Daud, A., Ras, F., Novitri, X., & Audia, C. P. (2019). Factors contributing to speaking anxiety: A case study of pre-service English teachers. *Journal of Educational Sciences, 3*(3), 412–422. https://doi.org/10.31258/jes.3.3.p.412-422

Dişli, A. (2020). *A mixed approach towards EFL teachers teaching anxiety* [Master's thesis, Çağ University]. https://acikbilim.yok.gov.tr/handle/20.500.12812/71336

Eren, G. (2020). *Foreign language teaching anxiety and self-efficacy perceptions of native and non-native EFL instructors at tertiary level institutions* [Master's thesis, Bilkent University].

https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=1gNB-ZW1XFx2Yi11S74g-A&no=XJN-Ow2LYY1Y9mLHOWQ4mA

Faqihi, M. A. (2024). Triggers for foreign language speaking anxiety: Perceptions of Saudi EFL college students. *English Language Teaching, 17*(2), 1–17. https://doi.org/10.5539/elt.v17n2p1

Ferguson, K., Frost, L., & Hall, D. (2012). Predicting teacher anxiety, depression, and job satisfaction. *Journal of Teaching and Learning, 8*(1). https://doi.org/10.22329/jtl.v8i1.2896

Francisco, W. S., & Alieto, E. O. (2022). Language learning in modular set-up: Challenges and coping mechanisms of secondary students. *Basic and Applied Education Research Journal, 3*(1), 41–55. http://doi.org/10.11594/baerj.03.01.04

Gürsoy, E., & Korkmaz, H. (2018). Speaking anxiety of freshmen and senior prospective ELT teachers. *Eğitimde Kuram ve Uygulama, 14*(1), 48–60. http://doi.org/10.17244/eku.346886

Horwitz, E. K. (1996). Even teachers get the blues: Recognizing and alleviating non-native teachers' feelings of foreign language anxiety. *Foreign Language Annals, 29*(3), 365–372. https://doi.org/10.1111/j.1944-9720.1996.tb01248.x

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21*, 112–127. http://doi.org/10.1017/S0267190501000071

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Journal, 70*(2), 125–132. https://doi.org/10.2307/327317

İpek, H. (2016). A qualitative study on foreign language teaching anxiety. *Journal of Qualitative Research in Education*, *4*(3)*,* 92–105. https://doi.org/10.14689/issn.2148-2624.1.4c3s5m

Karakaya, D. (2011). *Non-native EFL teachers' foreign language listening and speaking anxiety and their perceived competencies in teaching these skills* [Master's thesis, Middle East Technical University]. https://open.metu.edu.tr/handle/11511/21040

Knowles, M. (1975). *Self-directed learning: A guide for learners and teachers*. Association Press. https://doi.org/10.1177/105960117700200220

Kralova, Z., & Tirpakova, A. (2019). Non-native EFL teachers' speaking anxiety: Post-communist country context. *SAGE Open*, *9*(2). https://doi.org/10.1177/2158244019846698

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33*(2), 363–374. https://doi.org/10.2307/2529786

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815355

Liu, M. (2009). *Reticence and anxiety in oral English lessons*. Peter Lang. https://www.peterlang.com/document/1105859

Machida, T. (2016). Japanese elementary school teachers and English language anxiety. *TESOL Journal, 7*(1), 40–66. https://doi.org/10.1002/tesj.189

Merç, A. (2015). Foreign language teaching anxiety and self-efficacy beliefs of Turkish pre-service EFL teachers. *The International Journal of Research in Teacher Education, 6*(3), 40–58. https://ijrte.inased.org/makale/156

Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22-38). Routledge. https://www.scirp.org/reference/referencespapers?referenceid=2868375

Nazari, F., Ghanizadeh, A., & Mirzaee, S. (2022). EFL teachers' coping strategies amidst the COVID-19 virtual education and their association with work engagement and teacher apprehension. *Educational Research for Policy and Practice, 22*(1), 1–22. https://doi.org/10.1007/s10671-022-09317-0

Sammephet, B., & Wanphet P. (2013). Pre-service teachers' anxiety and anxiety management during the first encounter with students in EFL classroom. *Journal of Education and Practice, 4*(2), 78–87. https://doi.org/10.13140/RG.2.1.5016.4966

Schipor, M.-D., & Duca, D.-S. (2021). On-site vs. online teaching: Challenges and coping strategies. *Revista Romaneasca pentru Educatie Multidimensionala*, *13*, 123–136. https://doi.org/10.18662/rrem/13.1Sup1/388

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning, 28*(1), 129–142. https://doi.org/10.1111/j.1467-1770.1978.tb00309.x

Takahashi, A. (2009). Self-perception of English ability: Is it related to proficiency and/or class performance? *Niigata Studies in Foreign Languages and Cultures*, *14*, 39–48. https://www.semanticscholar.org/paper/Self-perception-of-English-Ability%3A-Is-it-related-Takahashi-%E9%AB%98%E6%A9%8B/98ed59be9cf961fa46ca4846b2569044e2f1f396

Woodrow, L. (2006). Anxiety and speaking English as a second language. *Regional Language Centre Journal*, 37(3), 308–328. https://doi.org/10.1177/0033688206071315

Wu, T. T., Hapsari, I. P., & Huang, Y. M. (2025). Effects of incorporating AI chatbots into think-pair-share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*, 1–39. https://doi.org/10.1080/09588221.2025.2478271

Yaşar, M. Ö., & Atay, D. (2021). Emergency remote teaching and the effects of online debate on L2 speaking anxiety: Pre-service English language teachers' perspectives. *English as a Foreign Language International Journal, 2*(6), 80–104. http://doi.org/10.56498/423262022

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal, 75*(4), 426–439. https://doi.org/10.2307/329492

Zarei, M. A., Alibakhshi, G., & Nezakatgoo, B. (2024). Strategies employed by EFL teachers to cope with language learners' classroom anxiety. *TESOL Journal, 15*(4), e843. https://doi.org/10.1002/tesj.843

Appendix

Differences Between ESA Factors and Demographic Variables

Variables		Statistical test	F1	F2	F3	F4	Total
			$\bar{X}$±SS	$\bar{X}$±SS	$\bar{X}$±SS	$\bar{X}$±SS	$\bar{X}$±SS
Gender	Female		2.33±0.89	3.87±0.84	1.95±0.84	2.16±1.07	2.58±0.57
	Male		2.13±0.93	3.87±0.89	1.91±1.00	2.08±0.94	2.5±0.66
		t-test	1.239	-0.040	0.307	0.464	0.778
		p	0.217	0.969	0.760	0.643	0.437
Age	30 or less		2.5±0.51	3.91±0.67	1.82±0.75	2.3±0.85	2.63±0.39
	31 to 40		2.43±0.95	3.81±0.88	1.92±0.83	2.13±1.1.0	2.57±0.61
	41 to 50		1.84±0.74	3.99±0.89	1.91±0.91	2.12±1.00	2.47±0.6
	51 or more		2.2±0.96	3.9±0.78	2.25±1.13	2.1±1.01	2.61±0.62
		F-test	4.619	0.462	0.993	0.172	0.489
		p	0.004*	0.709	0.397	0.915	0.690
		Bonferroni	1.2>3	-	-	-	-
Institution	MoE		2.51±0.96	3.6±0.96	2.25±0.92	2.33±1.05	2.67±0.63
	HE		2.09±0.79	4.11±0.65	1.67±0.74	1.98±1.00	2.46±0.53
		t-test	3.245	-4.122	4.572	2.235	2.371
		p	0.001*	0.000*	0.000*	0.027*	0.019*
Last completed degree	BA		2.36±0.84	3.75±0.95	2.06±0.87	2.25±1.02	2.6±0.61
	MA		2.16±0.98	4.02±0.69	1.68±0.66	1.94±1.02	2.45±0.5
	PhD		2.31±0.95	4.01±0.6	2.16±1.28	2.26±1.15	2.68±0.68
		F-test	0.974	2.346	4.243	1.758	1.739
		p	0.380	0.099	0.016*	0.175	0.179
		Bonferroni	-	-	1>2	-	-
Teaching experience							
Kindergarten	NoE		2.22±0.85	3.88±0.91	1.89±0.8	2.09±1.04	2.52±0.56
	E		2.57±1.08	3.79±0.5	2.21±1.14	2.39±0.99	2.74±0.67
		t-test	-1.689	0.772	-1.519	-1.483	-1.921
		p	0.099	0.442	0.137	0.140	0.056
Primary school	NoE		2.22±0.86	3.9±0.86	1.79±0.83	2.03±1.06	2.48±0.6
	E		2.38±0.95	3.82±0.84	2.16±0.9	2.3±0.99	2.67±0.56
		t-test	-1.190	0.611	-2.915	-1.705	-2.066
		p	0.235	0.542	0.004*	0.090	0.040*
Middle school	NoE		2.24±0.83	3.96±0.66	1.75±0.79	2.02±1.04	2.49±0.55
	E		2.32±0.95	3.81±0.95	2.07±0.91	2.23±1.03	2.61±0.61
		t-test	-0.587	1.289	-2.420	-1.302	-1.259
		p	0.558	0.199	0.017*	0.195	0.210
High school	NoE		2.24±0.85	3.93±0.72	1.73±0.7	2.00±1.03	2.48±0.51
	E		2.33±0.95	3.8±0.96	2.16±0.98	2.29±1.02	2.64±0.65
		t-test	-0.626	1.066	-3.405	-1.878	-1.927
		p	0.532	0.288	0.001*	0.062	0.056
Associate's degree	NoE		2.34±0.84	3.77±0.94	2.14±0.87	2.28±0.98	2.63±0.59
	E		2.19±0.99	4.04±0.62	1.59±0.79	1.91±1.11	2.43±0.56
		t-test	1.063	-2.353	4.169	2.330	2.181
		p	0.289	0.020*	0.000*	0.021*	0.030*
BA	NoE		2.42±0.89	3.62±0.98	2.08±0.9	2.3±1.06	2.61±0.65
	E		2.14±0.89	4.12±0.6	1.8±0.83	1.99±1.00	2.51±0.51
		t-test	2.124	-4.095	2.198	2.005	1.092
		p	0.035*	0.000*	0.029*	0.047*	0.277

Note. NoE indicates no experience; E represents experience.
*p <0.05

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Book Review: *Handbook of Open Universities Around the World*

Editors: Sanjaya Mishra and Santosh Panda (Routledge, 2025, 740 pages). ISBN: 978-1-032-76361-3 (hardback, $225.00); ISBN: 978-1-032-75405-5 (paperback, $46.49); ISBN: 978-1-003-47819-5 (ebook, $46.49), https://doi.org/10.4324/9781003478195

Reviewed by: Tian Belawati, *Universitas Terbuka, Indonesia*

Overview

The Handbook of Open Universities Around the World, edited by Sanjaya Mishra and Santosh Panda, offers both a panoramic survey and a reflective critique of what *openness* truly means in higher education today. Drawing together insights from more than 100 scholars and practitioners, the editors have curated an extraordinary compilation that maps the histories, organizational structures, and innovations of 47 open universities across Africa, the Americas, Asia, Europe, and Oceania. The result is not only a celebration of institutional achievement but also an invitation to confront difficult questions about equity, sustainability, and the future of open learning.

Open universities were originally conceived as democratic institutions designed to remove barriers of geography, class, gender, and prior schooling. They opened doors to learners traditionally excluded from mainstream education systems. In the current era of rapid digital transformation, when artificial intelligence (AI) and data-driven technologies are reshaping how education is delivered and experienced, the notion of openness demands fresh examination. The *Handbook* situates itself precisely at this critical juncture, bridging historical foundations with emerging digital realities.

Conceptual Anchors of Openness

The editors anchor the *Handbook* in two complementary conceptual frameworks. The first is Lord Geoffrey Crowther's classic articulation of the "four opens"—open to people, to places, to methods, and to ideas—cited by Sir John Daniel in his opening chapter. Daniel reminds readers that open universities were designed with the core goal of reaching the unreached and promoting social justice and accessibility, using modular, self-instructional courses and flexible delivery systems. Measured by this compass, many institutions in the *Handbook* have remained loyal to the founding vision. For instance, the Indira Gandhi National Open University (IGNOU) continues to uphold a policy of nonselective entry and massive reach, while the UK Open University welcomes learners regardless of prior qualifications and supports them through a sophisticated blend of digital and human tutoring.

A more contemporary framework comes from Sanjaya Mishra, whose 10-dimension model of openness, developed in earlier studies of technology-enhanced learning, offers a detailed lens through which to assess institutional practice. His model examines 10 key areas: entry requirements, location, time flexibility, curricular choice, pedagogy, technology use, resource licensing, assessment, credential recognition, and cost transparency. Viewed through this multifaceted lens, the degree of openness across institutions appears uneven. While access and geographic reach remain broadly inclusive, other dimensions—particularly curricular flexibility, the adoption of open educational resources (OER), and innovative assessment methods—show significant variability. For example, Open University Malaysia has integrated recognition of prior learning and modular credentials, but many other institutions still adhere to rigid curricula and conventional high-stakes examinations.

Empirical Breadth and Depth

One of the *Handbook*'s greatest strengths lies in its empirical depth and comparative breadth. Each institutional chapter is grounded in robust data, lending credibility and concreteness to the analyses. Enrollment figures, for example, vividly illustrate the scale and diversity of open universities. IGNOU, serving more than 4 million learners, stands as the largest university in the world, while the Open University of China follows closely with over 3.5 million students. Universitas Terbuka in Indonesia, though smaller in relative terms, now serves more than 700,000 learners dispersed across a vast archipelago, a testament to the logistical and pedagogical challenges of reaching geographically scattered populations. At the other end of the spectrum, smaller institutions such as the Open University of Cyprus enroll tens of thousands, not millions, demonstrating that the open university model can function effectively at different scales.

Financial and operational information further enriches the discussion by exposing how institutions sustain themselves. The UK Open University illustrates how economies of scale and distance-delivery models contribute to reduced unit costs per student, while Universitas Terbuka's experience shows that a careful mix of government subsidies and affordable tuition can ensure both accessibility and sustainability. These examples suggest that cost-effectiveness is not simply a by-product of scale but a result of strategic financial design and prudent governance.

Technological practices vary just as widely. In low-resource contexts such as the Zimbabwe Open University or the National Open University of Nigeria, print-based materials and radio broadcasts still play a dominant role, complemented by modest digital initiatives. Conversely, universities such as Athabasca University in Canada and Korea National Open University have invested heavily in digital ecosystems, including mobile integration and AI-supported analytics. These diverse cases give readers a nuanced appreciation of how open universities operate at the intersection of aspiration, constraint, and innovation.

Comparative Themes

Despite the enormous diversity among the 47 institutions, the *Handbook* highlights strikingly consistent themes. Every open university is guided by the mission to democratize higher education and to serve learners who are marginalized or excluded from conventional systems. The University of South Africa, for instance, supports multilingual learning and Indigenous outreach; Allama Iqbal Open University in Pakistan advances gender equity through large-scale female enrollment initiatives; and Universitas Terbuka in Indonesia focuses deliberately on remote and island communities.

However, institutional differences also reflect national priorities and local contexts. The curriculum at Open University Malaysia emphasizes vocational and workforce-oriented programs, while Athabasca University in Canada champions liberal arts and flexible postgraduate studies. Accreditation and quality assurance mechanisms also diverge. Bangladesh Open University aligns closely with national regulatory frameworks, whereas Korea National Open University enjoys greater academic autonomy. These variations reveal that openness is not a monolithic concept but a flexible, adaptive philosophy responsive to sociopolitical and cultural environments.

Digital Transformation and Artificial Intelligence

A recurring theme throughout the *Handbook* is the transformative power of technology. The evolution from correspondence-based instruction to complex online ecosystems is carefully chronicled, illustrating how technological progress has expanded educational opportunities while simultaneously exposing new inequities.

Artificial intelligence receives special attention as both a promise and a challenge. At Universitas Terbuka, for example, AI tools are being introduced to enhance learning systems and streamline assessment processes. Across the global landscape, contributors acknowledge AI's potential to personalize learning, predict student performance, and provide timely academic support. Yet they also sound a note of caution. Ethical concerns, which range from algorithmic bias and data privacy to the depersonalization of learning, underscore the need for deliberate, human-centered AI design. The editors advocate an approach that is both optimistic and critical: AI should augment, not replace, the relational and ethical dimensions of education. This balanced stance provides valuable guidance for open universities as they navigate the digital frontier.

Strengths, Limitations, and Gaps

The most conspicuous strength of the *Handbook* is its sheer comprehensiveness. No other volume in the field of open and distance learning has attempted to document so many institutions across such a wide geographical and cultural span. By blending quantitative indicators such as enrollment, budgetary allocations, and staffing with qualitative narratives that trace institutional histories and policy reforms, the editors have created a work that is both detailed and panoramic.

Equally impressive is the conceptual coherence that underpins the collection. By framing institutional case studies within Daniel's and Mishra's frameworks, the *Handbook* demonstrates intellectual continuity with the foundational philosophies of openness while also situating them within present-day technological and policy realities.

Nevertheless, the book is not without its shortcomings. Some chapters are uneven in depth, largely due to disparities in available data or contributor access. More substantively, several thematic gaps remain.

First, the learner's perspective, particularly regarding hidden costs, opportunity costs, and time burdens, is underexplored. For many students, the cost of mobile data and the necessity of traveling to centralized examination sites represent substantial barriers. Second, the discussion of open educational resources and open educational practices is inconsistent. While a few universities, such as the UK Open University, have institutionalized open licensing, many others remain hesitant or silent. Third, innovation in assessment receives limited attention; traditional proctored examinations continue to dominate even though they may not adequately measure adult learners' competencies. Fourth, the ethical, social, and environmental dimensions of open learning—issues like data privacy, vendor dependency, and the ecological implications of mass printing—are touched on only peripherally. Finally, the absence of learner voice and agency stands out as a missed opportunity, given that responsiveness to learners is at the very heart of the open education ethos.

Significance and Contribution

Despite these limitations, the *Handbook of Open Universities Around the World* represents a monumental scholarly and practical achievement. It reaffirms the enduring relevance of open universities as institutions that combine scale, equity, and flexibility in unique and powerful ways. Beyond serving as an encyclopedic reference, the volume offers comparative insights that allow policymakers and researchers to benchmark institutional performance and share best practices across regions.

Perhaps the *Handbook*'s most significant contribution lies in its portrayal of openness as an evolving philosophy rather than a static condition. Openness, the editors suggest, must be continually reinterpreted in response to technological change, shifting policy environments, and learners' changing expectations. By embedding rich empirical data within coherent conceptual frameworks, the *Handbook* succeeds both in documenting the current state of the field and in imagining its future trajectories.

Looking ahead, future editions could build upon this foundation by incorporating learner-centered narratives, especially from marginalized or underserved groups, to capture the lived experience of open learning. They might also provide more systematic analyses of open education policies and OER adoption; develop clearer benchmarks for institutional openness; and showcase innovative models of assessment and credentialing, including digital badges, e-portfolios, and competency-based recognition systems. Furthermore, addressing ethical and ecological concerns such as AI governance, data protection, and environmental sustainability would enhance the comprehensiveness of future research. Finally, longitudinal studies on graduate employability, social mobility, and community impact would offer valuable evidence of the real-world outcomes of openness.

Conclusion

In summary, the *Handbook of Open Universities Around the World* is an indispensable resource for researchers, practitioners, and policymakers in open and distance learning. Its combination of global scope, empirical rigor, and conceptual sophistication makes it a landmark contribution to literature. While not exhaustive in its treatment of openness, it convincingly demonstrates that open universities remain central to the future of higher education—that they are institutions uniquely capable of balancing access, quality, and innovation in an era of massification and digital disruption.

By situating open universities within both Daniel's foundational ideals and Mishra's multidimensional model, the *Handbook* reminds readers that openness is not merely the removal of barriers but an ongoing process of reimagining what higher education can and should be in a rapidly changing world.

Disclaimer

The writing of this review was done with the help of ChatGPT for outlining.

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

Regulation of Distance Learning Courses in Brazilian Higher Education: A Critical Review of Decree No. 12,456/2025 and Ordinance No. 378/2025

Thais Barbosa Reis[1], Nathalie Barbosa Reis Monteiro[2], Natacha Bertoia[2], Claudia Fernanda Franceschi Klement[2], Karen Perrotta Lopes de Almeida Prado[2], Alessandra Zago Dahmer[2], Miriam Rodrigues[2], and Eric Ferdinando Passone[1]
[1]Universidade Cidade de São Paulo (UNICID); [2]Universidade Presbiteriana Mackenzie (UPM)

Abstract

This field note examines the recent regulatory framework for distance higher education in Brazil, analyzing the implications of Decree 12,456/2025 and Ordinance 378/2025. Through critical analysis, we assessed the alignment of these measures and their potential impacts on educational quality, accessibility, and institutional accountability. We examined the measures designed to balance the expansion of access with the assurance of quality, including mandated percentages of in-person and synchronous activities, redefined faculty roles, and restrictions on institutional sharing. While acknowledging the potential to enhance academic rigor and curb low-quality programs, the analysis highlighted significant implementation challenges. These include increased operational costs, potential impacts on tuition, and concerns that restricting teacher education to blended or in-person modalities may exacerbate teacher shortages in remote areas. The study concluded that the new framework's ability to reduce inequalities and improve employability depends on financial support, vigilant oversight, and further research, offering a valuable case study for global debates on regulating digital higher education.

Keywords: distance education, regulation, higher education, educational quality, public policies

Regulation of Distance Learning Courses in Brazilian Higher Education: A Critical Review of Decree No. 12,456/2025 and Ordinance No. 378/2025

Decree No. 12,456 of May 19, 2025 (Government of Brazil, 2025a), together with Ordinance 378/2025 (Government of Brazil, 2025b), published on the same date, established a new regulatory framework for distance higher education (DE) in Brazil and a new public education policy for this modality. These legal instruments were intended to reconcile the expansion of access to higher education with the maintenance of quality standards, setting guidelines for the provision of courses through in-person, distance learning, and blended formats. In this article, we examined the key aspects of these regulations, highlighting their contributions, innovations, and potential challenges for practical implementation in the Brazilian educational system.

Decree 12,456/2025 (Government of Brazil, 2025a) had, as its central objectives, promoting access to quality higher education, reinforcing criteria such as social responsibility and transparency in the provision of DE courses, valorizing teaching, and defining specific rules for in-person support centers. Ordinance 378/2025 (Government of Brazil, 2025b) complemented these provisions by detailing course delivery formats, establishing minimum percentages of in-person and synchronous workload for each modality. This new regulatory framework aimed to ensure that the flexibility of distance learning courses did not compromise teaching quality, pedagogical practices, or dialogic communication among instructors, facilitators, and students, aligning with recently established guidelines for expanding access and improving professional training.

However, these regulations also introduced technical definitions that reflected methodological interventions that could be interpreted as constraints on the pedagogical autonomy guaranteed by the Brazilian Federal Constitution of 1988 (Government of Brazil, 1988). Among these were requirements such as (a) synchronous mediated activities, capped at 70 students per instructor with attendance monitoring, justified to ensure effective pedagogical interaction; and (b) the imposition of in-person exams, which favor discursive questions that demand critical analysis, synthesis skills, and account for one-third of the total evaluation weight.

The concept of DE support centers as decentralized units of higher education institutions (HEIs), with a prohibition on inter-institutional sharing, aimed to prevent the proliferation of substandard structures and ensure minimum infrastructure standards. While this measure was intended to prevent excessive sharing practices that could compromise infrastructure quality and ensure adequate operational conditions, it may pose operational challenges, particularly for small institutions or those in regions with limited internet service coverage. Furthermore, the new regulation (Decree No. 12,456/2025 and Ordinance No. 378/2025) reinforced the importance of curricular units as components linked to the course pedagogical project (CPP), that has “as a priority function to guide and intentionally conduct the pedagogical process in the daily life of the classroom” (Silva et. al., 2021, p. 468), thus emphasizing faculty responsibility in teaching and assessment processes.

The rules established by the decree and ordinance outline criteria for each instructional modality.

- In-person courses must offer at least 70% of contact hours (CH) as face-to-face activities.
- Blended/hybrid courses (a newly introduced category) require 30% face-to-face CH and 20% in either in-person or mediated synchronous activities.
- DE courses must include at least 10% in-person activities and 10% mediated synchronous classes, without exceeding the limits established for blended courses.

In-person delivery, exclusively, was to be maintained for programs such as Law, Medicine, and Nursing, reflecting concerns about practical training in these fields. Conversely, Engineering and Health Sciences programs may adopt the blended/hybrid format, provided they maintain 40% of CH in face-to-face instruction. This flexible approach signaled an attempt to balance increased flexibility with minimal structural requirements for fields that require equilibrium between theoretical and practical components.

The new regulatory framework established guidelines that emphasized the valorization of teaching roles and clearly distinguished among lead instructors, content specialists, and pedagogical facilitators; the latter require appropriate academic qualifications. Tutors, in turn, were restricted to administrative functions, a measure that, according to the regulation's objectives, may be associated with enhancing educational quality standards.

Decree 12,456/2025 and Ordinance 378/2025 constituted a significant milestone in consolidating DE in Brazil by establishing a balance between flexibility and academic rigor. However, their implementation may face substantial challenges, including the need for significant infrastructure investments and ongoing faculty development. Emerging evidence (though not yet systematically studied) has suggested that these regulatory requirements may increase operational costs, potentially affecting tuition fees.

Teacher education programs may encounter considerable impact; under the new framework, these programs have been restricted to blended or in-person modalities. This policy shift has raised concerns, as the distance education modality had previously enabled teacher training in remote and underserved areas, precisely those regions experiencing the most severe shortages of basic education teachers and education professionals.

Furthermore, the impact of these regulations on reducing educational inequalities and enhancing graduate employability warrants further study. While they may represent progress in improving the quality of DE programs in Brazil, their success will depend on financial support policies and ongoing monitoring. The effectiveness of these regulatory measures will be contingent upon institutions' capacity to internalize the normative guidelines, responsive oversight by regulatory bodies, and active civil society engagement in monitoring public policies. Table 1 summarizes the key changes introduced by Decree 12,456/2025 and Ordinance 378/2025 for distance education.

Table 1

Comparison of Previous Legislation and New Regulatory Framework

Category	Previous legislation	New regulatory framework
Program formats	In-person and DE	In-person, blended, and DE
CH in-person Programs	Maximum 40% DE content	Maximum 30% DE content
CH blended programs	Unregulated	30% in-person plus 20% in-person or mediated synchronous
CH DE programs	Maximum 30% in-person	Minimum 10% in-person plus 10% mediated synchronous activities
Institutional accreditation	Separate for in-person/DE	Unified accreditation
Faculty composition	Professors and tutors (pedagogical mediation)	Program coordinator, lead instructor, and content specialist. (Lead instructors may be assisted by qualified pedagogical facilitators. Tutors restricted to administrative roles.)
Learning assessment	Unregulated	Mandatory periodic in-person assessments with majority weight and 1/3 dedicated to discursive questions testing analysis/synthesis skills.
Campus sharing	Unlimited sharing permitted	Prohibited for different HEIs at same address (except for National Apprenticeship System).
Partnership campuses	Unrestricted	Permitted with conditions. Professors must be hired by main institution. Student contracts must be with main institution.

The Ministry of Education's quality benchmarks for distance undergraduate programs (Ministry of Education 2025) provided a comprehensive and updated framework on the essential elements for ensuring the academic quality of DE undergraduate programs. The previous document that addressed these benchmarks was nearly 20 years old (Ministry of Education, 2007).

Structured around key dimensions, including program design, student experience, teaching practices, learning methodologies, assessment, and infrastructure, the new framework emphasized student-centered learning at the core of educational development, with active and inclusive pedagogical mediation as a critical requirement. Key requirements included a CPP aligned with national curriculum guidelines, enhanced emphasis on interactivity and personalized learning, the structural role of in-person support centers and professional practice environments, as well as the intentional pedagogical integration of educational technologies.

Regarding teaching practice, the establishment of continuous faculty development programs have been recommended, focused on pedagogical mediation in synchronous environments and the design of critical discursive assessment. In terms of theory, studies to analyze the impact of these new regulations on dropout rates, learning quality, and the perpetuation of regional inequalities are will be imperative, thereby refining DE theoretical models in light of a more restrictive regulatory context.

In conclusion, the newly analyzed regulatory framework represents a significant endeavor to strike a balance between the flexibility of distance education and the assurance of academic quality in Brazil. For Brazilian educators, these regulations have profound implications and entail significant changes, as they necessitate curricular adaptation, mastery of new competencies for hybrid and synchronous pedagogical mediation, and a redefined, more specialized role within distance learning ecosystems.

Beyond the national context, this case presents a valuable study for other nations facing the challenges of regulating digital higher education. Attempts to curb the proliferation of low-quality programs, establish clear interaction parameters, and enforce critical in-person assessments serve as a crucial experiment. Outcomes such as costs, access, and effectiveness will be closely observed by policymakers, institutions, and educators worldwide, thereby contributing to the global discourse on how to expand access to higher education without compromising educational rigor and integrity.

Furthermore, the new framework has reaffirmed commitments to universal access and digital inclusion, sustainable educational practices, and institutional governance. These proposed standards aim not only to expand DE offerings but also to establish distance education as a rigorous and legitimate modality, as well as a socially responsible pathway within Brazilian higher education.

Acknowledgements

While writing this field note, the authors made use of Grammarly (Pro) to polish some language in the text. We confirm that we have reviewed and edited the content as necessary, and we take full responsibility for it.

References

Government of Brazil. (1988). *Constituição da República Federativa do Brasil de 1988* [Constitution of the Federative Republic of Brazil of 1988]. Official Gazette of the Union. https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Government of Brazil. (2025a). *Decreto nº 12.456, de 19 de maio de 2025* [Decree No. 12,456 of May 19, 2025]. Official Gazette of the Union. https://www.in.gov.br/en/web/dou/-/decreto-n-12.456-de-19-de-maio-de-2025-630398639

Government of Brazil. (2025b). *Portaria MEC Nº 378, de 19 de maio de 2025* [MEC Ordinance No. 378, of May 19, 2025]**.** Official Gazette of the Union. https://www.in.gov.br/en/web/dou/-/portaria-mec-n-378-de-19-de-maio-de-2025-630395302.

Ministry of Education. (2007). *Referenciais de qualidade para educação superior a distância* [Quality benchmarks for distance higher education]. Ministry of Education/Distance Education Secretariat. https://portal.mec.gov.br/seed/arquivos/pdf/legislacao/refead1.pdf.

Ministry of Education. (2025). *Referenciais de qualidade de cursos de graduação com oferta a distância* [Quality benchmarks for undergraduate courses offered via distance learning]. Ministry of Education/Distance Education Secretariat. https://www.gov.br/mec/pt-br/politica-regulacao-supervisao-educacao-superior/ead/documentos/referenciais_qualidade.pdf.

Silva, B. C.; Oliveira, L. M. S. R.; Moreira, M. B. (2021). Pedagogical project of undergraduate courses and teacher education. *International Journal of Advanced Engineering Research Science (IJAERS)*, 8(11), 468-473. https://dx.doi.org/10.22161/ijaers.811.49

International Review of Research in Open and Distributed Learning
Volume 27, Number 1

February – 2026

Digital Literacy in Enhancing Collaborative Teaching: A Systematic Review

Walton Wider[1,2,3,4], Abidah Saad[5]*, Nor Asiah Mahmood[6], Suhaida Ishak[7], Muhammad Ridzuan Abdul Aziz[8], Changhe Wu[1], Hao Wu[1], and Jem Cloyd M. Tanucan[9]
*[1]Faculty of Business and Communications, INTI International University, Nilai, Malaysia; [2]Department of Applied Economic Sciences, Wekerle Sándor Üzleti Főiskola, Budapest, Hungary; [3]Faculty of Management, Shinawatra University, Pathum Thani, Thailand; [4]Department of Economics, Management, and Law, International Institute of Management and Business, Minsk, Belarus; [5]Faculty of Business and Management, University Teknologi MARA, Kedah Branch, Kampus Sungai Petani, Malaysia; [6]College of Business Administration, Prince Sultan University, Riyadh, Saudi Arabia; [7]Faculty of Business, Information & Human Sciences, Infrastructure University Kuala Lumpur, Kajang, Malaysia; [8]Faculty of Business and Management, Universiti Teknologi MARA, Selangor Branch, Puncak Alam Campus, Malaysia; [9]School of Education, Norfolk State University, Norfolk, United States; *Corresponding Author*

Abstract

Digital literacy is central to collaborative teaching in technology-mediated environments, particularly open and distributed learning. Guided by the Community of Inquiry and TPACK (Technological Pedagogical Content Knowledge) frameworks, this systematic review examines how digital literacy enables educators to codesign instruction, sustain interaction, and support reflective practice while addressing structural and contextual barriers. Following PRISMA 2020, comprehensive searches in Scopus and the Web of Science identified 32 peer-reviewed articles published in 2024. Thematic synthesis produced three strands: (a) integration of digital literacy in education, highlighting links to teaching presence, professional development, and instructional design; (b) digital literacy in response to educational challenges, demonstrating its role in resilience, equity, and socio-emotional support across remote and hybrid contexts; and (c) advancing learning through digital competencies, detailing gains in collaboration, critical inquiry, and innovative use of augmented reality, virtual reality, data analytics, and emerging AI tools alongside ethical considerations. Evidence indicates that digital literacy functions as a pedagogical capacity rather than solely a technical skill and yields the strongest outcomes when aligned with institutional culture, curriculum design, and continuous professional learning. Policy recommendations include sustained investment in equitable infrastructure, structured capacity building aligned with UNESCO's Digital Literacy Global Framework and ICT (Information and Communication Technology) Competency Framework for Teachers, and explicit attention to ethics and inclusion. Future research should adopt longitudinal and comparative designs to trace the impact on educator identity, collaboration, and learner outcomes.

Keywords: digital literacy, collaborative teaching, systematic review, Community of Inquiry, TPACK, open and distributed learning, educational policy

Introduction

Rapid advancements in digital technology have profoundly reshaped the educational landscape, particularly in collaborative teaching practices (Rahimi & Oh, 2024). In contemporary higher education, digital tools and platforms have become essential for promoting communication, innovation, and active participation among educators and learners (Zorko, 2024; Zou et al., 2025). Collaborative teaching, which involves educators working together to design, implement, and assess instructional activities, increasingly depends on digital environments that enable co-construction of knowledge and continuous professional development (Ramos et al., 2022).

Despite the widespread adoption of technology, many educators continue to face challenges in effectively integrating digital literacy into collaborative teaching. Insufficient digital competence, unequal access to technological infrastructure, and limited institutional support hinder educators from maximizing the pedagogical potential of digital platforms (Irwandi, 2023; Jony & Sultana, 2023). In addition, variations in digital proficiency across educational levels and geographic regions contribute to inconsistent implementation of technology-mediated teaching (Nuhu et al., 2021; Oliiar & Fomin, 2022). Existing professional development programs often focus on technical skills rather than the pedagogical, ethical, and reflective aspects of digital engagement, which are essential for collaborative and innovative instruction (Çetinkaya, 2024; Kwiatkowska & Wiśniewska-Nogaj, 2022).

The COVID-19 pandemic intensified the demand for digital competence as institutions rapidly transitioned to remote and hybrid learning models (Alqahtani, 2024). This shift exposed disparities in technological readiness and pedagogical adaptability, emphasizing digital literacy as a key determinant of educational resilience, equity, and continuity (Blau et al., 2020; Perifanou et al., 2021). While prior studies have highlighted the importance of digital literacy in facilitating collaborative learning and reflective practice, few studies have systematically synthesized how established digital competence frameworks can inform effective collaborative teaching within open and distributed learning environments (Haryaka et al., 2025; Huachara-Martinez et al., 2023).

This study addresses this gap by examining how digital literacy functions in both a pedagogical and a collaborative capacity that empowers educators to design inclusive, interactive, and sustainable learning environments. Drawing on the Community of Inquiry and TPACK (Technological Pedagogical and Content Knowledge) frameworks, this review explores how digital literacy enhances collaboration, communication, and reflective practice in technology-enhanced education. The findings aim to contribute to theory and practice by guiding educators, policymakers, and researchers to strengthen professional collaboration in digital contexts.

Theoretical Framework

The Community of Inquiry (CoI) framework proposed by Garrison et al. (2001) explains that meaningful learning in online and blended environments results from the interaction of three interdependent elements: cognitive, social, and teaching presence. Digital literacy supports this presence by allowing educators to design and manage interactive environments (teaching presence), engage learners in inquiry and reflection (cognitive presence), and foster authentic communication and collaboration (social presence). Empirical research confirms the applicability of CoI across disciplines and learning contexts, demonstrating that its presence is essential for developing engagement and achieving higher-order learning outcomes (Cleveland-Innes, 2019; Micsky & Foels, 2019; Swan, 2019). In particular, teaching presence has been linked to student satisfaction and instructional quality in remote and emergency online learning (Patwardhan et al., 2020). More recent developments have introduced the concept of learning presence, which incorporates self-regulation and digital competence as critical elements that influence cognitive engagement (Wertz, 2022). Together, these findings suggest that digital literacy strengthens educators' ability to create learning communities characterized by critical thinking, collaboration, and shared knowledge construction (Purwandari et al., 2022).

The TPACK framework developed by Mishra and Koehler (2006) complements CoI by emphasizing the integration of technological, pedagogical, and content knowledge for effective digital instruction. Digital literacy serves as the foundation for this integration, enabling educators to align digital tools with pedagogical strategies and curricular goals. Teachers with higher digital literacy competencies are better equipped to select and adapt digital technologies to enhance their instructional quality and collaboration (Barjestesh et al., 2025). This process is closely related to the teaching presence and instructional design principles derived from the CoI model (Richardson et al., 2012).

Together, CoI and TPACK provide a robust theoretical foundation for understanding digital literacy as both a pedagogical and an ethical construct that promotes collaboration in distributed learning environments. These frameworks have been successfully applied in teacher education programs to encourage reflective practices, professional discourse, and collaboration in technology-rich settings (Papanikolaou et al., 2014). TPACK also guides curriculum development that integrates digital literacy into teaching, fostering critical thinking, creativity, communication, and collaboration between educators and students (Jordan et al., 2025). Both models align with global initiatives such as UNESCO's Digital Literacy Global Framework (DLGF) and the ICT (Information and Communication Technology) Competency Framework for Teachers (ICT-CFT), which emphasize inclusivity, accessibility, and responsible technology use in education (Asagar, 2025; Choudhary, 2024). The integration of these theoretical and policy perspectives underscores the ethical dimension of digital literacy as a driver of digital citizenship, equity, and empowerment in education (Raza & Akhter, 2024).

Literature Review

According to Law et al. (2018), *digital literacy* is the ability to access, manage, understand, integrate, communicate, evaluate, and create information safely and appropriately using digital technologies for employment, decent work, and entrepreneurship. This definition captures its multidimensional character, encompassing the technical, cognitive, and socio-emotional competencies required for effective participation in digital environments (Tinmaz et al., 2023). Ng (2012) expanded on this by identifying three interrelated dimensions—technical, cognitive, and socio-emotional—that enable individuals to use and create digital content meaningfully and ethically. These competencies are not limited to operational skills but extend to reflective practice, collaboration, and ethical engagement with digital tools (Rodríguez-García et al., 2022).

Zhao et al. (2021) contextualized digital literacy and digital competence within the transformation from industrial and information-based economies to knowledge societies. They argue that digital competence distinguishes the knowledge society by emphasizing the application of digital tools for generating and using knowledge, rather than for merely accessing information. The COVID-19 pandemic further accelerated this shift, making digital competence essential for sustaining teaching and learning across educational levels. Within this framework, *digital competence* is defined as the confident, critical, and responsible use of digital technologies in communication, learning, and work (Sánchez-Canut et al., 2023).

Although digital literacy and competence are conceptually related, they differ in their focus. Digital literacy emphasizes understanding and interpreting digital information, whereas digital competence extends to critical awareness, lifelong learning, and practical applications. These concepts are operationalized in the European Digital Competence Framework and Spain's Common Digital Competence Framework for Teachers, which identifies 21 specific competencies and proficiency levels for assessment and training (Mattar et al., 2022). UNESCO's global digital literacy framework complements these efforts by establishing internationally recognized benchmarks that align digital competencies with broader educational and socioeconomic goals (Gabriel et al., 2022; Jung et al., 2024).

Empirical research has consistently demonstrated that digital competence contributes to effective collaboration, critical thinking, and innovation in higher education (Männistö et al., 2020; Ohle-Peters et al., 2024). Studies integrating cooperative learning and flipped classroom models report significant improvements in preservice teachers' digital literacy, pedagogical competence, and 21st-century skills (Aslan, 2022). Similarly, structured environments such as Future Classroom Labs promote digital competence through guided collaboration, provided sustained teacher support is available (Lazareva & Tømte, 2024).

Digital literacy also underpins open and distributed learning systems by enhancing equity, learner autonomy, and reflective engagement (Huachara-Martinez et al., 2023). Within these contexts, digital literacy acts as both an enabler of collaboration and a determinant of instructional quality. However, these challenges persist. Variations in digital proficiency, inadequate access to technology, and limited institutional investment continue to restrict educators' ability to effectively integrate digital literacy (Irwandi, 2023; Jony & Sultana, 2023). Furthermore, ongoing technological evolution requires educators to continually update their competencies to maintain relevance and pedagogical quality (Nuhu et al., 2021; Oliiar & Fomin, 2022).

Institutions that integrate digital literacy into professional development frameworks have reported stronger collaboration and improved educational outcomes (Lora et al., 2021; Perifanou et al., 2021). Structured training programs not only enhance educators' technical capabilities but also cultivate ethical awareness and reflective practice. In the post-pandemic context, digital literacy is increasingly recognized as a pedagogical capacity rather than merely a technical skill. This enables educators to design inclusive learning environments that emphasize engagement, creativity, and communication (Azzahro et al., 2023; Darmaji et al., 2023; Ervianti et al., 2023). Nevertheless, disparities in professional training and access to it remain significant, especially in vocational and secondary education settings (Adeleye et al., 2024; Qiu et al., 2024). Addressing these inequities requires targeted interventions that integrate both technical and collaborative competencies to build inclusive participatory learning ecosystems (Karroum & Elshaiekh, 2023).

Scope and Relevance

This systematic review synthesizes recent studies that examine the intersection of digital literacy and collaborative teaching across secondary, vocational, and higher education contexts. Anchored in the CoI and TPACK frameworks, digital literacy is conceptualized as both a pedagogical and technological foundation for effective collaboration, communication, and reflective teaching within open and distributed learning (ODL) environments.

The CoI framework provides analytical perspectives on how cognitive, social, and teaching presence are strengthened through digital literacy. These presences serve as lenses for examining how educators codesign learning activities, sustain engagement, and facilitate dialogue in distributed contexts. The TPACK framework complements this approach by demonstrating how technological, pedagogical, and content knowledge interact to shape effective digital instruction. Together, these frameworks explain how digital literacy enables educators to align technology with pedagogical goals, thereby enhancing collaboration and instruction quality.

This review includes studies published in recent years that address digital literacy in post-pandemic educational contexts. The pandemic has accelerated digital adoption and highlighted the urgent need for inclusive, resilient, and sustainable teaching practices. Therefore, this review focuses on evidence illustrating how digital literacy enhances adaptability, innovation, and collaboration among educators operating in technology-mediated environments.

The findings are expected to inform educators, policymakers, and researchers by providing evidence-based strategies to strengthen digital literacy training and collaborative teaching practices consistent with the CoI and TPACK principles. The review advances the theoretical and practical understanding of how digital literacy underpins professional growth, collaborative engagement, and pedagogical innovation in contemporary education. It also contributes to global conversations on teacher readiness and equity in technology-enhanced learning, reinforcing the importance of digital competence as a foundation for lifelong learning and educational transformation.

Material and Methods

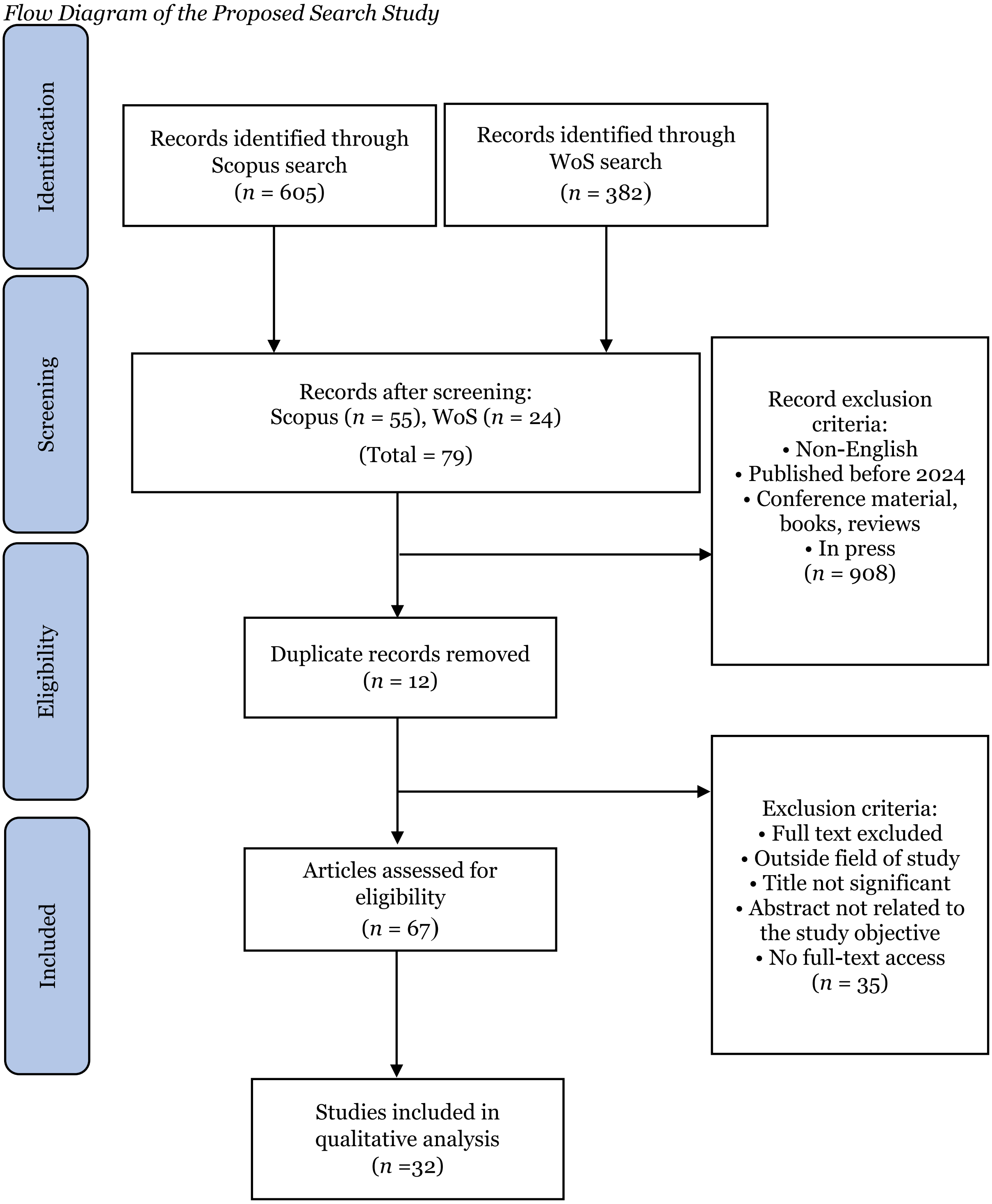
Identification

This study followed the PRISMA 2020 guidelines for systematic reviews to ensure methodological transparency and reproducibility (Figure 1). In the identification phase, comprehensive search strings derived from key concepts related to digital literacy and collaborative teaching were developed. Keywords and their synonyms were expanded through the use of dictionaries, thesauri, and prior literature on technology-enhanced learning and ODL.

Two major databases, Scopus and Web of Science (WoS), were selected because of their wide coverage of peer-reviewed educational research. Finalized search strings (Table 1) were applied to both databases in January 2025. To capture the most current post-pandemic developments, the review included studies published from 2024 onward, a period marked by significant transitions to online and hybrid teaching. Pre-2024 literature was excluded because earlier works predated the rapid technological and pedagogical transformations catalyzed by the COVID-19 pandemic. The initial search yielded 987 publications relevant to digital literacy and collaborative teaching.

Figure 1

Flow Diagram of the Proposed Search Study



Note. WoS = Web of Science.

Table 1

Search Strings Used for Scopus and Web of Science (WoS)

Database	Search string
Scopus	TITLE-ABS-KEY ((“Digital Literacy” OR “Computer Proficiency”) AND (“online teach*” OR (collab* AND teach*)))
WoS	(“Digital Literacy” OR “Computer Proficiency”) AND (“online teach*” OR (collab* AND teach*))

Screening

The screening stage aimed to refine the dataset to include only those studies that aligned with the research objectives. During this phase, duplicate records were removed, and the titles, abstracts, and keywords of the remaining articles were carefully assessed for their relevance to digital literacy and collaborative teaching. Of the 987 initially retrieved records, 908 were excluded because they did not meet the predefined inclusion criteria. Studies were retained if they satisfied the following conditions: written in English, published from 2024 onward, categorized as peer-reviewed journal articles, and available in their final published versions. Non-journal sources, including books, book chapters, reviews, meta-analyses, and conference proceedings, were excluded to ensure the inclusion of empirical and up-to-date evidence. Following the removal of duplicates and ineligible records, 79 studies proceeded to the eligibility stage.

Eligibility

During the eligibility phase, the titles, abstracts, and full texts of the remaining publications were critically reviewed to ensure alignment with the study’s conceptual focus. This evaluation followed the principles of both the CoI and TPACK frameworks, emphasizing research that addressed cognitive, social, and teaching presence, or technological–pedagogical integration within collaborative contexts.

Of the 79 studies screened, 32 met the inclusion criteria (Table 2). The excluded articles were out of scope, lacked empirical grounding, or did not provide sufficient detail for the analysis. Several publications were excluded owing to restricted access to the full text, which is essential for reliable synthesis and interpretation.

Table 2

Number and Details of Primary Studies Database

No.	Author(s) (Year)	Title	Journal	Scopus	WoS
1	Sohel et al. (2024)	E-Learning Experience of Indigenous Rural Communities in the Face of COVID-19 Crisis in Chittagong Hills Tracts Region, Bangladesh: A Qualitative Investigation	*International Journal of Community Well-Being*	✓	
2	López-Meri et al. (2024)	Digital Competencies in Verifying Fake News: Assessing the Knowledge and Abilities of Journalism Students	*Societies*	✓	✓
3	Deiniatur et al. (2024)	English Teachers' Beliefs and Practices in Integrating Digital Literacy in the Language Classroom	*International Journal of Evaluation and Research in Education*	✓	
4	Kee et al. (2024)	An Empirical Study on Immersive Technology in Synchronous Hybrid Learning in Design Education	*International Journal of Technology and Design Education*	✓	
5	Zhang et al. (2024)	Digital Literacy Competence, Digital Literacy Practices and Teacher Identity Among Pre-Service Teachers	*Journal of Education for Teaching*	✓	
6	Nevrelova et al. (2024)	Enhancing Digital Literacy in Primary Education Through Augmented Reality	*Frontiers in Education*	✓	✓
7	Rasdiana et al. (2024)	Mediation of Digital Literacy in Investigating the Effect of School Culture on Teacher Performance: Implication for Educational Policy	*Journal of Infrastructure, Policy and Development*	✓	
8	Deiniatur & Cahyono (2024)	Digital Literacy Practices of Novice English as a Foreign Language Teacher in Writing Research Articles for Publication	*Journal of Education and Learning*	✓	
9	Kurniawan (2024)	Enhancing 21st-Century Writing Skills Through a Reflective Collaborative Learning Model Based on Critical Thinking	*Pakistan Journal of Life and Social Sciences*	✓	

10	Achruh et al. (2024)	Challenges and Opportunities of Artificial Intelligence Adoption in Islamic Education in Indonesian Higher Education Institutions	*International Journal of Learning, Teaching and Educational Research*	✓	
11	Xu & Tan (2024)	Beyond Words: L2 Writing Teachers' Visual Conceptualizations of ChatGPT in Teaching and Learning	*Journal of Second Language Writing*	✓	✓
12	Tangkish et al. (2024)	Digital Mind and Human Consciousness: Integration of Digital Technology in Shaping Learning Experiences	*Perspektivy Nauki i Obrazovania*	✓	
13	Molina-Torres (2024)	Flipped Classroom to Teach Digital Skills During COVID-19	*Journal of Technology and Science Education*	✓	
14	Nash (2024)	Critical Inquiry in (and About) Media Environments: Examining an Asset-Based Digital Literacy Curriculum	*Journal of Literacy Research*	✓	
15	Folabit & Jita (2024)	Are Academics Adapting to Students' Technology Learning Preferences? A South African Study of Teaching Identities	*Issues in Educational Research*	✓	✓
16	Kumalasari et al. (2024)	Comparative Analysis of Generation Z's Digital History Literacy in History Education Majors on Java Island: A Study of History Digital Literacy	*Journal of Education and E-Learning Research*	✓	
17	Zakariya et al. (2024)	Affordances and Constraints of a Blended Learning Course: Experience of Pre-Service Teachers in an African Context	*Humanities and Social Sciences Communications*	✓	✓
18	Bergstrom (2024)	"I Thought It Was an Accident": Digital Literacy and MLL Use of Collaborative Writing Software in Post-Secondary Composition Classes	*Computer-Assisted Language Learning Electronic Journal*	✓	
19	Fathali et al. (2024)	Digital Literacy and EFL Teachers' Anxiety with Teaching Online via Virtual Classroom Software	*The JALT CALL Journal*	✓	

20	Hasanah et al. (2024)	Exploring the Correlation of Self-Perception on the Use of Digital Literacy in Learning	*International Journal of Evaluation and Research in Education*	✓	
21	Bruckhaus et al. (2024)	Evaluation of Students' Digital Literacy Through an Immersive University–High School Collaboration	*Frontiers in Education*	✓	
22	Momdjian et al. (2024)	A Comparison of Perceptions of Digital Competences of Schoolteachers to School Leaders in Lebanon	*Social Sciences and Humanities Open*	✓	
23	Cabaron (2024)	Exploring the Impact of Digital Literacy on the Self-Efficacy of Maritime Education Faculty	*International Journal of Advanced and Applied Sciences*	✓	✓
24	Mhlongo et al. (2024)	Mathematics Teachers' Experiences of Using Online Teaching Resources for Professional Learning in a Context of Disadvantage	*International Journal of Learning, Teaching and Educational Research*	✓	
25	Li & Zhang (2024)	Embedding Digital Literacies in the Language Teacher Education Curriculum: Pre-Service and In-Service Teachers' Perspectives	*CALICO Journal*	✓	
26	Alqahtani (2024)	Nurse Educators and Faculty Members Challenges Towards Online Teaching During COVID-19 Pandemic Crisis: A Qualitative Descriptive Study	*Nursing Open*	✓	✓
27	Kiryakova & Kozhuharova (2024)	The Digital Competences Necessary for the Successful Pedagogical Practice of Teachers in the Digital Age	*Education Sciences*	✓	✓
28	Dabengwa et al. (2024)	Exploring Digital Competences in Zimbabwean Secondary Schools Using a Multimodal View: A Hermeneutical Phenomenography Study	*Cogent Education*	✓	✓
29	van Staden & Lotz-Sisitka (2024)	E-Learning as a Mediating Tool to Support Interactive Professional Learning of Teacher Educators	*Interactive Learning Environments*	✓	

30	Kurniawati et al. (2024)	A Case Study of Millennial English Teachers' Awareness of Digital Media in EFL Classrooms	*Studies in Media and Communication*	✓	
31	Brante et al. (2024)	Eight-Year-Olds Engaging in Guided Information Searches with iPads: Dimensions of Reading Competence	*International Journal of Child–Computer Interaction*	✓	
32	Dharmayanti et al. (2024)	Digital Literacy Competence for Scientific Writing: Students' Perceptions and Skills	*Journal of Language Teaching and Research*	✓	

Note. WoS = Web of Science.

Data Abstraction and Analysis

The data synthesis phase employed an integrative thematic analysis approach that combined qualitative coding with descriptive synthesis to identify emerging themes and subthemes. Each of the 32 included studies was systematically examined to extract information on the research objectives, methodologies, and findings related to digital literacy, collaborative teaching, and ODL principles. The analysis was guided by the CoI and TPACK frameworks, ensuring that the derived themes captured both the pedagogical and technological dimensions of collaboration.

The coding process was conducted over several stages. The first stage involved familiarization with each study to identify recurring concepts associated with digital literacy and collaboration. Thematic categorization was carried out, where the data were organized into subthemes aligned with the three presences of the CoI framework—cognitive, social, and teaching—as well as the intersections of technological, pedagogical, and content knowledge emphasized in the TPACK framework. In the refinement and integration stage, overlapping themes were merged and discrepancies were resolved through collective author discussions to achieve conceptual coherence.

To ensure transparency and rigor, a reflective audit log was maintained throughout the process to record coding decisions, interpretations, and emerging insights. An expert validation phase was subsequently conducted to strengthen the reliability and domain validity of the thematic analysis. Three academic experts with more than 5 years of experience in educational technology and social sciences independently reviewed the thematic structure and assessed its clarity, consistency, and conceptual alignment. Their feedback informed minor revisions that enhanced the thematic and interpretive depth.

This rigorous synthesis process enabled the identification of the central research question guiding the review: How does digital literacy support and enhance collaborative teaching practices in ODL environments? The final themes were interpreted through the lens of the CoI and TPACK frameworks to illustrate how digital literacy facilitates interaction, pedagogical innovation, and reflective collaboration among educators in technology-rich educational settings.

Results and Discussion

A synthesis of 32 studies identified three overarching themes that illustrate how digital literacy supports collaborative teaching and learning within the context of ODL: (a) integration of digital literacy in education, (b) digital literacy in response to educational challenges, and (c) advancing learning through digital competencies. Each theme aligns with the CoI and TPACK frameworks' theoretical principles. The findings demonstrate that digital literacy enhances communication, collaboration, and pedagogical innovation, but they also reveal the structural and contextual challenges that hinder its full potential.

Integration of Digital Literacy in Education

The integration of digital literacy into educational practices significantly influences teachers' professional development, instructional design, and collaborative engagement. Consistent with the teaching presence in the CoI framework, teachers with higher digital literacy are more capable of managing and facilitating interactive learning environments. Deiniatur et al. (2024) reported that English as a foreign language (EFL) teachers with strong behavioral and control beliefs regarding technology demonstrated greater creativity, communication, and critical thinking in their instructional practices. Similarly, Zhang et al. (2024) found that preservice teachers with strong digital communication and collaboration competencies exhibited higher professional confidence and pedagogical effectiveness. These results are consistent with the TPACK framework, in which the integration of technological and pedagogical knowledge enhances teachers' capacity to deliver meaningful and engaging instructions.

Despite its benefits, digital literacy integration remains uneven across educational contexts. Mhlongo et al. (2024) and Dabengwa et al. (2024) found that inadequate infrastructure, poor connectivity, and limited technical support continued to impede effective digital adoption in South African and Zimbabwean schools, particularly in rural areas. These challenges highlight the persistence of the digital divide, the need for systemic investment in ICT resources, and ongoing professional training.

Simultaneously, digital literacy contributes to social presence in the CoI model by promoting collaboration and shared learning among educators. Li and Zhang (2024) showed that embedding digital tasks in teacher education programs strengthened communication and creativity among preservice teachers. Zakariya et al. (2024) found that blended learning courses encouraged collaborative work and personalized learning through shared digital platforms despite technical constraints. Kurniawati et al. (2024) added that millennial English teachers effectively used digital tools in classroom instruction but required continuous training to sustain innovation. These studies affirm that digital literacy, when supported by institutional resources and policies, fosters innovation and cooperation and improves teaching outcomes.

Digital Literacy in Response to Educational Challenges

The second theme highlights the role of digital literacy as an adaptive mechanism in addressing systemic challenges in education, particularly during the COVID-19 pandemic and the subsequent shift toward remote and hybrid instruction (Alqahtani, 2024). Sohel et al. (2024) examined the experiences of Indigenous students in Bangladesh and reported that limited device ownership, poor Internet connectivity, and low technical literacy severely constrained online participation. These findings reflect broader issues of inequality within ODL environments, highlighting the importance of inclusive strategies that bridge socio-technical divides.

Institutional culture also influences the success of digital integration. Rasdiana et al. (2024) found that supportive and collaborative school environments enhanced teacher performance and digital literacy, whereas rigid institutional structures restricted innovation. Similarly, Molina-Torres (2024) observed that flipped learning methodologies, when implemented through digital platforms such as Moodle, improved digital competence and teaching quality. These findings align with the teaching presence dimension of CoI, emphasizing that pedagogical design and institutional support are central to effective online collaboration.

The socio-emotional dimension of digital literacy has also emerged as a significant factor that influences teaching resilience. Fathali et al. (2024) found that teachers with higher digital literacy experienced lower levels of anxiety in managing online instruction. Teachers who lacked sufficient technical proficiency reported stress and diminished confidence, which reflected the need for targeted professional development. Folabit and Jita (2024) similarly highlighted the evolving nature of teacher identity in digital contexts where educators are required to align their practices with students' technological expectations.

From a broader perspective, digital literacy enhances social and cognitive presence by fostering collaboration and reflective practices. Van Staden and Lotz-Sisitka (2024) found that e-learning tools in teacher education supported transformative collaboration, whereas Momdjian et al. (2024) revealed discrepancies between teachers' and administrators' perceptions of digital competence. Teachers demonstrated stronger practical application of digital tools, whereas institutional leaders required further training to promote systemic digital literacy. These results reaffirm the importance of the alignment between institutional vision, teacher capacity, and learner support in sustaining collaborative ODL ecosystems.

Advancing Learning Through Digital Competencies

The final theme illustrates how digital competencies advance pedagogical innovation, critical thinking, and interactive learning (Kiryakova & Kozhuharova, 2024). López-Meri et al. (2024) reported that digital literacy enhanced students' ability to detect misinformation and develop fact-checking skills, while Dharmayanti et al. (2024) found that students demonstrated proficiency in information evaluation and critical thinking but required improvement in creativity and collaboration. These studies support the cognitive presence dimension of CoI, in which digital competence facilitates inquiry, reflection, and knowledge construction.

Emerging technologies have transformed the nature of collaborative and experiential learning (Bruckhaus et al., 2024). Kee et al. (2024) and Nevrelova et al. (2024) showed that augmented and virtual reality (AR/VR) applications increased learner engagement, motivation, and teamwork. These findings align with the TPACK framework, which posits that integrating technological tools with sound pedagogical strategies can enhance authentic learning experiences. Similarly, Kurniawan (2024) demonstrated that collaborative learning models grounded in critical thinking improved students' writing and digital literacy skills, preparing them for future professional demands. Xu and Tan (2024) further emphasized the emerging potential of artificial intelligence tools, such as ChatGPT, in developing critical digital literacy among teachers of writing, although they noted challenges in integrating these tools effectively into curricula.

At the institutional level, Cabaron (2024) observed that digital training initiatives for maritime faculty members improved teaching confidence and instructional effectiveness, while Hasanah et al. (2024) found moderate digital literacy among Indonesian teachers with evident need for data and communication skills. Brante et al. (2024) identified similar competency gaps among younger learners who were adept at navigation but struggled to evaluate digital information sources. These findings underscore the necessity for structured digital literacy programs that address differentiated needs across learner groups and professional levels.

The growing use of artificial intelligence in education further illustrates opportunities and ethical concerns. Achruh et al. (2024) reported that while AI can enhance personalization and efficiency in Islamic education, it also raises the issues of digital inequality and data ethics. Bergstrom (2024) added that unfamiliarity with digital writing platforms may hinder effective collaboration and accessibility, reinforcing the importance of digital readiness and inclusive design. These studies highlight that while digital competencies are essential for advancing learning, their effective application requires a balanced approach that considers ethical, cultural, and contextual factors.

Implications and Future Research

The synthesis of findings highlights that advancing digital literacy within ODL requires coherent integration of policy reform, pedagogical innovation, and empirical inquiry. The interaction between the CoI and TPACK frameworks provides a strong theoretical foundation to guide these developments. Both frameworks emphasize that meaningful learning in distributed environments emerges when digital literacy enables educators to create cognitively engaging, socially connected, and pedagogically coherent content (Aslan et al., 2025). The implications derived from this analysis can inform institutional policy, educational practices, and future research on advancing digital literacy.

At the policy level, the review stresses the importance of national and institutional strategies that embed digital literacy within broader educational reform agendas. Governments and higher education institutions should ensure sustained investment in ICT infrastructure, particularly in rural and underserved regions, where digital inequities remain the most severe (Welesilassie & Gerencheal, 2025). Equitable access to digital tools and resources must be regarded as a central element of educational justice, ensuring that all teachers and learners can fully participate in ODL ecosystems. Policies should also require continuous professional development in digital literacy guided by UNESCO's DLGF and ICT-CFT. Additionally, incorporating ethical and socio-emotional competencies into digital education policies is essential to prepare educators for challenges associated with online safety, data privacy, and responsible use of emerging technologies (Jordan et al., 2025).

From a practical perspective, educational leaders and practitioners should develop capacity-building programs that strengthen both their technological proficiency and pedagogical adaptability. The TPACK framework can serve as a foundation for modular training programs that connect technical knowledge with effective teaching strategies and subject content (Aslan et al., 2025). Professional learning communities, peer mentoring, and collaborative teaching initiatives can nurture a culture of innovation and shared practice. Institutions should also promote reflective teaching practices that emphasize the cognitive, social, and teaching presence described in the CoI model (Anderson & Dron, 2011). Integrating digital literacy into curriculum design and assessment processes will further equip educators to foster critical thinking, creativity, and collaboration among learners (Niswah & Dewi, 2024; Suriani et al., 2024).

Inclusive pedagogical approaches must also account for learners' varied technological experiences and emotional well-being. Embedding socio-emotional dimensions in digital learning environments can reduce anxiety related to technological adaptation and strengthen learners' confidence in online collaborations. Such practices ensure that digital literacy development moves beyond functional competence to embrace the human and relational aspects of education (Pramusti et al., 2024).

These findings also point to several important directions for future research. Longitudinal studies are needed to examine how digital literacy training shapes teachers' professional identity, instructional design, and long-term engagement with ODL (Fazilla et al., 2022). Comparative studies across educational systems and cultural contexts can offer valuable insights into how institutional readiness and national policies influence digital literacy outcomes (Welesilassie & Gerencheal, 2025). Further research should explore the integration of AI, data analytics, and immersive technologies within the CoI and TPACK frameworks to understand how these tools transform cognitive and social presence in online learning (Yuliardi et al., 2024). Finally, studies using mixed-method designs that combine bibliometric, qualitative, and experimental approaches can provide stronger evidence of the theoretical connections linking digital literacy, collaboration, and learning effectiveness.

Conclusion

This systematic review has examined how digital literacy supports and enhances collaborative teaching across diverse educational contexts. Guided by the CoI and TPACK frameworks, this study addressed the persistent challenge of effectively integrating digital literacy into collaborative pedagogical practices. Using the PRISMA 2020 protocol, data were systematically retrieved from the Scopus and WoS databases, resulting in 32 peer-reviewed studies that met the inclusion criteria. These studies were analyzed thematically to identify key trends, theoretical intersections, and implications for educational policy and practice.

The review revealed three main findings. First, digital literacy strengthens teaching and social and cognitive presence by enabling educators to codesign instructional activities, engage in reflective dialogue, and sustain meaningful interactions in digital environments. Second, it functions as an adaptive mechanism that supports instructional resilience and pedagogical innovation during times of disruption such as the COVID-19 pandemic. However, disparities in access to technology, uneven skill development, and limited institutional support continue to constrain their potential. Third, developing advanced digital competencies, such as information evaluation, data literacy, and the ability to use collaborative tools, enhances innovation, critical thinking, and learner engagement. However, creativity and ethical practices remain areas requiring further attention.

This synthesis demonstrates that digital literacy should be viewed as a pedagogical capacity rather than a purely technical skill. Its transformative impact is most evident when technological integration aligns with the institutional culture, professional learning, and curriculum design. However, fragmented implementation and socio-technical inequities continue to impede sustainable progress in collaborative teaching.

At the policy level, the findings underscore the importance of sustained investment in equitable infrastructure, accessible digital platforms, and continuous professional development aligned with international frameworks such as UNESCO's DLGF and ICT-CFT. At institutional and pedagogical levels, digital literacy should be embedded within teacher training programs that emphasize the interrelationship between technological, pedagogical, and content knowledge. Establishing professional learning communities, peer mentoring initiatives, and reflective teaching practices can help to cultivate a collaborative and innovative digital learning culture.

Future research should employ longitudinal and comparative designs to examine how digital literacy development influences educators' professional growth, collaborative engagement, and student outcomes. Further inquiry into the integration of artificial intelligence, data analytics, and immersive technologies within CoI- and TPACK-informed frameworks will deepen our understanding of how these emerging tools shape teaching, social, and cognitive interactions in digital education.

In conclusion, digital literacy serves as the foundation for effective and equitable collaborative teaching. When supported by coherent policies, robust professional learning, and accessible technological infrastructure, it enables educators to design innovative, inclusive, and reflective learning environments. Therefore, strengthening digital literacy is both an educational necessity and a strategic pathway toward building sustainable, learner-centered, and future-ready education systems.

References

Achruh, A., Rapi, M., Rusdi, M., & Idris, R. (2024). Challenges and opportunities of artificial intelligence adoption in Islamic education in Indonesian higher education institutions. *International Journal of Learning, Teaching and Educational Research*, *23*(11), 423–443. https://doi.org/10.26803/ijlter.23.11.22

Adeleye, O., Eden, C., & Adeniyi, I. (2024). Educational technology and the digital divide: A conceptual framework for technical literacy inclusion. *International Journal of Science and Research Archive, 12*(1), 150–156. https://doi.org/10.30574/ijsra.2024.12.1.0405

Alqahtani, F. (2024). Nurse educators and faculty members' challenges towards online teaching during COVID-19 pandemic crisis: A qualitative descriptive study. *Nursing Open*, *11*(11), Article e70062. https://doi.org/10.1002/nop2.70062

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distributed Learning*, *12*(3), 80–97. https://doi.org/10.19173/irrodl.v12i3.890

Asagar, M. S. (2025). Digital competence in education: A comparative analysis of frameworks and conceptual foundations. *Synergy: International Journal of Multidisciplinary Studies*, *2*(1), 9–23. http://doi.org/10.2139/ssrn.5172207

Aslan, S. (2022). Using cooperative learning and the flipped classroom model with prospective teachers to increase digital literacy self-efficacy, technopedagogical education, and 21st-century skills competence. *International Journal of Progressive Education*, *18*(3), 121–137. https://doi.org/10.29329/ijpe.2022.439.9

Aslan, S., Alanoğlu, M., & Karabatak, S. (2025). Enhancing 21st-century teaching competencies: The key role of digital literacy in connecting pre-service teachers' TPACK. *Information Development*. Advance online publication. https://doi.org/10.1177/02666669251315841

Azzahro, F., Norra, B., & Achmad, C. (2023). The relationship of digital literacy ability with students' cognitive learning outcomes. *Jurnal BIOEDUIN: Program Studi Pendidikan Biologi*, *13*(1), 29–35. https://doi.org/10.15575/bioeduin.v13i1.24364

Barjestesh, H., Vijayaratnam, P., Sabzevari, M., Fatehi Rad, N., Rabani, K., & Manoochehrzadeh, M. (2025). Digital literacy of Iranian English as a foreign language (EFL) teachers: Teaching experience in focus. *Forum for Linguistic Studies*, *7*(1), 163–171. https://doi.org/10.30564/fls.v7i1.7244

Bergstrom, T. (2024). "I thought it was an accident": Digital literacy and MLL use of collaborative writing software in post-secondary composition classes. *Computer-Assisted Language Learning Electronic Journal*, *25*(1), 185–200. https://callej.org/index.php/journal/article/view/98

Blau, I., Shamir-Inbal, T., & Avdiel, O. (2020). How does the pedagogical design of a technology-enhanced collaborative academic course promote digital literacies, self-regulation, and perceived learning of students? *Internet and Higher Education*, *45*, Article 100722. https://doi.org/10.1016/j.iheduc.2019.100722

Brante, E. W., Walldén, R., & Ridell, K. (2024). Eight-year-olds engaging in guided information searches with iPads: Dimensions of reading competence. *International Journal of Child–Computer Interaction*, *40*, Article 100658. https://doi.org/10.1016/j.ijcci.2024.100658

Bruckhaus, A. A., Bennett, A., Brawer-Cohen, M., Sinclair, M., Ramirez-De La Cruz, G., Ragusa, G., & Duncan, D. (2024). Evaluation of students' digital literacy through an immersive university–high school collaboration. *Frontiers in Education*, *9*, Article 1429893. https://doi.org/10.3389/feduc.2024.1429893

Cabaron, R. R. (2024). Exploring the impact of digital literacy on the self-efficacy of maritime education faculty. *International Journal of Advanced and Applied Sciences*, *11*(7), 176–181. https://doi.org/10.21833/ijaas.2024.07.019

Çetinkaya, S. (2024). Field experience and digital technology-based literacy instruction. *International Journal of Educational Research Review*, *9*(2), 77–92. https://doi.org/10.24331/ijere.1404019

Choudhary, H. (2024). Building bridges to digital inclusion: Implications for curriculum development of digital literacy training programs. *International Journal of Technology Enhanced Learning*, *16*(3), 282–296. https://doi.org/10.1504/IJTEL.2024.139706

Cleveland-Innes, M. (2019). The community of inquiry theoretical framework: Designing collaborative online and blended learning. In H. Beetham & R. Sharpe (Eds.), *Rethinking pedagogy for a digital age* (3rd ed., pp. 85–102). Routledge. https://doi.org/10.4324/9781351252805

Dabengwa, I. M., Moyo, S., Ncube, S., Gashirai, T. B., Makaza, D., Makoni, P., Pasipamire, N., Chademana, G. K., Mafoti, M., Mapfumo, S., & Mandaza, D. (2024). Exploring digital competences in Zimbabwean secondary schools using a multimodal view: A hermeneutical phenomenography study. *Cogent Education*, *11*(1), Article 2387911. https://doi.org/10.1080/2331186X.2024.2387911

Darmaji, D., Astalini, A., Kurniawan, D., Nursulistiyo, E., Perdana, R., & Prameswari, N. (2023). The role of digital literacy and generic skills on student physics learning outcomes. *Jurnal Pendidikan Sains Indonesia*, *11*(4), 794–807. https://doi.org/10.24815/jpsi.v11i4.30560

Deiniatur, M., & Cahyono, B. Y. (2024). Digital literacy practices of novice English as a foreign language teacher in writing research articles for publication. *Journal of Education and Learning*, *18*(1), 165–172. https://doi.org/10.11591/edulearn.v18i1.20899

Deiniatur, M., Cahyono, B. Y., Ivone, F. M., & Prayogo, J. A. (2024). English teachers' beliefs and practices in integrating digital literacy in the language classroom. *International Journal of Evaluation and Research in Education*, *13*(2), 1242–1251. https://doi.org/10.11591/ijere.v13i2.25733

Dharmayanti, P. A. P., Padmadewi, N. N., Utami, I. G. A. L. P., & Suarcaya, P. (2024). Digital literacy competence for scientific writing: Students' perceptions and skills. *Journal of Language Teaching and Research*, *15*(5), 1550–1560. https://doi.org/10.17507/jltr.1505.16

Ervianti, E., Sampelolo, R., & Pratama, M. (2023). The influence of digital literacy on student learning. *Klasikal: Journal of Education, Language Teaching and Science*, *5*(2), 358–365. https://doi.org/10.52208/klasikal.v5i2.878

Fathali, S., Emadi, A., & Jebeli, S. L. (2024). Digital literacy and EFL teachers' anxiety with teaching online via virtual classroom software. *The JALT CALL Journal*, *20*(2), 1–21. https://doi.org/10.29140/jaltcall.v20n2.1154

Fazilla, S., Yus, A., & Muthmainnah, M. (2022). Digital literacy and TPACK's impact on preservice elementary teachers' ability to develop science learning tools. *Profesi Pendidikan Dasar*, *9*(1), 71–80. https://doi.org/10.23917/ppd.v7i1.9652

Folabit, N. L., & Jita, L. C. (2024). Are academics adapting to students' technology learning preferences? A South African study of teaching identities. *Issues in Educational Research*, *34*(4), 1241–1259. https://www.iier.org.au/iier34/folabit.pdf

Gabriel, F., Marrone, R., Van Sebille, Y., Kovanovic, V., & de Laat, M. (2022). Digital education strategies around the world: Practices and policies. *Irish Educational Studies*, *41*(1), 85–106. https://doi.org/10.1080/03323315.2021.2022513

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. https://doi.org/10.1080/08923640109527071

Haryaka, U., Razak, N. K., Rachman, F., Tung, K. Y., & Judijanto, L. (2025). Integrating digital literacy, critical thinking, and collaborative learning: Addressing contemporary challenges in 21st-century education. *Journal of Hunan University Natural Sciences*, *52*(3), 101–112. https://doi.org/10.55463/issn.1674-2974.52.3.9

Hasanah, U., Santoso, A., Pratiwi, Y., & Farid, M. (2024). Exploring the correlation of self-perception on the use of digital literacy in learning. *International Journal of Evaluation and Research in Education*, *13*(6), 4354–4364. https://doi.org/10.11591/ijere.v13i6.27114

Huachara-Martinez, E., Erazo-Moreno, M. M., Paz-Checa, D. P., Chomba-Sung, S. D. R., & Nina-Cuchillo, J. (2023). Digital competencies in collaborative learning of students in a public university in Lima. *Journal of Higher Education Theory and Practice*, *23*(15), 121–129. https://doi.org/10.33423/jhetp.v23i15.6413

Irwandi, I. (2023). Enhancing digital literacy through English language teaching: A needs analysis for English students at Muhammadiyah University of Mataram. *Jurnal Ilmiah Mandala Education*, *9*(3), 2055–2060. https://doi.org/10.58258/jime.v9i3.5782

Jony, M., & Sultana, S. (2023). Parents' perception towards online education. *International Journal of Educational Innovation and Research*, *2*(2), 151–166. https://doi.org/10.31949/ijeir.v2i2.5039

Jordan, A., Julianto, A., & Firmansyah, M. A. (2025). Integrating digital literacy into curriculum design: A framework for 21st-century learning. *Journal of Technology, Education & Teaching (J-TECH)*, *1*(2), 79–85. https://doi.org/10.62734/jtech.v1i2.418

Jung, J., Choi, S., & Fanguy, M. (2024). Exploring teachers' digital literacy experiences. *International Review of Research in Open and Distributed Learning*, *25*(2), 41–59. https://doi.org/10.19173/irrodl.v25i2.7572

Karroum, S., & Elshaiekh, N. (2023). Digital transformation in education: Discovering the barriers that prevent teachers from adopting emerging technologies. In *Proceedings of the 2023 24th International Arab Conference on Information Technology (ACIT)* (pp. 166–170). IEEE. https://doi.org/10.1109/ACIT58888.2023.10453908

Kee, T., Zhang, H., & King, R. B. (2024). An empirical study on immersive technology in synchronous hybrid learning in design education. *International Journal of Technology and Design Education*, *34*(3), 1243–1273. https://doi.org/10.1007/s10798-023-09855-5

Kiryakova, G., & Kozhuharova, D. (2024). The digital competences necessary for the successful pedagogical practice of teachers in the digital age. *Education Sciences, 14*(5), Article 507. https://doi.org/10.3390/educsci14050507

Kumalasari, D., Purwanta, H., & Aw, S. (2024). Comparative analysis of Generation Z's digital history literacy in history education majors on Java Island: A study of history digital literacy. *Journal of Education and E-Learning Research*, *11*(1), 90–96. https://doi.org/10.20448/jeelr.v11i1.5342

Kurniawan, K. (2024). Enhancing 21st-century writing skills through a reflective collaborative learning model based on critical thinking. *Pakistan Journal of Life and Social Sciences*, *22*(2), 12393–12400. https://doi.org/10.57239/PJLSS-2024-22.2.00886

Kurniawati, N., Mayuni, I., & Lustyantie, N. (2024). A case study of millennial English teachers' awareness of digital media in EFL classrooms. *Studies in Media and Communication*, *12*(1), 286–297. https://doi.org/10.11114/smc.v12i1.6670

Kwiatkowska, W., & Wiśniewska-Nogaj, L. (2022). Digital skills and online collaborative learning: The study report. *Electronic Journal of E-Learning*, *20*(5), 510–522. https://doi.org/10.34190/eje1.20.5.2412

Law, N., Woo, D., & Wong, G. (2018). *A global framework of reference on digital literacy skills for indicator 4.4.2* (UNESCO Institute for Statistics [UIS] information paper no. 51). UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000265403

Lazareva, A., & Tømte, C. E. (2024). Flexible learning spaces as an arena for developing professional digital competence through small group collaboration. *Teachers and Teaching*, *30*(4), 489–508. https://doi.org/10.1080/13540602.2024.2308893

Li, M., & Zhang, M. (2024). Embedding digital literacies in the language teacher education curriculum: Pre-service and in-service teachers' perspectives. *CALICO Journal*, *41*(3), 273–296. https://doi.org/10.1558/cj.25987

López-Meri, A., Doménech-Fabregat, H., & Marcos-García, S. (2024). Digital competencies in verifying fake news: Assessing the knowledge and abilities of journalism students. *Societies*, *14*(5), Article 66. https://doi.org/10.3390/soc14050066

Lora, J., López-Bravo, C., & Mosquera-Pérez, C. (2021). Virtual learning tools in higher education: EDIEDPAT, an interactive platform for urban heritage teaching. In *EDULEARN21 Proceedings* (pp. 8104–8112). IATED Digital Library. https://doi.org/10.21125/edulearn.2021.1642

Männistö, M., Mikkonen, K., Kuivila, H. M., Koskinen, C., Koivula, M., Sjögren, T., & Kääriäinen, M. (2020). Health and social care educators' competence in digital collaborative learning: A cross-sectional survey. *SAGE Open*, *10*(4). https://doi.org/10.1177/2158244020962780

Mattar, J., Santos, C. C., & Cuque, L. M. (2022). Analysis and comparison of international digital competence frameworks for education. *Education Sciences*, *12*(12), Article 932. https://doi.org/10.3390/educsci12120932

Mhlongo, S., Ndlovu, M., & Chirinda, B. (2024). Mathematics teachers' experiences of using online teaching resources for professional learning in a context of disadvantage. *International Journal of Learning, Teaching and Educational Research*, *23*(10), 27–44. https://doi.org/10.26803/ijlter.23.10.2

Micsky, T., & Foels, L. (2019). Community of inquiry (CoI): A framework for social work distance educators. *Journal of Teaching in Social Work*, *39*(4–5), 293–307. https://doi.org/10.1080/08841233.2019.1642976

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

Molina-Torres, M. P. (2024). Flipped classroom to teach digital skills during COVID-19. *Journal of Technology and Science Education*, *14*(1), 158–168. https://doi.org/10.3926/jotse.2256

Momdjian, L., Manegre, M., & Gutiérrez-Colón, M. (2024). A comparison of perceptions of digital competences of schoolteachers to school leaders in Lebanon. *Social Sciences and Humanities Open*, *10*, Article 100937. https://doi.org/10.1016/j.ssaho.2024.100937

Nash, B. (2024). Critical inquiry in (and about) media environments: Examining an asset-based digital literacy curriculum. *Journal of Literacy Research*, *56*(2), 133–156. https://doi.org/10.1177/1086296X241244700

Nevrelova, N., Korenova, L., Lavicza, Z., Bruzkova, N., & Schmid, A. (2024). Enhancing digital literacy in primary education through augmented reality. *Frontiers in Education*, *9*, Article 1390491. https://doi.org/10.3389/feduc.2024.1390491

Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, *59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

Niswah, P. U., & Dewi, N. R. (2024). Development of e-worksheet with the TPACK approach to train students' digital literacy and higher order thinking skills. *Jurnal Penelitian Pendidikan IPA*, *10*(6), 3473–3485. https://doi.org/10.29303/jppipa.v10i6.4850

Nuhu, K., Abdulfatai, D., & Onojah, A. (2021). Undergraduate awareness and perception on the use of digital collaborative tools in facilitating learning in selected universities within the Ilorin metropolis. *Indonesian Journal of Educational Research and Technology*, *1*(3), 95–104. https://doi.org/10.17509/ijert.v1i3.38735

Ohle-Peters, A., Ludewig, U., & McElvany, N. (2024). Can we foster pre-service teachers' competences for digital collaboration? *Frontiers in Education*, *9*, Article 1455074. https://doi.org/10.3389/feduc.2024.1455074

Oliiar, M., & Fomin, K. (2022). Blended learning as a means of future primary school teacher professional development. *Mountain School of Ukrainian Carpaty*, (26), 90–94. https://doi.org/10.15330/msuc.2022.26.90-94

Papanikolaou, K., Gouli, E., & Makri, K. (2014). Designing pre-service teacher training based on a combination of TPACK and communities of inquiry. *Procedia—Social and Behavioral Sciences*, *116*, 3437–3442. https://doi.org/10.1016/j.sbspro.2014.01.779

Patwardhan, V., Rao, S., & Prabhu, N. (2020). Community of Inquiry (CoI) framework and course design as predictors of satisfaction in emergency remote teaching: Perspectives of hospitality management students. *Journal of E-Learning and Knowledge Society*, *16*(4), 94–103. https://doi.org/10.20368/1971-8829/1135315

Perifanou, M., Economides, A., & Tzafilkou, K. (2021). Teachers' digital skills readiness during COVID-19 pandemic. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(8), 238–251. https://doi.org/10.3991/ijet.v16i08.21011

Pramusti, R. A., Ulum, B., & Kumalasani, M. P. (2024). Digital literacy transformation of learners with special needs through TPACK approach in IPAS learning in elementary school. *Journal of Learning Improvement and Lesson Study*, *4*(2), 58–68. https://doi.org/10.24036/jlils.v4i2.122

Purwandari, E. P., Junus, K., & Santoso, H. B. (2022). Exploring e-learning community of inquiry framework for engineering education. *International Journal of Instruction*, *15*(1), 619–632. https://e-iji.net/ats/index.php/pub/article/view/465

Qiu, Y., Li, L., & Wang, X. (2024). Quantitative analysis digital literacy of secondary and higher vocational school students in the digital economy background: Preliminary empirical research based on 181 samples from Zhejiang, China. *INTI Journal*, *2024*(5). https://doi.org/10.61453/intij.202405

Rahimi, R. A., & Oh, G. S. (2024). Rethinking the role of educators in the 21st century: Navigating globalization, technology, and pandemics. *Journal of Marketing Analytics*, *12*(2), 182–197. https://doi.org/10.1057/s41270-024-00303-4

Ramos, J. L., Cattaneo, A. A., de Jong, F. P., & Espadeiro, R. G. (2022). Pedagogical models for the facilitation of teacher professional development via video-supported collaborative learning. A review of the state of the art. *Journal of Research on Technology in Education*, *54*(5), 695–718. https://doi.org/10.1080/15391523.2021.1911720

Rasdiana, R., Mauludin, I., Yahya, A., Putri, D. E., Machrus, M. A., Marbun, M., Sholikhah, A., Sinusi, N. S., Fathonah, S., Salmayda, S., Pawartani, T., & Ridwan, A. (2024). Mediation of digital literacy in investigating the effect of school culture on teacher performance: Implication for educational policy. *Journal of Infrastructure, Policy and Development*, *8*(12), Article 9117. https://doi.org/10.24294/jipd.v8i12.9117

Raza, T., & Akhter, N. (2024). Harnessing ICT resources: Empowering educators and enhancing student skills through teacher training programs. *Journal of Asian Development Studies*, *13*(4), 673–688. https://doi.org/10.62345/jads.2024.13.4.55

Richardson, J. C., Arbaugh, J. B., Cleveland-Innes, M., Ice, P., Swan, K. P., & Garrison, D. R. (2012). Using the Community of Inquiry framework to inform effective instructional design. In L. Moller & J. B. Huett (Eds.), *The next generation of distance education: Unconstrained learning* (pp. 97–125). Springer. https://doi.org/10.1007/978-1-4614-1785-9_7

Rodríguez-García, A. M., Cardoso-Pulido, M. J., De la Cruz-Campos, J. C., & Martínez-Heredia, N. (2022). Communicating and collaborating with others through digital competence: A self-perception study based on teacher trainees' gender. *Education Sciences*, *12*(8), Article 534. https://doi.org/10.3390/educsci12080534

Sánchez-Canut, S., Usart-Rodríguez, M., Grimalt-Álvaro, C., Martínez-Requejo, S., & Lores-Gómez, B. (2023). Professional digital competence: Definition, frameworks, measurement, and gender differences: A systematic literature review. *Human Behavior and Emerging Technologies*, *2023*(1), Article 8897227. https://doi.org/10.1155/2023/8897227

Sohel, M. S., Sifullah, M. K., Hossain, B., Hossain, M. A., Sarker, M. F. H., Hossain, I., Hossain, M. E., & Uddin, M. R. (2024). E-learning experience of Indigenous rural communities in the face of COVID-19 crisis in Chittagong Hills Tracts Region, Bangladesh: A qualitative investigation. *International Journal of Community Well-Being*, *7*(2), 291–314. https://doi.org/10.1007/s42413-024-00207-2

Suriani, A., Sukma, E., Chandra, C., & Dasa Putri, L. (2024). TPACK-integrated learning media assisted by Appypie to improve the literacy skills of elementary school teacher education (ESTE) students. *Jurnal Inovasi Pendidikan dan Pembelajaran Sekolah Dasar*, *8*(1). https://doi.org/10.24036/jippsd.v8i1.128990

Swan, K. (2019). Social construction of knowledge and the community of inquiry framework. In I. Jung (Ed.), *Open and distance education theory revisited: Implications for the digital era* (pp. 57–65). Springer Singapore. https://doi.org/10.1007/978-981-13-7740-2_7

Tangkish, N., Tumashbay, T., Shaldarbekova, A., Rakhimshikova, M., Aripzhan, G., & Yerbota, A. (2024). Digital mind and human consciousness: Integration of digital technology in shaping learning experiences. *Perspektivy Nauki i Obrazovania*, *69*(3), 58–75. https://doi.org/10.32744/pse.2024.3.4

Tinmaz, H., Fanea-Ivanovici, M., & Baber, H. (2023). A snapshot of digital literacy. *Library Hi Tech News*, *40*(1), 20–23. https://doi.org/10.1108/LHTN-12-2021-0095

van Staden, W., & Lotz-Sisitka, H. (2024). E-learning as a mediating tool to support interactive professional learning of teacher educators. *Interactive Learning Environments*, *32*(7), 3226–3239. https://doi.org/10.1080/10494820.2023.2170423

Welesilassie, M. W., & Gerencheal, B. (2025). Digital language instruction in Ethiopian high schools: Digital literacy, utilization, and challenges. *European Journal of Interactive Multimedia and Education*, *6*(1), Article e02501. https://doi.org/10.30935/ejimed/15722

Wertz, R. E. (2022). Learning presence within the Community of Inquiry framework: An alternative measurement survey for a four-factor model. *The Internet and Higher Education*, *52*, Article 100832. https://doi.org/10.1016/j.iheduc.2021.100832

Xu, W., & Tan, X. (2024). Beyond words: L2 writing teachers' visual conceptualizations of ChatGPT in teaching and learning. *Journal of Second Language Writing*, *64*, Article 101110. https://doi.org/10.1016/j.jslw.2024.101110

Yuliardi, R., Kusumah, Y. S., Nurjanah, N., Juandi, D., & Said Husain, S. K. (2024). The relationship between information literacy, learning space and TPACK on prospective teachers' digital mathematics literacy. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, *9*(3), 144–163. https://doi.org/10.23917/jramathedu.v9i3.5361

Zakariya, Y. F., Danlami, K. B., & Shogbesan, Y. O. (2024). Affordances and constraints of a blended learning course: Experience of pre-service teachers in an African context. *Humanities and Social Sciences Communications*, *11*(1), Article 1596. https://doi.org/10.1057/s41599-024-04136-5

Zhang, S., Gu, M. M., Sun, W., & Jin, T. (2024). Digital literacy competence, digital literacy practices and teacher identity among pre-service teachers. *Journal of Education for Teaching*, *50*(3), 464–478. https://doi.org/10.1080/02607476.2023.2283426

Zhao, Y., Llorente, A. M. P., & Gómez, M. C. S. (2021). Digital competence in higher education research: A systematic literature review. *Computers & Education*, *168*, Article 104212. https://doi.org/10.1016/j.compedu.2021.104212

Zorko, T. (2024). Digital transformation through cooperative learning. *Švietimas: Politika, Vadyba, Kokybė/Education Policy, Management and Quality*, *16*(1), 69–79. https://doi.org/10.48127/spvk-epmq/24.16.69

Zou, Y., Kuek, F., Feng, W., & Cheng, X. (2025). Digital learning in the 21st century: Trends, challenges, and innovations in technology integration. *Frontiers in Education*, *10*, Article 1562391. https://doi.org/10.3389/feduc.2025.1562391

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

MOOCs Reshaping Undergraduate Health Education: A Systematic Review

Melike Aydemir Arslan[1*], Asiye Ata[1], and Sevda Kucuk[1]
[1]Ataturk University, Turkey; [*]Corresponding Author

Abstract

Given the growing demand for flexible and accessible health education, massive open online courses (MOOCs) have been recognized as instrumental in expanding undergraduate learning. This systematic review was conducted to investigate the use of MOOCs in undergraduate health education, focusing on publication trends, geographic distribution, and key research variables. A total of 31 peer-reviewed articles were reviewed, and data were sourced from six international databases: Web of Science, Scopus, ERIC, EBSCOHost, ScienceDirect, and PubMed. It was found that MOOCs have been integrated into undergraduate health education since 2014, with a notable increase in publications observed after 2022. The highest number of studies was published in China. Student satisfaction was identified as the most frequently studied variable, and medical education was reported as the dominant field. Quantitative research were predominantly used, with sample sizes between 101 and 300 participants. Questionnaires were commonly employed as a data collection tool, and many studies were based on custom-developed MOOCs for their research. Courses were typically between 4 and 6 weeks duration. Improved clinical skills were frequently reported as outcomes, while the lack of practical experience in MOOC-based learning was identified as a major limitation. More practice-oriented teaching approaches were recommended by most studies. To enhance the effectiveness of MOOCs in health education, more innovative and practical implementation strategies are needed. Future research is encouraged to address these gaps and strengthen the impact of MOOCs on undergraduate health programs. The growing role of MOOCs in health education is highlighted, particularly the need to integrate practical components for greater educational impact.

Keywords: MOOCs, undergraduate health education, medical education, systematic review, educational evaluation, clinical skills

Introduction

In recent years, digitization and opportunities created by information and communication technologies have brought about transformative changes for learners and teachers. With the widespread use of e-learning, systems for both synchronous and asynchronous education have been developed; text materials have been supplemented by audio, animation, and video. Massive open online courses (MOOCs), whose popularity has increased rapidly in recent years, are now an important tool and environment for higher education (Zhu et al. 2018).

For more than two decades, MOOCs have played a key role in the rise of education supported by information and communication technologies, and have been considered a revolutionary development in the field. Open access and scalability are the main principles of MOOCs. Open access refers to courses that are open to anyone who wants to learn and generally refers to being free online. Scalability refers to the fact that courses are designed to support a large number of participants and respond to individual applications (Yuan & Powell, 2013). MOOCs have three main features: (a) they are generally free to use, (b) they provide high-quality education on a variety of subjects, and (c) they are pioneers in lifelong learning. Students have opportunities to learn without any prerequisites or time and space limitations. Thus, it can be said that MOOCs are one of the biggest supporters of lifelong learning (Goh et al., 2017).

MOOCs are widely used in many fields, such as computer science, teaching, engineering, politics, economics, and communication sciences, as well as health studies, especially medical education. In health education, MOOCs have been most common in nursing, pharmacy, and medicine. They have been used to support medical training and interprofessional learning, especially during the COVID-19 pandemic (Dedeilia et al., 2020).

The use of MOOCs in medical education has been seen as an important step towards improving health services (Pouladi et al., 2023). Including perspectives from developing countries can enhance the global relevance of MOOCs, as they have offered scalable and cost-effective solutions in resource-limited settings. However, contextual barriers such as language, technology access, and cultural differences should also be considered (Barteit et al., 2019). MOOCs offer exciting opportunities for learners on a global scale. Carefully selected and collaboratively designed MOOCs can provide nearly unlimited access to undergraduate and postgraduate health education research, and studies examining MOOCs have helped fill knowledge gaps by employing various research models (Setia et al., 2019). The literature has highlighted the effectiveness of e-learning in health education, enhancing knowledge, skills, and satisfaction. Systematic reviews have explored its applications in postgraduate medical education (de Leeuw et al., 2019; Hopcan et al., 2024), low- and middle-income countries (Barteit et al., 2020), and among nurses and medical students (Delungahawatta et al., 2022; Lahti et al., 2014). Common findings have emphasized the importance of well-designed interventions tailored to learners' needs (Naciri et al., 2021; Regmi & Jones, 2020).

However, despite these advancements, important research gaps remain, particularly in the context of undergraduate health education. Most existing studies focused on postgraduate levels (Nieder et al., 2022; Rowe et al., 2019). As a result, the role of MOOCs in supporting practical, hands-on training at the undergraduate level remains underexplored. This is a significant limitation, as clinical and applied experiences are essential components of health education. However, current MOOC implementations often lack strong frameworks for simulating or supporting real-world practice. These include areas such as patient interaction, procedural training, and interprofessional collaboration. Recent evidence as shown that MOOCs can enhance medical knowledge and clinical skills. In some cases, they even outperform traditional teaching methods (Gao et al., 2021; Yang et al., 2023). Nevertheless, further research is needed to evaluate their effectiveness in delivering comprehensive, practice-oriented education at the undergraduate level.

Despite the growing popularity of MOOCs, their use and efficacy in undergraduate health education remain underexplored. This study aimed to address this gap by conducting a comprehensive systematic review focused on MOOCs in undergraduate health education. It analyzed publication trends, methodologies, key variables, and MOOC characteristics. By bridging this gap, the study provided valuable insights to guide future research and practical applications, shaping the role of MOOCs in meeting the needs of educators and students. The following research questions guided this study.

1. What were the publication trends of articles on using MOOCs in undergraduate health education?
    - publication years
    - country
    - variables
    - research topic
    - research methods
    - sample
    - data collection tools

2. What MOOC platforms have been used in undergraduate health education?

3. How have MOOCs used materials, assessment types, and implementation periods in undergraduate health education?

4. What main implications have been derived from the articles on using MOOCs in undergraduate health education?
    - strengths
    - weaknesses
    - suggestions

Methodology

The purpose of this systematic review was to examine the role of MOOCs in undergraduate health education. The concept of MOOCs in this study was based on definitions widely accepted in the relevant literature. While some studies involved small rather than massive sample groups, this was mostly due to methodological limitations, and these courses still fulfilled the core principles of MOOCs—open access, online delivery, and scalability. Our systematic review was an explicitly and systematically conducted literature review that answered specific questions through the application of a replicable search strategy, with studies included or excluded based on explicit criteria (Gough et al., 2017). Following the identification of potential studies, we applied the PRISMA guidelines (Moher et al., 2015) to ensure a consistent and transparent review process.

Search Strategy and Selection Procedure

This systematic review targeted peer-reviewed articles in English, focusing on the use of MOOCs in undergraduate health education, and indexed in six international databases: Web of Science, Scopus, ERIC, EBSCOHost, ScienceDirect, and PubMed. The search was conducted without any time limitations. The search was undertaken in April 2024. The search strings were (“medical education” OR “health education” OR “nurse education”) AND (“MOOC” OR “massive open online course”). Table 1 summarizes the search strategy for each database.

Table 1

Search Strategies According to Database

Database	Search strategy
WOS	TOPIC=[(“medical education” OR “health education” OR “nurse education”) AND ((“MOOC”) OR “massive open online course”)]
ERIC	(“medical education” OR “health education” OR “nurse education”) AND (“MOOC” OR “massive open online course”)
SCOPUS	TITLE-ABS-KEY [(“health education” OR “medical education” OR “nurse education”) AND (“MOOC” OR “massive open online course”)]
Science Direct	Title, abstract, keywords: (“medical education” OR “health education” OR “nurse education”) AND (“MOOC” OR “massive open online course”)
EbscoHost	KW TI AB (medical education or health education or nurse education) AND KW TI AB (mooc or massive open online courses)
PubMed	(“medical education”[Title/Abstract] OR “health education”[Title/Abstract] OR “nurse education”[Title/Abstract]) AND (“MOOC”[Title/Abstract] OR “massive open online course”[Title/Abstract])

In the initial search, we identified 611 records. After removing 314 duplicates, we used Web-based review software to facilitate data extraction and analysis. The final selection process involved a thorough screening to ensure that studies specifically addressing the applications of MOOCs in undergraduate health education were included (Table 2).

Table 2

Inclusion and Exclusion Criteria

Inclusion criteria	Exclusion criteria
Studies published up to April 2024	-
Journal articles	Book, book chapters, conference papers, dissertations, editorials
Application of MOOC in health education	Not about MOOC in health education
English language	Not English language
Undergraduate education	Graduate education, professional education, public education
	Full text inaccessible

Quality Assessment

The selection of articles to be included in the review was based on a two-dimensional evaluation approach (Vissenberg et al., 2022). The first dimension, methodological trustworthiness, involved assessing the studies according to criteria such as design, context, sampling, data collection, data analysis, and claims and evidence. The second dimension, relevance to the review, focused on evaluating the extent to which each study aligned with the subject matter of the review (Vissenberg et al., 2022). Each study was rated as low, medium, or high based on these criteria (Table 3).

Table 3

Checklist Quality Assessment

Quality dimension	Criteria	Considerations
Methodological trustworthiness	Design	Is the study design appropriate and rigorous?
	Context	Is the study context clearly described?
	Sampling	Is the sampling strategy clearly appropriate?
	Data collection	Are data collection methods clearly appropriate?
	Data analysis	Is the data analysis clearly appropriate?
	Claims/Evidence	Are claims supported by sufficient evidence?
Relevance to review		Does the study directly address key concepts relevant to the review's objectives?

Table 4 presents the quality assessment results for the selected studies. A structured rating system was employed based on the two key dimensions of methodological trustworthiness and relevance to the review. Each study was independently evaluated against these dimensions and rated for quality as high (i.e., met both criteria at a high level), medium (i.e., demonstrated moderate alignment) or low (i.e., clear methodological or relevance-related limitations). This classification enabled a more transparent and consistent comparison of the strengths and applicability of the evidence presented. Details are given in Table 4.

Table 4

Quality Assessment on the Selected Studies

Article title	Methodological trustworthiness	Relevance to the review
Delivering a medical school elective with massive open online course (MOOC)	High	High
Comparison of the influence of massive open online courses and traditional teaching methods in medical education in China: A meta-analysis	High	High
Uncovering motivation and self-regulated learning skills in integrated medical MOOC learning: A mixed methods research protocol	Medium	Medium

Article title	Methodological trustworthiness	Relevance to the review
The Integration of an anatomy massive open online course (MOOC) into a medical anatomy curriculum	High	High
Use of H5P interactive learning content in a self-paced MOOC for learning activity preferences and acceptance in an Indonesian medical elective module	Medium	Medium
Assessing the effectiveness of massive open online courses on improving clinical skills in medical education in China: A meta-analysis	High	High
Creating a 'choose your topic' massive open online course: An innovative and flexible approach to delivering injury prevention education	Low	Low
Application of online teaching mode combining case studies and the MOOC platform in obstetrics and gynecology probation teaching	Medium	Medium
Designing a cassive open online Course (MOOC) in face-to-face sessions. A blended design to teach practical histology	High	High
The distance teaching practice of combined mode of massive open online course micro-video for interns in emergency department during the COVID-19 epidemic period	High	High
Science education regarding oral health: General health for nonmedical undergraduates applying a SPOC teaching model	Medium	Medium
The efficacy of the "Talk-to-Me" suicide prevention and mental health education program for tertiary students: A crossover randomised control trial	High	High
Understanding the utility of 'Talk-to-Me' an online suicide prevention program	Medium	Medium
Effect on exam performance through massive open online courses versus face-to-face classrooms among Indian medical students: An analytic study	High	High
Application of a new multi-element integrated teaching mode based on bite-sized teaching, flipped classroom, and MOOC in clinical teaching of obstetrics and gynaecology	High	High
Improving undergraduate education of occupational health and occupational medicine applying massive open online courses & problem-based learning	Medium	Medium
The 'Talk-to-Me' MOOC intervention for suicide prevention and mental health education among tertiary students: Protocol of a multi-site cross-over randomised controlled trial	High	High
Teaching effects of the online and offline flipped classroom model (FCM) in the post-epidemic era: Development and feasibility study	Medium	Medium

Article title	Methodological trustworthiness	Relevance to the review
Preventing chronic pain: A human systems approach—Results from a massive open online course	Medium	Medium
Health care professionals from developing countries report educational benefits after an online diabetes course	High	High
Perception and use of massive open online courses among medical students in a developing country: Multicentre cross-sectional study	Medium	Medium
A massive open online course for teaching physiotherapy students and physiotherapists about spinal cord injuries	High	High
MOOCs, an innovative alternative to teach first aid and emergency treatment: A practical study	Medium	Medium
Implications of a new form of online education	Low	Low
The comparison of teaching efficiency between massive open online courses and traditional courses in medicine education: A systematic review and meta-analysis	Medium	Medium
A test of the first course (Emergency Medicine) that is globally available for credit and for free	High	High
Behavior analysis and formative assessments in online oral medicine education during the COVID-19 pandemic	Medium	Medium
Considering the use of massive open online courses (MOOCs) in medical education	Medium	Medium
Cyberincivility in the massive open online course learning environment: Data-mining study	Low	Low
Delivering an online course in emergency nursing education during the pandemic: What are the effects on students' learning?	High	High
MOOCs and medical education: Hope or hype?	Medium	Medium

Screening and Inter-Rater Reliability

The screening process involved a team of two coders to evaluate 611 titles and abstracts. During this initial screening stage, sensitivity was prioritized over specificity, meaning that papers were more likely to be included than excluded. To reach a consensus, the reasons for inclusion and exclusion were discussed for the first 40 articles at regular meetings. Subsequently, articles were allocated randomly and reviewed based on predefined inclusion and exclusion criteria.

A third researcher randomly reviewed 50 of the articles to ensure consistency and accuracy in coding decisions. Inter-rater reliability was determined using Cohen's kappa (κ; Cohen, 1960), which measures the degree of consistency among raters. Kappa values are characterized as fair (.40 to .60), good (.60 to .75), and excellent (over .75; Bakeman & Gottman, 1997). In our study, the inter-rater reliability was calculated as $\kappa = 1$, indicating excellent agreement among the coders.

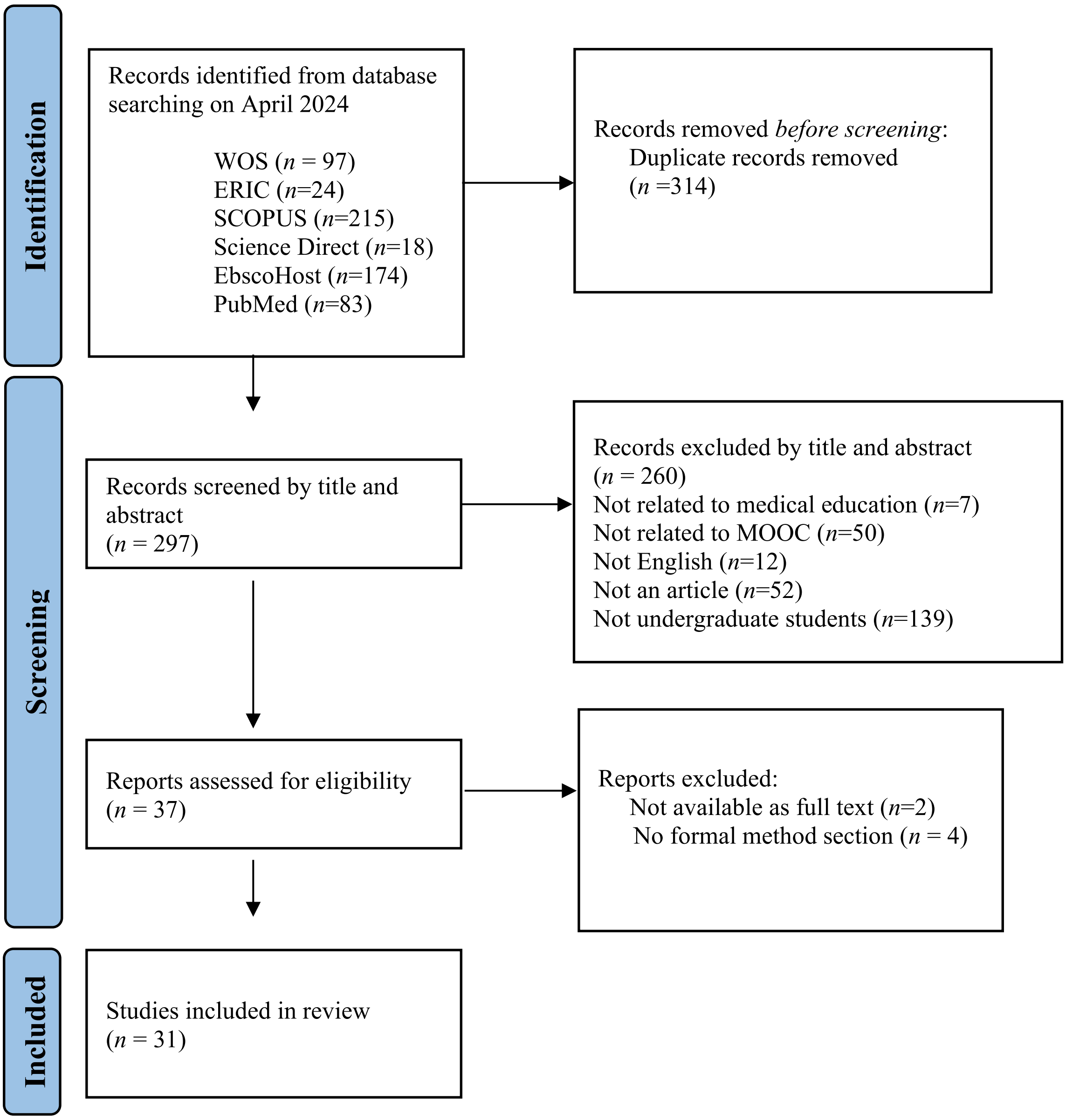
After the initial screening, 351 articles were excluded because they were not (a) related to health education, (b) focused on MOOCs, (c) written in English, (d) journal articles, or (e) centered on undergraduate students. This process left us with 37 reports for full-text assessment. Of these, six reports were excluded because two could not be accessed in full text, and four lacked a formal method section. Consequently, 31 studies were included in the final review for synthesis (Figure 1).

ARRefMan software was used to systematically analyse each study. We created a detailed form designed to capture various aspects of the studies based on our research questions. For the method section analysis, we used the Educational Technology Paper Classification Form developed by Kucuk et al. (2013). This classification helped categorize the studies based on their methodological approaches, allowing for more structured and systematic analysis.

Our systematic review on the use of MOOCs in undergraduate health education analyzed publication trends, including publication year, country of origin, and key variables investigated. We examined the topics covered and research methodologies, along with sample characteristics, sample sizes, and data collection tools. We also identified MOOC platforms and targeted courses. Learning materials and assessment systems were documented. Key implications highlighted the strengths, weaknesses, and suggestions for future research and practice. To identify the strengths and weaknesses of the studies, two researchers independently performed thematic coding by reviewing relevant textual content. Statements were then categorized and grouped through consensus, and recurring themes were extracted to ensure consistency and transparency.

Figure 1

PRISMA Flow Chart



Findings on Using MOOCs in Undergraduate Health Education

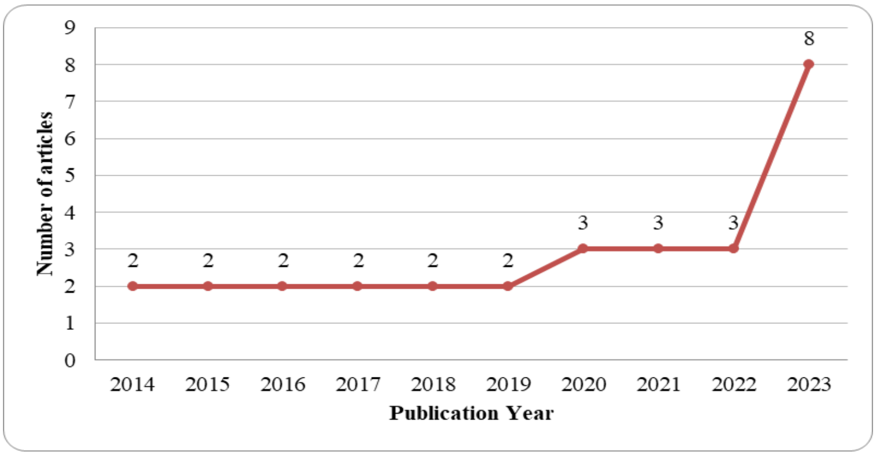
The research findings were analyzed under seven headings based on the research questions. The findings obtained for each question are presented below.

Publication Trends of Articles

We analyzed the distribution of year of publication for the articles on using MOOCs in undergraduate health education. Since there were no publications before 2014 and the scope of our study was limited to until April 2024, the year 2024 was not included. The distribution of publications between 2014 and 2023 is shown in Figure 2.

Figure 2

Distribution by Year of Publication

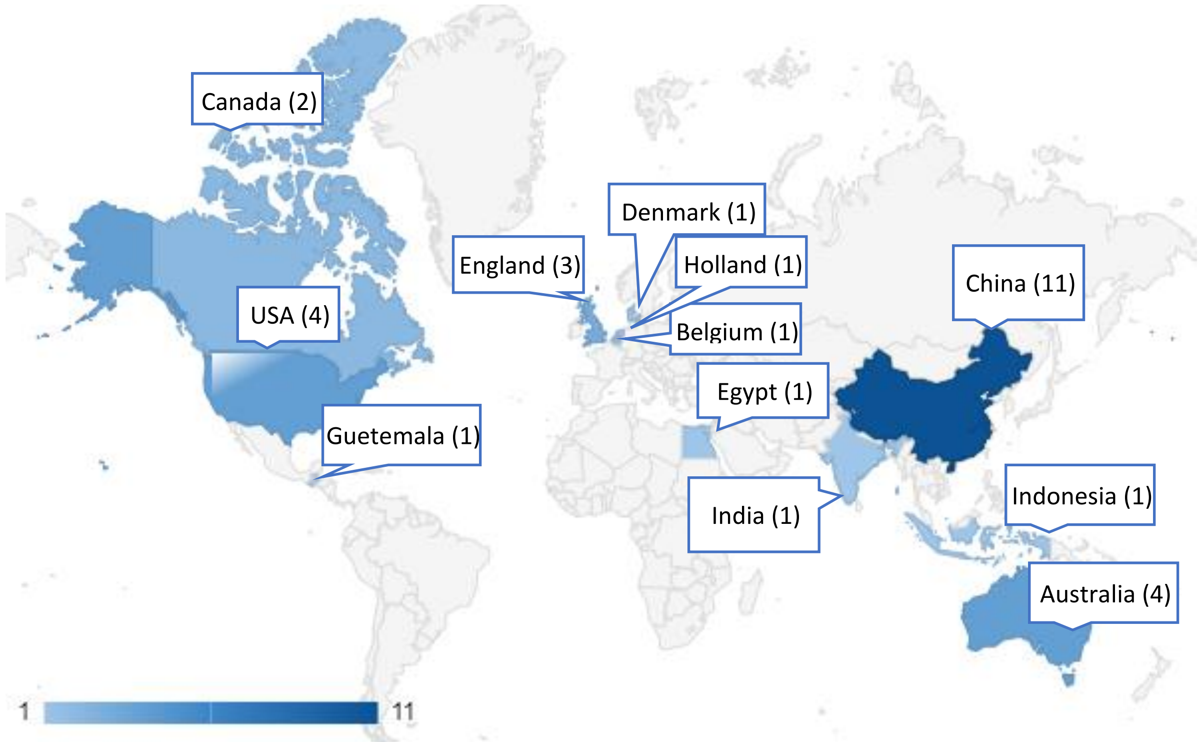


When the distribution of publications by year was examined, it was found that MOOCs have been used in undergraduate health education since 2014. There were two publications between 2014 and 2019. After 2022, a significant increase in the number of publications was observed, with the highest number of publications in 2023 (n = 8). As of April 2024, two publications were identified.

Figure 3 presents a heat map that illustrates the distribution of publications examining the use of MOOCs in undergraduate health education by country. The countries were determined based on the country of the corresponding author. In the studies on using MOOCs in health education, China (n = 11) had the highest number of publications.

Figure 3

Distribution of Publications by Country



The research topics of the studies were identified and are presented in Table 5. A total of 19 different research topics were identified. The research topics were categorized based on the primary focus areas explicitly stated in each study, such as clinical skills, oral health, or public health, under the broader scope of undergraduate medical education. Medical education in general (n = 5) was the most common research topic, followed by emergency nursing/nursing (n = 3), and suicide prevention and mental health education (n = 3).

Table 5

Research Topic

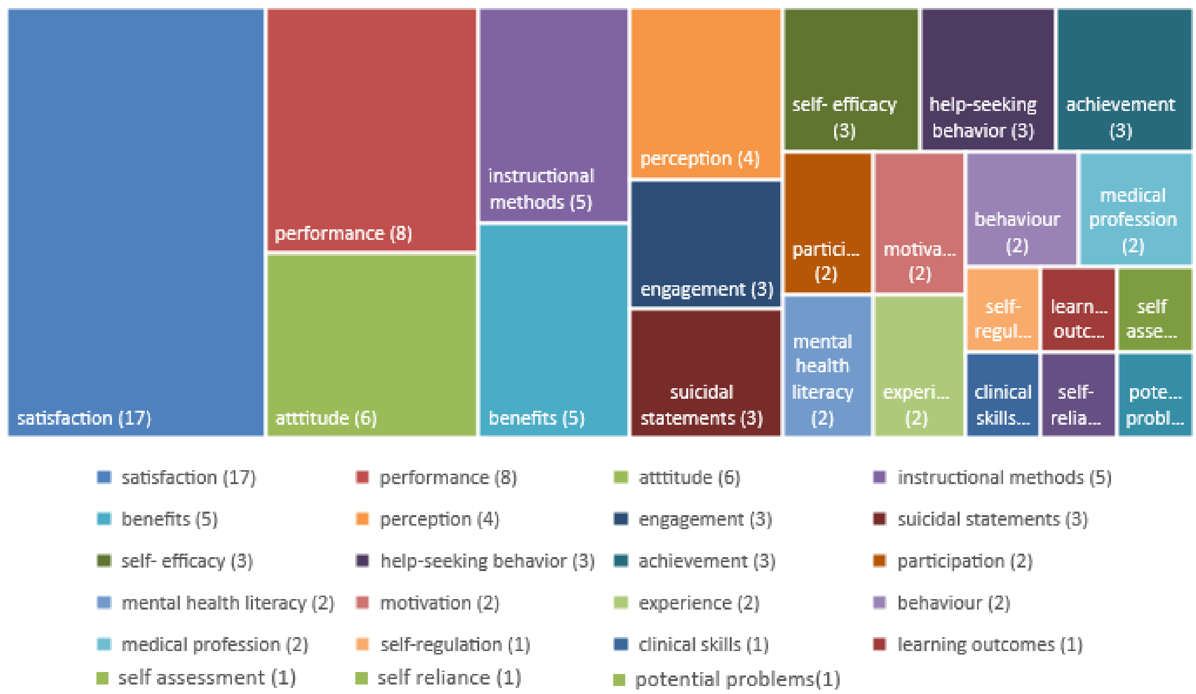
Research topic	n
Medical education	5
Emergency nursing/nursing	3
Suicide prevention and mental health education	3
General internal medicine	2
Anatomy	2
Emergency medicine	2
Obstetrics and gynecology	2
Oral health, dentistry	2
Herbal medicine	1

Research topic	*n*
Injury prevention	1
Practical histology	1
Business of medicine	1
Chronic pain	1
Cyber incivility	1
General health	1
Medical genetics education	1
Occupational health/ medicine	1
Diabetes	1
Physiotherapy	1

The variables used in publications on using MOOCs in undergraduate health education were examined, and their distribution is presented in Figure 4. We found that 23 different variables were used in the studies. The most analyzed variable was satisfaction (n = 17), followed by performance (n = 8).

Figure 4

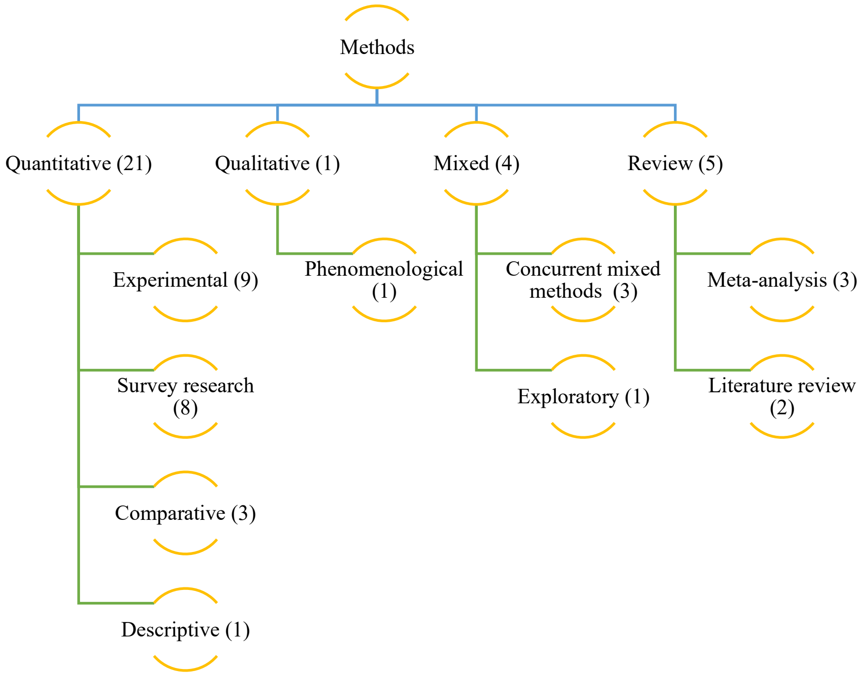
Distribution of Variables Used in Articles



The research methods used in the studies were identified; their distribution is shown in Figure 5. Quantitative methods (n = 21) were the most commonly used in studies on using MOOCs in health education, followed by reviews (n = 5) and mixed methods (n = 4).

Figure 5

Research Methods



The sample populations and sample sizes of the studies were analyzed and are presented in the cross table (Table 6). Due to the focus of our study, samples from the medical field were studied most often (*n* = 23), followed by samples from the non-medical field (*n* = 6), nursing (*n* = 4), and other health fields (*n* = 2). Most studies were conducted with a sample size between 101 and 300 (*n* = 12).

Table 6

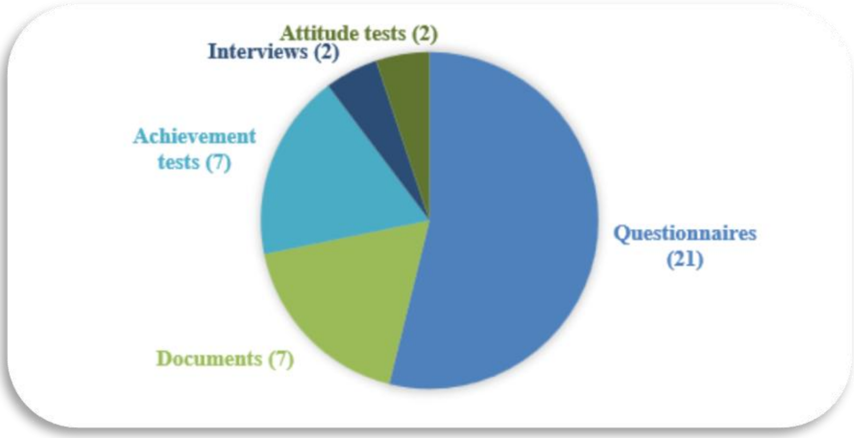
Distribution of Sample Sizes by Type of Sample Population

	Sample size							
Population	1–10	11–30	31–100	101–300	301–1,000	1,000+	Unspecified	Total
Medical	1	1	5	6	6	3	1	23
Non-medical			1	4		1		6
Nursing			3				1	4
Other health				2				2
Biomedical					1			1
Dentistry					1			1
Physiotherapy						1		1
Total	1	1	9	12	8	5	2	

The data collection tools used in the analyzed studies were examined, and the results are presented in Figure 6. Questionnaires (*n* = 21) were used the most frequently, followed by documents (*n* = 7) and achievement tests (n = 7). Interviews (n = 2) and attitude, perception, personality, and aptitude tests (n = 2) were the least used.

Figure 6

Data Collection Tools



MOOC Platform Used in Research

We examined the use of MOOC to explore how differences in platform design and functionality (e.g., user interface, content delivery methods, learner support features) influenced the pedagogical effectiveness and accessibility of MOOCs in undergraduate health education. MOOC platforms used in the field of health were analyzed (see Table 7). In most studies, MOOCs specifically developed for the study (n = 8) were used. Following this, ready-made platforms were used, though the specific platform was not mentioned in some studies (n = 6). Among the ready-made platforms, EdX (n = 4) was the most frequently used.

Table 7

MOOC Platforms Used in Health Education

MOOC platform	*n*
Specific	8
Existing but not written	6
EdX	4
Coursera	3
FutureLearn	3
Udemy	1
TalkToMe	1
Udacity	1
China University MOOC	1
NextGenU’s	1
MOOC Medical Neuroscience	1
Not given	3

How MOOCs are Used in Health Education

The materials used in MOOCs in health education, assessment types, and MOOC implementation periods were analyzed and are presented in Figure 7. The most common assessment types were completed graded course assessments (n = 11) and online quizzes (n = 11), followed by final course score (n = 8), certification (n = 6), objective structured video examinations (n = 5), and online tasks (n = 4). Theoretical assessment (n = 2), practical assessment (n = 2), mentor assessment (n = 2), and evaluation and assessment systems (n = 2) were used less frequently. Clinical decision-making (n = 1), peer assessment (n = 1), and self-assessment (n = 1) were the assessment types used least.

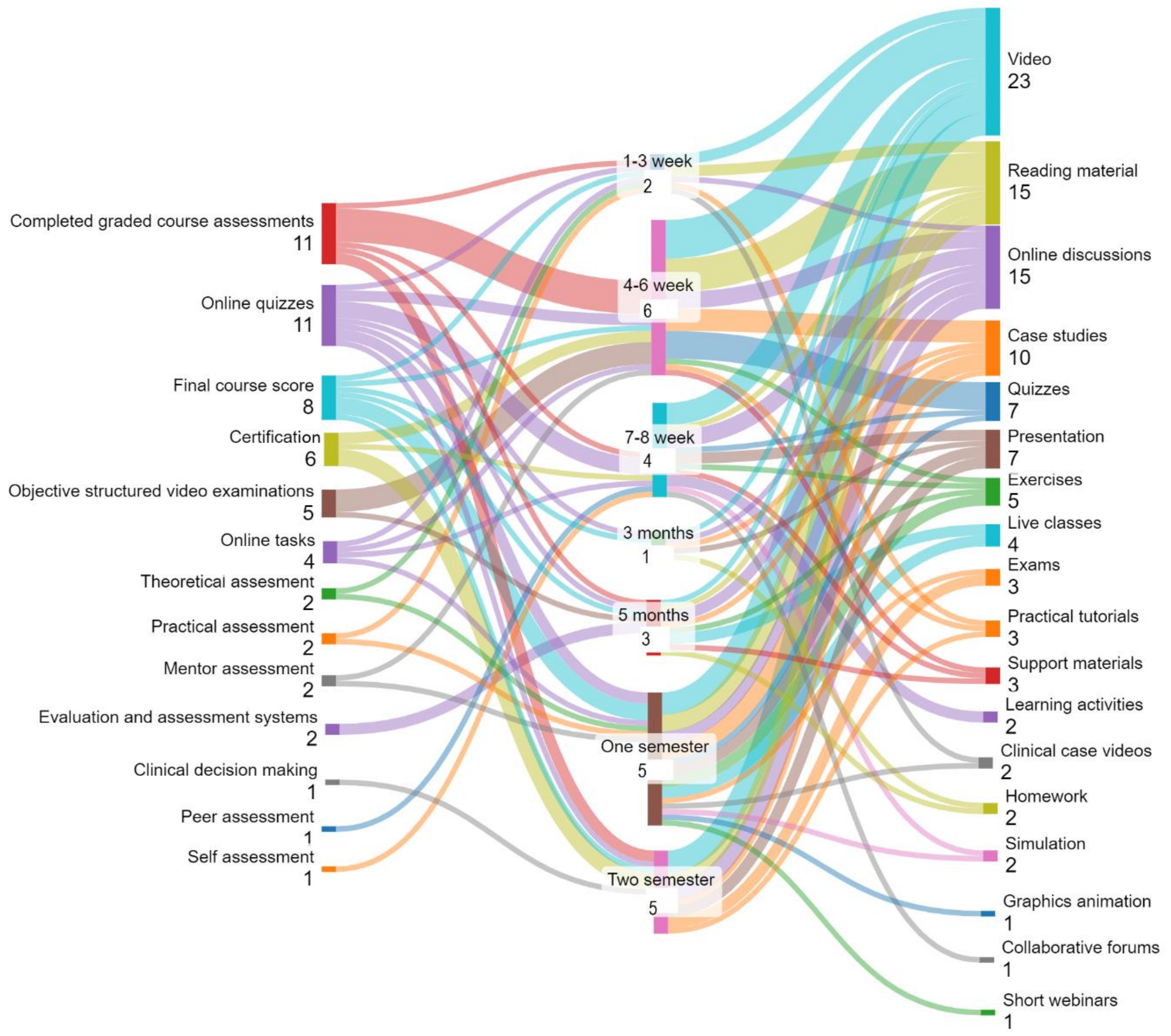
When the materials used in MOOCs for undergraduate health education were examined, it was found that video (n = 23) was the most preferred. This was followed by reading materials (n = 15), online discussions (n = 15), case studies (n = 10), quizzes (n = 7), presentations (n = 7), exercises (n = 5), live classes (n = 4), exams (n = 3), practical tutorials (n = 3), support materials (n = 3), learning activities (n = 2), clinical case videos (n = 2), homework (n = 2), and simulations (n = 2). The methods used least were graphic animations (n = 1), collaborative forums (n = 1), and short webinars (n = 1).

The most common duration for undergraduate health education MOOCs was between four and six weeks (n = 6), followed by studies conducted over one semester (n = 5) and two semesters (n = 5). Less frequent were studies with a duration of seven to eight weeks (n = 4), five months (n = 3), and one to three weeks (n = 2), respectively. Just one study examined a course of over three months duration (n = 1).

In the one to three week range, video (n = 2) and reading material (n = 2) were the course materials used most often, with different assessment types being used. In the four to six week range, completed graded course assessments (n = 6) and objective structured video examinations (n = 4) were the most common assessment types, while video (n = 7), reading material (n = 6), and quizzes (n = 5) were the materials used most. For courses in the seven to eight week range, online quizzes (n = 3) were the most common assessment type, while video (n = 4) and online discussions (n = 3) were the materials used most. In the three-month range, a range of different assessment types and materials were used equally. In the five-month range, evaluation and assessment systems (n = 2) were the assessments used most, with live classes (n = 2) and online discussions (n = 2) the materials used most. For one semester courses, final course scores (n = 3) and video (n = 4) were used as assessment types and materials, respectively. In the two-semester range, certification (n = 3) and video materials (n = 4) were the most commonly used.

Figure 7

Distribution of the Use of MOOCs in Health Education

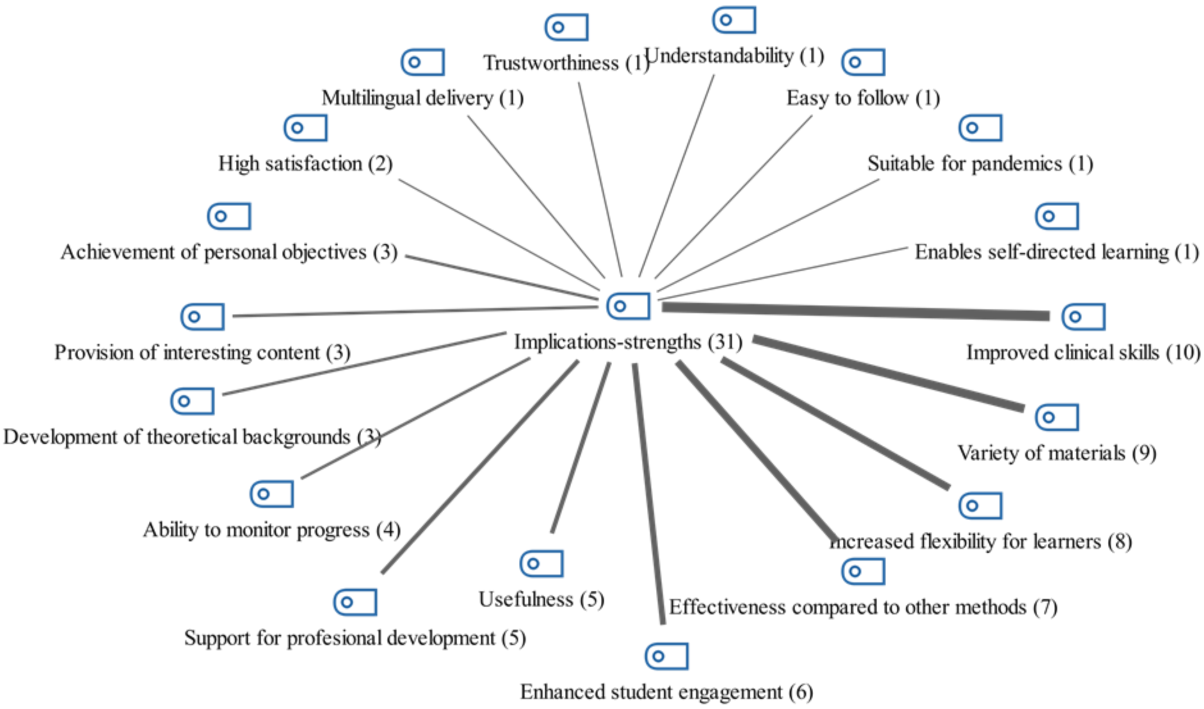


Main Implications for Using MOOCs in Medical Education

The strengths identified in the use of MOOCs for undergraduate health education were analyzed, and a general distribution was created. The distribution of these strengths is shown in Figure 8. The most common result was improved clinical skills (n = 10). Other positive aspects included variety of materials (n = 9), increased flexibility for learners (n = 8), effectiveness compared to other methods (n = 7), and enhanced student engagement (n = 6). Additional positive results are outlined in Figure 8.

Figure 8

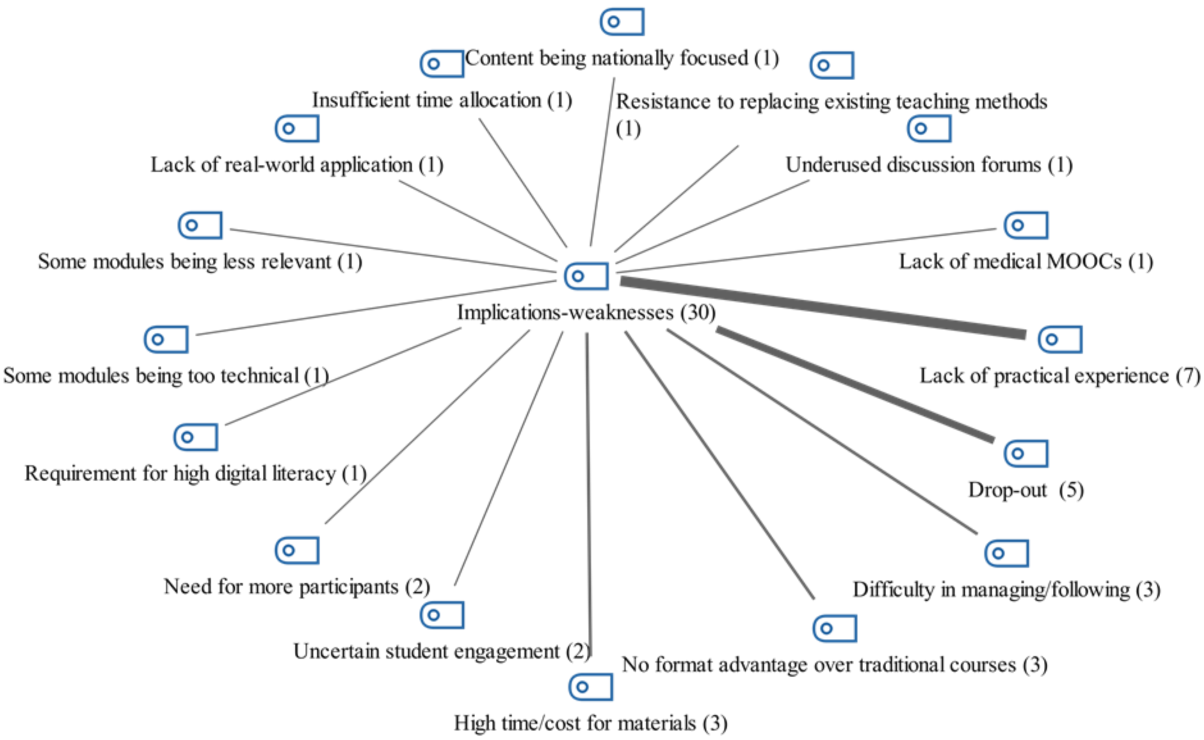
Strengths of Using MOOCs in Health Education



The weaknesses identified in the use of MOOCs for undergraduate health education were also analyzed, and a general distribution was created. The distribution of these weaknesses is shown in Figure 9. The most common weakness was a lack of practical experience (n = 7). Other noted weaknesses included participants' drop-out (n = 5), difficulty in managing/following (n = 3), no format advantage over traditional courses (n = 3), high time/cost for materials (n = 3), uncertain student engagement (n = 2), and need for more participants (n = 2). Additional weaknesses are outlined in Figure 9.

Figure 9

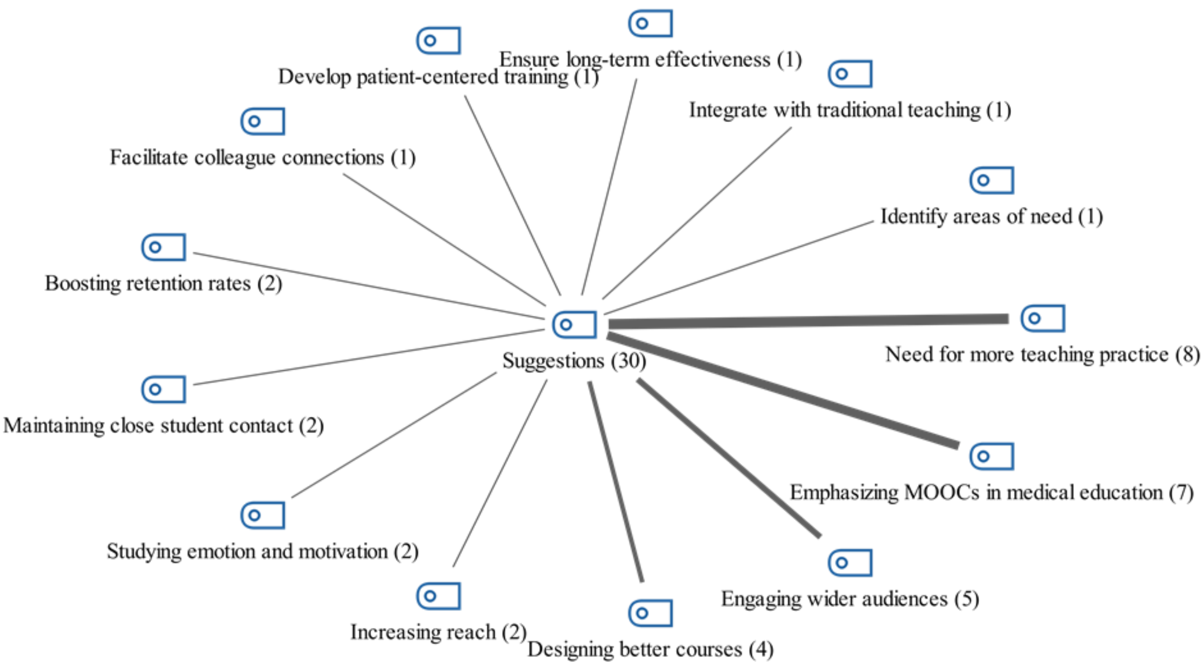
Weaknesses of Using MOOCs in Health Education



The suggestions in the studies on using MOOCs in health education were analyzed, and a general distribution was created (see Figure 10). The most common suggestion was the need for more teaching practice (n = 8). Other suggestions included emphasizing MOOCs in medical education (n = 7), engaging wider audiences (n = 5), designing better courses (n = 4), increasing reach (n = 2), studying emotion and motivation (n = 2), maintaining close student contact (n = 2), and boosting retention rates (n = 2). Additional suggestions are outlined in Figure 10.

Figure 10

Suggestions for Using MOOCs in Health Education



Discussion

Publication Trends of Articles on Using MOOCs in Undergraduate Health Education

MOOCs have become a significant tool in medical education, offering a flexible and accessible platform for undergraduate health students, health professionals, and the general public (Longhini et al., 2021; Nieder et al., 2022; Setia et al., 2019). In this study, the place of MOOCs in undergraduate health education has been presented in the context of research methodology, the platform used to host MOOCs and how it is used, and the research results obtained. MOOCs have been used in various fields since 2012, and in undergraduate health education since 2014 (Clark et al., 2017; Liyanagunawardena & Williams, 2014. They have become increasingly popular in online learning in recent years, especially after 2022. The popularity of MOOCs in the medical education field can be explained by the fact that it is a suitable method to support and increase the medical education and interprofessional learning required during COVID-19 (Bettiol et al., 2022; Dedeilia et al., 2020). The upward trend in publications, particularly the sharp increase in 2023, reflected the growing academic interest in leveraging MOOCs for undergraduate health education in the post-pandemic era (Nieder et al., 2022).

While the USA and UK were pioneers in medical MOOCs (Nieder et al., 2022; Rowe et al., 2019), China has emerged as a leader in this field. The first MOOC platform managed by universities in China started in 2013 (Guo, 2014), and medical MOOCs in China have been deeply involved in higher education and seen as important resources (Gong, 2018; Yang et al., 2023). In addition, due to the large population of China, it may be that China has preferred MOOCs for mass education at the undergraduate health education level. The dominance of studies from developed regions highlighted a need for broader geographic inclusion, as contextual differences may significantly influence the effectiveness and design of MOOCs.

When we examined studies of undergraduate health education MOOCs, we determined that satisfaction was the variable studied most often (Gao et al., 2021; Nieder et al., 2022; Rowe et al., 2019; Yang et al., 2023). This trend suggested that researchers prioritized learners' subjective experiences as a key outcome in evaluating the effectiveness of MOOCs. However, the frequent focus on satisfaction also pointed to a potential gap in assessing deeper learning outcomes such as knowledge retention and skill acquisition, which future studies could explore.

Performance is another key variable; research showed that MOOCs can significantly improve academic performance. A study involving health sciences undergraduates found that those who participated in a MOOC showed better academic performance compared to those who did not (Martín-Valero et al., 2021). Achievement, closely related to performance, has often been measured through course completion and grades. The ability to predict student success in MOOCs and for-credit courses using online interactions has been demonstrated, suggesting that similar factors influence achievement in both contexts (Almeda et al., 2018). Still, it remains unclear whether these performance indicators align with actual skill development in clinical contexts, which has long been a critical consideration in undergraduate health education. Attitude towards MOOCs can affect participation and completion rates, and we found that didactical interventions positively changed learners' attitudes toward online learning (Khalil & Ebner, 2018). This suggested that pedagogical strategies play a significant role in shaping engagement outcomes. Instructional methods in MOOCs are evolving, with a focus on integrating digital technologies to enhance learning. The post-pandemic shift in health education has highlighted the potential of MOOCs to deliver synchronous and asynchronous learning effectively (Almeda et al., 2018). The benefits of MOOCs, including accessibility and the ability to reach a wide audience, have been demonstrated by the diverse range of health education MOOCs available during the pandemic (Dolores-Maldonado et al., 2021).

The studies we examined focused on a wide range of research topics. Overall, the number of review studies was high. Most of these addressed topics related to medical education, while nursing was the second most studied field. This finding is similar to other studies (Gao et al., 2021; Rowe et al., 2019; Yang et al., 2023).

When the research methods in the studies were analyzed, it was found that mostly quantitative studies were preferred, and among these, experimental studies were the most common research design. Other studies have observed that descriptive/explorative studies (Longhini et al., 2021) and narrative/opinion designs (Rowe et al., 2019) were preferred in research related to medical MOOCs. Survey comparative and descriptive studies were prominent in collecting data from large numbers of students. It was also observed that the number of review studies has increased. This diversity of research methods indicates that different aspects of the use of medical MOOCs continue to be explored. The rapid spread of MOOCs requires continuous use of diverse research methods in order to evaluate their effectiveness and impact on health education outcomes (Bettiol et al., 2022).

In both medical and non-health fields, sample sizes in the range of 101 to 300 are preferred, so studies in the undergraduate health education fields have generally been limited to medium-sized groups. Jordan (2014) stated that despite the high enrolment rates in MOOCs, many students did not complete the course, so sample sizes can be further reduced for effective analyses. More comprehensive analyses may be required to evaluate the effectiveness of MOOCs with a large participant population (Kizilcec & Schneider, 2015). Most of the included studies recruited participants from the medical-related programs, suggesting that MOOCs are more commonly used in medical education than in other health disciplines. The predominance of medical-related samples supports previous studies indicating that the use of MOOCs in medical education has become widespread (Liyanagunawardena & Williams, 2014). The trend toward smaller sample sizes may have reflected feasibility concerns and dropout-related limitations rather than methodological preference, suggesting the need for more diverse and scalable designs in future MOOC research.

Questionnaires were the most commonly used data collection tool in the studies we examined, and this may have been expected because they facilitate the receipt of feedback from participants in online education environments that appeal to large masses, such as MOOCs. Questionnaires enable easy and fast data collection from a large number of participants. In the literature, it was emphasized that questionnaires were frequently used in MOOC research and were ideal for measuring participants' experiences, satisfaction, and perceptions of learning processes (Longhini et al., 2021; Nieder et al., 2022).

MOOC Platforms Used in Undergraduate Health Education Research

Most of the MOOC platforms used in the field of health were specially developed. Educational needs in the field of health can be very specific, and therefore, general-purpose MOOC platforms may not always be sufficient. Health education requires applied knowledge as well as interdisciplinary skills, so specially designed platforms may be more effective. Among the ready-made MOOCs, EdX, Coursera, and FutureLearn were more preferred. This finding is similar to the studies in the literature (Kononowicz et al., 2015; Longhini et al., 2021)

The analysis of MOOCs in undergraduate health education revealed a variety of materials, assessment types, and implementation times, reflecting the adaptability and potential of MOOCs to enhance learning in this field. MOOCs have become an important tool in health education by offering a wide range of learning resources and assessment methods tailored to different learner needs and contexts (Setia et al., 2019). Various types of assessment have been used in health education MOOCs, with completed graded course assessments and online quizzes the most common. It may be that completed graded course assessments were preferred because they are seen as a criterion in determining the success of MOOCs (Nieder et al., 2022). The variety of assessment types shows that MOOC designers intentionally selected methods that align with the structure and learning objectives of the course (Villarroel et al., 2018).

MOOCs in undergraduate health education have used a variety of materials to facilitate learning. These include lecture videos, interactive discussions, and case studies, which are crucial for engaging learners and enhancing their understanding of complex topics (Alturkistani et al., 2018; Nwameme et al., 2023). Videos have been the most preferred material (Longhini et al., 2021), consistent with the importance of audio-visual learning in online courses (Guo, 2014). High-quality course content and the sharing of educational resources online have been emphasized to improve the learning experience and ensure the accessibility of materials to a broad audience (Cui, 2022).

For MOOCs in undergraduate health education, courses lasting four to six weeks were the most common, which suggested that a medium-duration course was preferred in the field of health. MOOCs are generally planned for short-term and intensive knowledge transfer; flexible and modular approaches are at the forefront of online education (Hew & Cheung, 2014). However, courses spread over one or two semesters were also quite common. Longer courses may be preferred when more comprehensive and in-depth topics need to be addressed. While these findings provide flexibility in online learning, they point to the necessity of longer programs in areas that require intensive knowledge, such as health education (Hauer et al., 2023).

The evaluation types and materials used in MOOCs for health education varied depending on the course duration. While videos and reading materials were generally preferred in short-term MOOCs, more comprehensive evaluation methods and materials were used in long-term courses (Hew & Cheung, 2014). While short-term courses aimed for intensive knowledge transfer, long-term courses allowed students to learn the subject in depth and develop their higher-order skills with a wider variety of materials and assessment methods. This highlighted the flexibility of online education and showed that MOOCs in health education have been structured according to different learning needs (Schettino & Capone, 2022).

The Main Implications for Using MOOCs in Undergraduate Health Education

The use of MOOCs in undergraduate health education offered several strengths, with the most prominent being the improvement of clinical skills. A meta-analysis conducted in China demonstrated that students participating in MOOC-based courses exhibited improved clinical skill scores (Yang et al., 2023). Other notable benefits have included the provision of diverse materials, increased flexibility for learners, and enhanced student engagement. MOOCs offer a wide range of materials that cater to different learning needs and preferences, which is crucial for health education where diverse content is necessary (Eglseer, 2023). The flexibility of MOOCs allows learners to access content at their own pace and convenience, making it easier for them to balance their studies with other commitments (Eglseer, 2023; Guerrero-Quiñonez et al., 2023; Iniesto et al., 2016). This flexibility is particularly beneficial during unprecedented times, such as the COVID-19 pandemic, where traditional learning methods were disrupted (Rulinawaty et al., 2023). We found that these strengths were distributed across various aspects of learning, contributing to the overall effectiveness of MOOCs in health education.

Analyses of the use of MOOCs in undergraduate health education have revealed several weaknesses. The most frequently cited problem was that online platforms may be insufficient for developing practical skills in health education (Regmi & Jones, 2020). In addition, declining participation revealed that it is difficult to ensure long-term student motivation (Jordan, 2015). Structural weaknesses, such as the difficulty of management and follow-up, as well as the lack of format advantage compared to traditional courses, indicated that MOOCs were inadequate in providing individual feedback while appealing to large audiences. MOOCs cannot fully meet the interaction and practice opportunities offered by face-to-face education (Margaryan et al., 2015). In addition, challenges such as high digital literacy requirements made it difficult for some students to use these platforms effectively due to differences in learners' digital skills (Kizilcec et al., 2013).

Suggestions for using MOOCs in undergraduate health education have focused on improving both the theoretical and practical aspects of the platforms. The suggestion to provide more teaching practice emphasized the need to support theoretical knowledge with practical applications, and reinforced that practical skills are critical due to the nature of health education (Regmi & Jones, 2020). This reflected a common criticism (Olivares Olivares et al., 2021) that MOOCs may be particularly lacking in terms of developing clinical skills. At the same time, suggestions for greater use of MOOCs in medical education and their appeal to wider audiences imply that the accessibility and impact of these platforms can be increased. Research has shown that MOOCs have the potential to reach more students, especially because they are low-cost and widely accessible (Hollands & Tirthali, 2014; Laurillard & Kennedy, 2017; Liyanagunawardena et al., 2013). To enhance the applicability of these recommendations, integrating hands-on components such as virtual simulations, scenario-based tasks, and peer feedback into MOOC content is essential (Cook et al., 2010; Kononowicz et al., 2015). Additionally, small-scale pilot programs in collaboration with medical institutions could help evaluate the effectiveness and refine practical elements before wider implementation.

Besides this access, course designs need to be improved, which is an important requirement for students to experience more interaction and in-depth learning (Hew & Cheung, 2014). In addition, examining motivational and affective factors may be the key to increasing students' levels of commitment and engagement in the online learning process (Garrison, 2016).

Conclusion and Suggestions

We systematically analyzed 31 MOOC studies on undergraduate health education from six databases. However, the study had certain limitations that should be acknowledged. First, only English-language publications were included, which may have led to the exclusion of relevant studies published in other languages. Second, the analysis was limited to a specific set of six databases, which may not fully represent all relevant literature in the field. These constraints should be considered when interpreting the findings, as they may affect the generalizability of the results particularly in non-English-speaking or underrepresented regions.

Several implications were revealed. It has been observed that there has been a remarkable increase in the number of studies on undergraduate health education in recent years, especially after the COVID-19 pandemic. Researchers can turn to work in this field. It has been observed that developed countries in the USA, Australia, and Europe, and especially China in Asia, have been active in this field, but there are very few studies in developing countries. Cooperative studies can be carried out with developing countries under the leadership of developed countries. Although many variables were investigated in these studies, the number of studies in which variables specific to medical education were examined was rare. In order to contribute more to the literature on medical education, variables studies less often, yet specific to medical education, could be focused on. We determined that there were more topics on basic medical sciences and general health education and fewer topics on practical skills. MOOCs could incorporate cutting-edge technologies, such as virtual reality, augmented reality, artificial intelligence, and so on, with MOOCs designed for topics that include practical skills.

There were few mixed-method studies in this field. Researchers can contribute to the development of a more comprehensive and in-depth understanding of undergraduate health education by conducting mixed-method studies. In MOOCs for undergraduate health education, data can be collected from a large number of participants and analyzed with machine learning techniques, which allows for individualization of learning processes and in-depth insights to improve educational success. Universities can contribute to a more widespread impact by opening their own MOOCs to other universities and sharing them on other recognized platforms.

In order to support individualized learning in MOOCs, material diversity, course duration, and assessment types may vary according to needs. Increasing the variety of materials, keeping course durations flexible, and using different types of assessment (e.g., self-assessment, peer assessment, practice tests) to support individualized learning in undergraduate health MOOCs can significantly improve student motivation and learning efficiency. Implications based on the literature can be taken into account in future studies. To increase the effectiveness of MOOCs in undergraduate health education, an innovative and holistic approach should be adopted in terms of both content and implementation. Future research could explore the integration of immersive technologies and conduct longitudinal or case-based studies with detailed statistical analyses to assess the long-term effectiveness of MOOCs, particularly in developing practical clinical skills in undergraduate health education.

References

Almeda, M. V., Zuech, J., Utz, C., Higgins, G., Reynolds, R., & Baker, R. S. (2018). Comparing the factors that predict completion and grades among for-credit and open/MOOC students in online learning. *Online Learning*, *22*(1), 1–18. https://doi.org/10.24059/olj.v22i1.1060

Alturkistani, A., Car, J., Majeed, A., Brindley, D., Wells, G., & Meinert, E. (2018). Determining the effectiveness of a massive open online course in data science for health. In *12th International Conference on e-Learning 2018* (pp. 27–36), Madrid, Spain. https://files.eric.ed.gov/fulltext/ED590297.pdf

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press.

Barteit, S., Guzek, D., Jahn, A., Bärnighausen, T., Jorge, M. M., & Neuhann, F. (2020). Evaluation of e-learning for medical education in low-and middle-income countries: A systematic review. *Computers & Education*, *145*, 103726. https://doi.org/10.1016/j.compedu.2019.103726

Barteit, S., Sié, A., Yé, M., Depoux, A., Louis, V. R., & Sauerborn, R. (2019). Lessons learned on teaching a global audience with massive open online courses (MOOCs) on health impacts of climate change: a commentary. *Globalization and Health, 15*, 1–5. https://doi.org/10.1186/s12992-019-0494-6

Bettiol, S., Psereckis, R., & MacIntyre, K. (2022). A perspective of massive open online courses (MOOCs) and public health. *Frontiers in Public Health, 10*. https://doi.org/10.3389/fpubh.2022.1058383

Clark, K. R., Vealé, B. L., & Watts, L. K. (2017). A review of the use of massive open online courses (MOOCs) in medical imaging education. *Internet Journal of Allied Health Sciences and Practice*, *15*(2), 1. https://doi.org/10.46743/1540-580X/2017.1659

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46. https://doi.org/10.1177/001316446002000104

Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2010). Instructional design variations in Internet-based learning for health professions education: A systematic review and meta-analysis. *Academic Medicine, 85*(5), 909–922. https://doi.org/10.1097/ACM.0b013e3181d6c319

Cui, S. (2022). Analysis and quality improvement research of MOOC. In *Proceedings of the 6th International Conference on Digital Technology in Education* (pp. 193–199). https://doi.org/10.1145/3568739.3568772

de Leeuw, R., De Soet, A., Van Der Horst, S., Walsh, K., Westerman, M., & Scheele, F. (2019). How we evaluate postgraduate medical e-learning: Systematic review. *JMIR Medical Education*, *5*(1), e13128. https://doi.org/10.2196/13128

Dedeilia, A., Sotiropoulos, M. G., Hanrahan, J., Janga, D., Dedeilias, P., & Sideris, M. (2020). Medical and surgical education challenges and innovations in the COVID-19 era: A systematic review. *In Vivo, 34*, 1603–1611. https://doi.org/10.21873/INVIVO.11950

Delungahawatta, T., Dunne, S. S., Hyde, S., Halpenny, L., McGrath, D., O'Regan, A., & Dunne, C. P. (2022). Advances in e-learning in undergraduate clinical medicine: A systematic review. *BMC Medical Education*, *22*(1), 711. https://doi.org/10.1186/s12909-022-03773-1

Dolores-Maldonado, G., Cañari-Casaño, J. L., Montero-Romainville, R., & Malaga, G. (2021). Massive open online course (MOOC) opportunities in health education (HE) in a mandatory social isolation context. *F1000Research*, *10*. https://doi.org/10.12688/f1000research.52049.2

Eglseer, D. (2023). Development and evaluation of a massive open online course (MOOC) for healthcare professionals on malnutrition in older adults. *Nurse Education Today*, *123*, 105741. https://doi.org/10.1016/j.nedt.2023.105741

Gao, J., Yang, L., Zou, J., & Fan, X. (2021). Comparison of the influence of massive open online courses and traditional teaching methods in medical education in China: A meta-analysis. *Biochemistry and Molecular Biology Education*, *49*(4), 639–651. https://doi.org/10.1002/bmb.21523

Garrison, D. R. (2016). *E-learning in the 21st century: A community of inquiry framework for research and practice*. Routledge. https://doi.org/10.4324/9781315667263

Goh, W., Ayub, E., Wong, S. Y., & Lim, C. L. (2017). The importance of teacher's presence and engagement in MOOC learning environment: A case study. In *2017 IEEE Conference on e-Learning, e-Management and e-Services* (pp. 127-132). https://doi.org/10.1109/IC3e.2017.8409250

Gong, Z. (2018). The development of medical MOOCs in China: Current situation and challenges. *Medical Education Online*, *23*(1), 1527624. https://doi.org/10.1080/10872981.2018.1527624

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews*. Sage.

Guerrero-Quiñonez, A. J., Bedoya-Flores, M. C., Mosquera-Quiñonez, E. F., Ango-Ramos, E. D., & Lara-Tambaco, R. M. (2023). The MOOC as an alternative model for university education. *Ibero-American Journal of Education & Society Research*, *3*(1), 280–286. https://doi.org/10.56183/iberoeds.v3i1.630

Guo, Y. (2014). MOOCs and the future of China's higher education. *Journal of Higher Education Management*, *8*(5), 29–33. https://doi.org/10.13316/j.cnki.jhem.2014.05.002

Hauer, K. E., Chang, A., van Schaik, S. M., Lucey, C., Cowell, T., & Teherani, A. (2023). It's all about the trust and building a foundation: Evaluation of a longitudinal medical student coaching program. *Teaching and Learning in Medicine*, *35*(5), 550–564. https://doi.org/10.1080/10401334.2022.2111570

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45–58. https://doi.org/10.1016/j.edurev.2014.05.001

Hollands, F. M., & Tirthali, D. (2014). Why do institutions offer MOOCs? *Online Learning*, *18*(3), n3. http://dx.doi.org/10.24059/olj.v18i3.464

Hopcan, S., Polat, E., & Albayrak, E. (2024). Research trends in e-learning practices for postgraduate medical education: A systematic review. *Education and Information Technologies*, *29*(5), 5921–5945. https://doi.org/10.1007/s10639-023-12035-6

Iniesto, F., McAndrew, P., Minocha, S., & Coughlan, T. (2016). Accessibility of MOOCs: Understanding the provider perspective. *Journal of Interactive Media in Education*, *2016*(1). https://doi.org/10.5334/jime.430

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *International Review of Research in Open and Distributed Learning*, *15*(1), 133–160. https://doi.org/10.19173/irrodl.v15i1.1651

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *International Review of Research in Open and Distributed Learning*, *16*(3), 341–358. https://doi.org/10.19173/irrodl.v16i3.2112

Khalil, M., & Ebner, M. (2018). Can learning analytics find success in didactical measurements? Results from a MOOC case study. In D. Ifenthaler (Ed.), *Digital workplace learning: Bridging formal and informal learning with digital technologies* (pp. 211–225). Springer. https://doi.org/10.1007/978-3-319-46215-8_12

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2013). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. *Computers & Education, 104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction*, *22*(2), 1–24. https://doi.org/10.1145/2699735

Kononowicz, A. A., Berman, A. H., Stathakarou, N., McGrath, C., Bartyński, T., Nowakowski, P., Malawski, M., & Zary, N. (2015). Virtual patients in a behavioral medicine massive open online course (MOOC): A case-based analysis of technical capacity and user navigation pathways. *JMIR Medical Education, 1*(2), e8. https://doi.org/10.2196/mededu.4394

Kucuk, S., Aydemir, M., Yildirim, G., Arpacik, O., & Goktas, Y. (2013). Educational technology research trends in Turkey from 1990 to 2011. *Computers & Education*, *68*, 42–50. https://doi.org/10.1016/j.compedu.2013.04.016

Lahti, M., Hätönen, H., & Välimäki, M. (2014). Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: A systematic review and meta-analysis. *International Journal of Nursing Studies*, *51*(1), 136–149. https://doi.org/10.1016/j.ijnurstu.2012.12.017

Laurillard, D., & Kennedy, E. (2017). The potential of MOOCs for learning at scale in the Global South. *Center for Global Higher Education*, *31*. http://dx.doi.org/10.5287/ora-gagqxyjje

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008–2012. *The International Review of Research in Open and Distributed Learning, 14*(3), 202–227. https://doi.org/10.19173/irrodl.v14i3.1455

Liyanagunawardena, T. R., & Williams, S. A. (2014). Massive open online courses on health and medicine. *Journal of Medical Internet Research*, *16*(8), e191. https://doi.org/10.2196/jmir.3439

Longhini, J., De Colle, B., Rossettini, G., & Palese, A. (2021). What knowledge is available on massive open online courses in nursing and academic healthcare sciences education? A rapid review. *Nurse Education Today*, *99*, 104812. https://doi.org/10.1016/j.nedt.2021.104812

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education*, 80, 77–83. https://doi.org/10.1016/j.compedu.2014.08.005

Martín-Valero, R., Pastora-Bernal, J. M., Ortiz-Ortigosa, L., Casuso-Holgado, M. J., Pérez-Cabezas, V., & Ruiz-Párraga, G. T. (2021). The usefulness of a massive open online course about postural and technological adaptations to enhance academic performance and empathy in health sciences undergraduates. *International Journal of Environmental Research and Public Health*, *18*(20), 10672. https://doi.org/10.3390/ijerph182010672

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1–9. https://doi.org/10.1186/2046-4053-4-1

Naciri, A., Radid, M., Kharbach, A., & Chemsi, G. (2021). E-learning in health professions education during the COVID-19 pandemic: A systematic review. *Journal of Educational Evaluation for Health Professions*, *18*. https://doi.org/10.3352/jeehp.2021.18.27

Nieder, J., Nayna Schwerdtle, P., Sauerborn, R., & Barteit, S. (2022). Massive open online courses for health worker education in low-and middle-income countries: A scoping review. *Frontiers in Public Health*, *10*, 891987. https://doi.org/10.3389/fpubh.2022.891987

Nwameme, A., Dako-Gyeke, P., Asampong, E., Allotey, P., Reidpath, D. D., Certain, E., Vahedi, M., Ross, B., Maher, D., & Launois, P. (2023). Improving understanding of disease control implementation research through a MOOC with participants from low-and middle-income countries: Evaluating participant reactions and learning. *PLoS Neglected Tropical Diseases*, *17*(3), e0011139. https://doi.org/10.1371/journal.pntd.0011139

Olivares Olivares, S. L., Hernández, R. I. E., Corolla, M. L. T., Alvarez, J. P. N., & Sánchez-Mendiola, M. (2021). MOOC learning assessment in clinical settings: Analysis from quality dimensions. *Medical Science Educator*, *31*(2), 447–455. https://doi.org/10.1007/s40670-020-01178-7

Pouladi, E., Yazdi, M. T., Salehi, M., & Moradi, S. (2023). Use of massive online open courses (MOOCs) model in virtual education development in first macro-region universities of medical sciences in countrywide territorial planning. *Journal of Education and Health Promotion*, *12*, 326. https://doi.org/10.4103/jehp.jehp_1221_22

Regmi, K., & Jones, L. (2020). A systematic review of the factors—enablers and barriers—affecting e-learning in health sciences education. *BMC Medical Education, 20*(1), 1–18. https://doi.org/10.1186/s12909-020-02007-6

Rowe, M., Osadnik, C. R., Pritchard, S., & Maloney, S. (2019). These may not be the courses you are seeking: a systematic review of open online courses in health professions education. *BMC Medical Education*, *19*, 1–11. https://doi.org/10.1186/s12909-019-1774-9

Rulinawaty, R., Priyanto, A., Kuncoro, S., Rahmawaty, D., & Wijaya, A. (2023). Massive open online courses (MOOCs) as catalysts of change in education during unprecedented times: A narrative review. *Jurnal Penelitian Pendidikan IPA*, *9*(Special issue), 53–63. https://doi.org/10.29303/jppipa.v9iSpecialIssue.6697

Schettino, G., & Capone, V. (2022). Learning design strategies in MOOCs for physicians' training: A scoping review. *International Journal of Environmental Research and Public Health*, *19*(21), 14247. https://doi.org/10.3390/ijerph192114247

Setia, S., Tay, J. C., Chia, Y. C., & Subramaniam, K. (2019). Massive open online courses (MOOCs) for continuing medical education: Why and how? *Advances in Medical Education and Practice*, *2019*(10), 805–812. https://doi.org/10.2147/AMEP.S219104

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, *43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Vissenberg, J., d'Haenens, L., & Livingstone, S. (2022). Digital literacy and online resilience as facilitators of young people's wellbeing? A systematic review. *European Psychologist*, *27*(2), 76–85. https://doi.org/10.1027/1016-9040/a000478

Yang, L., Zou, J., Gao, J., & Fan, X. (2023). Assessing the effectiveness of massive open online courses on improving clinical skills in medical education in China: A meta-analysis. *Heliyon*, *9*(8), e19263. https://doi.org/10.1016/j.heliyon.2023.e19263

Yuan, L., & Powell, S. (2013). *MOOCs and open education: Implications for higher education*. CETIS. https://e-space.mmu.ac.uk/619735/

Zhu, M., Sari, A., & Lee, M. M. (2018). A systematic review of research methods and topics of the empirical MOOC literature (2014–2016). *The Internet and Higher Education*, *37*, 31–39. https://doi.org/10.1016/j.iheduc.2018.01.002

International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 27, Number 1

February – 2026

A Meta-Analysis of ChatGPT's Influence on Learning Achievement

Min Young Doo[1] and Yeonjeong Park[2]
[1]College of Education, Kangwon National University; [2]College of Education, Jeonbuk National University

Abstract

This meta-analysis synthesized empirical findings on the influence of ChatGPT on learning achievement. An electronic database search using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines was conducted with relevant keywords to identify eligible research studies published between November 2022 and December 2024. A total of 22 eligible publications that met our inclusion criteria were reviewed. ChatGPT had a moderate positive effect on learning achievement (g = 0.573), indicating its potential to enhance learning outcomes. Subgroup analysis revealed that ChatGPT had a larger effect on middle and high school students (g = 0.928) than on undergraduates (g = 0.538), although the difference was not statistically significant. This finding highlights the importance for instructors and educational practitioners to consider the applications of ChatGPT in middle and high school settings. No significant statistical differences were found among the cognitive, affective, and metacognitive domains. Given that nearly half the studies focused on the cognitive domain, it is important to diversify the application of generative AI across a variety of subjects in different learning domains. The most frequently used instructional approaches with ChatGPT applications were lectures (22.1%) and self-regulated learning (16.3%). The effect of ChatGPT was largest on self-regulated learning (g = 1.115), followed by case-based learning (g = 0.836), while its effect was smallest on game-based learning (g = 0.092). This study was conducted within two years of ChatGPT's emergence, limiting our ability to analyze a large number of publications. Nevertheless, this study offers meaningful implications for future research on the application of ChatGPT for educational purposes.

Keywords: ChatGPT, learning achievement, generative AI, artificial intelligence in education, AIED

Introduction

Artificial intelligence (AI) has been defined by Popenici et al. (2017) as "computing systems that are able to engage in human-like processes such as learning, adapting, synthesizing, self-correction, and the use of data for complex processing tasks" (p. 2). ChatGPT by OpenAI, a representative commercially available generative AI, has rapidly gained attention since its emergence in November 2022. It is expected to bring about fundamental and unprecedented changes in society. As generative AI is evolving at a fast speed and is becoming more user friendly (e.g., from ChatGPT 3.0 in 2022 to GPT-4o in 2024), far more people have begun to use it in their daily lives. Reuters (2024) reported that OpenAI estimated the number of weekly ChatGPT users would exceed 200 million by the summer of 2024. The popularity of OpenAI has spread to education fields as well. Many organizations, such as UNESCO (Sabzalieva & Valentini, 2023) and the Council of Europe (2022) released special reports or issue papers about the application of AI in education and expected change in the field. Growing academic interest in artificial intelligence in education (AIED) has also prompted many top-tier journals to publish special issues on generative AI applications in education. In addition, new journals have recently launched that exclusively focus on AIED.

Díaz and Nussbaum (2024) defined AIED as "an educational technology capable of detecting patterns in existing or in vivo data and making automatic instructional decisions that are developed or implemented for pedagogical purposes to enhance the teaching and learning process" (p. 2). As Gagne's taxonomy of learning delineates (Gagne & Briggs, 1984), students traditionally have been expected to memorize definitions and concepts, apply rules and principles, understand cause and effect, and solve well-defined or ill-defined problems in the learning process. In contrast, generative AI, such as ChatGPT, requires students to proactively ask questions instead of passively answering given questions. Rospigliosi (2023) described the interaction or conversation between ChatGPT and its users, where a thread of user questions and ChatGPT responses enables an interactive learning process. This process encourages students to ask follow-up questions and challenge the responses of ChatGPT leading to enhanced learning engagement. This is the most distinctive feature of generative AI compared to other instructional media or intelligent tutoring systems (ITS) that have been adopted in education (Lo et al., 2024).

Educators have high expectations that generative AI can facilitate personalized learning, characterized by tailoring learning experiences for individuals and flexible learning paths (Gunawardena et al., 2024). Generative AI is also expected to accelerate the shift toward a learner-centered paradigm (Lee & Moore, 2024) and enhance learning effectiveness and efficiency (Alneyadi & Wardat, 2023; Hsu, 2024; Teng, 2024). UNESCO delineated potential areas for the application of AI in higher education, including teaching and learning, research, administration, and community engagement (Sabzalieva & Valentini, 2023). Crompton and Burke (2023) identified applications of AI in higher education with a more focused emphasis on teaching and learning, including assessment/evaluation, prediction, AI assistants, ITS, and managing student learning. Similarly, Díaz and Nussbaum's (2024) findings included ITS, prediction, diagnostics, adaptive systems, and assessment. These results imply that generative AI including ChatGPT is expected to have an impact across all areas of education.

Despite the high expectations of generative AI for pedagogical purposes, educators are also concerned about side effects and negative influences of ChatGPT applications on learning. UNESCO raised ethical concerns about using ChatGPT in education, such as academic integrity, cognitive bias, accessibility, and gender and diversity issues (Sabzalieva & Valentini, 2023). In their systematic review, Lo et al.

(2024) found that AI may reduce critical thinking and lead to overreliance on ChatGPT, contributing to student disengagement. They also noted an increase in plagiarism and cheating due to the use of ChatGPT in education. Dikilitaş et al. (2024) also raised concerns that excessive use of ChatGPT in writing or problem-solving activities may lower students' self-regulated learning. Many researchers have recently conducted rigorous research to examine the effects of ChatGPT on learning outcomes with high expectations of using ChatGPT for learning.

The purpose of this study was to synthesize current empirical research findings on the influence of ChatGPT on learning achievement using a meta-analysis method. Wu and Yu (2024) criticized the lack of meta-analyses synthesizing quantitative studies on the effects of generative AI, in contrast to the dozens of systematic review studies and substantial number of conceptual papers on the same theme (e.g., Crompton & Burke, 2023; Díaz & Nussbaum, 2024; Jeon et al., 2023).

A meta-analysis enables researchers to synthesize individual quantitative research findings and provide more robust and convincing conclusions (Borenstein et al., 2009). It may be premature to conduct a meta-analysis to synthesize empirical studies examining the influence of ChatGPT on learning achievement, as only two years had passed since the emergence of ChatGPT at the time this study was conducted. However, given that ChatGPT continues to evolve with growing attention and widespread use in education, it is worthwhile to conduct a meta-analysis to understand the status of ChatGPT application in education in the early stage of ChatGPT and to provide direction for future research.

Theoretical Framework

Driscoll (1993) defined learning as "a persisting change in human performance or performance potential" resulting from the learner's experience and interaction with the world. In line with this perspective, the present study conceptualizes learning achievement as measurable changes in students' attitudes, cognition, and behaviors. This broad definition acknowledges that achievement encompasses not only academic performance but also affective dispositions and metacognitive strategies that shape long-term learning potential.

Historically, the integration of new technologies into education has been accompanied by debates over their effects on learning outcomes. Saettler (1990) traced the evolution of instructional technology research alongside the emergence of new media, establishing the tradition of media comparison research. Within this paradigm, scholars have investigated whether novel technologies produce superior learning outcomes relative to earlier tools or traditional instruction. For instance, Steenbergen-Hu and Cooper (2014) synthesized findings on ITS and reported small to moderate effects on student achievement. Yet this tradition has also faced criticism. Warnick and Burbules (2007) cautioned that conflating media with instructional methods can lead to misleading conclusions, a critique that continues to resonate. More recently, Buchner and Kerres (2023) observed that media comparison research remains prevalent in fields such as augmented reality, suggesting that interest in technology–learning effects persists despite its contested status. As a cutting-edge technology, ChatGPT is increasingly integrated into teaching and learning, and there is growing interest in investigating its effects on learning achievement despite the controversy in media comparison research.

To better understand the potential influence of ChatGPT, we considered the following theoretical perspectives. Bloom's taxonomy and its revision by Anderson and Krathwohl (2001) highlighted the multiple dimensions of learning achievement, ranging from factual recall to higher-order reasoning and self-regulation. Cognitive load theory (Sweller, 1988) has been supported by several recent studies (e.g., Becker et al., 2025; Martin et al., 2025; Patac & Patac, 2025), which have suggested that ChatGPT may reduce extraneous cognitive demands by providing immediate explanations, thereby allowing learners to focus on deeper processing. From a sociocultural perspective (Vygotsky, 1978), ChatGPT can be viewed as a form of digital scaffolding that supports learners in their zone of proximal development through interactive prompts and adaptive dialogue. Finally, frameworks such as the Technology Acceptance Model (Davis, 1989) and the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003) explain how learners' perceptions of usefulness and ease of use shape their willingness to integrate ChatGPT into their study practices, thereby influencing whether its affordances translate into measurable achievement gains. Taken together, these perspectives form a conceptual foundation for the present meta-analysis. While this study did not test a single overarching theory, the combination of media comparison traditions, educational taxonomies, cognitive and sociocultural learning theories, and technology adoption frameworks provides a multifaceted lens for interpreting the synthesized empirical findings on ChatGPT's influence on learning achievement.

Previous Reviews Examining ChatGPT's Influence on Learning

Several meta-analyses have recently examined the impact of ChatGPT and related AI chatbots on student learning, yet their findings reveal both promising effects and important limitations. For example, Deng et al. (2025) conducted one of the most comprehensive reviews, synthesizing 69 experimental studies published between 2022 and 2024. They reported strong positive effects of ChatGPT on academic achievement, affective states, and higher-order thinking, while finding weaker or inconsistent results for self-efficacy and mental effort. Their analysis further indicated that contextual factors such as intervention setting and duration moderated the outcomes, with classroom-based and longer interventions producing stronger effects. However, their study was restricted to experimental designs, thereby excluding correlational evidence and limiting the scope of generalization.

Building on this foundation, Wu and Yu (2024) expanded the focus to AI chatbots more broadly, analyzing 24 empirical studies across diverse educational settings. Their results also indicated large effect sizes for performance, motivation, self-efficacy, and interest, but a negative association with anxiety. Unlike Deng et al. (2025), Wu and Yu did not examine how instructional approaches or learning domains might shape these outcomes, which makes it difficult to translate their findings into concrete pedagogical practices. Similarly, Sun and Zhou (2024) narrowed their scope to college students, showing medium overall effects of generative AI on achievement and highlighting the effectiveness of text-based and independent learning conditions. Yet their review excluded younger learners and did not clarify how differences across school levels or disciplines might influence the impact of generative AI. Together, these three meta-analyses underscore the educational potential of ChatGPT and related tools, but each is constrained by its limited scope of participants, contexts, or moderating variables.

Beyond meta-analytic evidence, a number of systematic reviews have provided complementary insights into how ChatGPT and generative AI are being used in education. For instance, Lo et al. (2024) synthesized 72 studies on student engagement and found mixed outcomes, with relatively stronger evidence for behavioral engagement but weaker or inconsistent patterns for cognitive and affective engagement. Their review illustrated the breadth of ChatGPT applications but, lacking effect-size estimates, could not quantify the magnitude of such effects. Similarly, Dikilitaş et al. (2024) reviewed a small number of early empirical studies in higher education and identified themes such as skill development, feedback, motivation, and ethical concerns. Although valuable, their review was necessarily limited by the short timeframe since ChatGPT's release. At the K–12 level, Díaz and Nussbaum (2024) offered a theoretically grounded review of 183 AI application studies using the Human-Centered AI framework, classifying interventions by learning theories such as constructivism and experiential learning. While their approach has enriched theoretical understanding, it does not provide quantitative evidence on learning outcomes. Taken together, these reviews highlight the rapid growth of ChatGPT-related research but also indicate the absence of systematic synthesis across learner groups, learning domains, and instructional approaches.

Earlier work on ITS has provided additional historical context. Steenbergen-Hu and Cooper's (2014) meta-analysis revealed that ITS had small to moderate effects on undergraduates' academic learning, with stronger outcomes compared to self-reliant learning but weaker results relative to human tutoring. Although situated in a pre-generative AI era, their findings point to the importance of considering both instructional context and comparator conditions when evaluating AI-based educational technologies. This insight remains relevant for assessing ChatGPT's role in contemporary classrooms.

Collectively, the existing body of reviews has provided important but partial insights. Meta-analyses have demonstrated substantial overall effects but have been often constrained by narrow participant groups, selective outcome measures, or limited consideration of moderating factors. Systematic reviews have deepened theoretical perspectives and mapped emerging trends but lack the empirical synthesis needed to quantify impact. These limitations underscore the need for a broader and more integrative meta-analysis that synthesizes recent empirical studies on ChatGPT, incorporates diverse learner populations, and explicitly examines the moderating effects of instructional approaches, learning outcomes, and research methods. The present study aims to address these gaps by offering a comprehensive meta-analysis of ChatGPT's influence on learning achievement across educational contexts.

Method

Given that the purpose of this study was to examine the effects of ChatGPT on learning achievement, we formulated the following research questions:

1. What are the overall effects of ChatGPT on learning achievement?
2. Are the effects of ChatGPT on learning achievement moderated by the school levels of participants?
3. How do the types of learning outcomes moderate the effects of ChatGPT on learning achievement?
4. Are the effects of ChatGPT on learning achievement moderated by instructional approaches?
5. Are the effects of ChatGPT on learning achievement moderated by research methods?

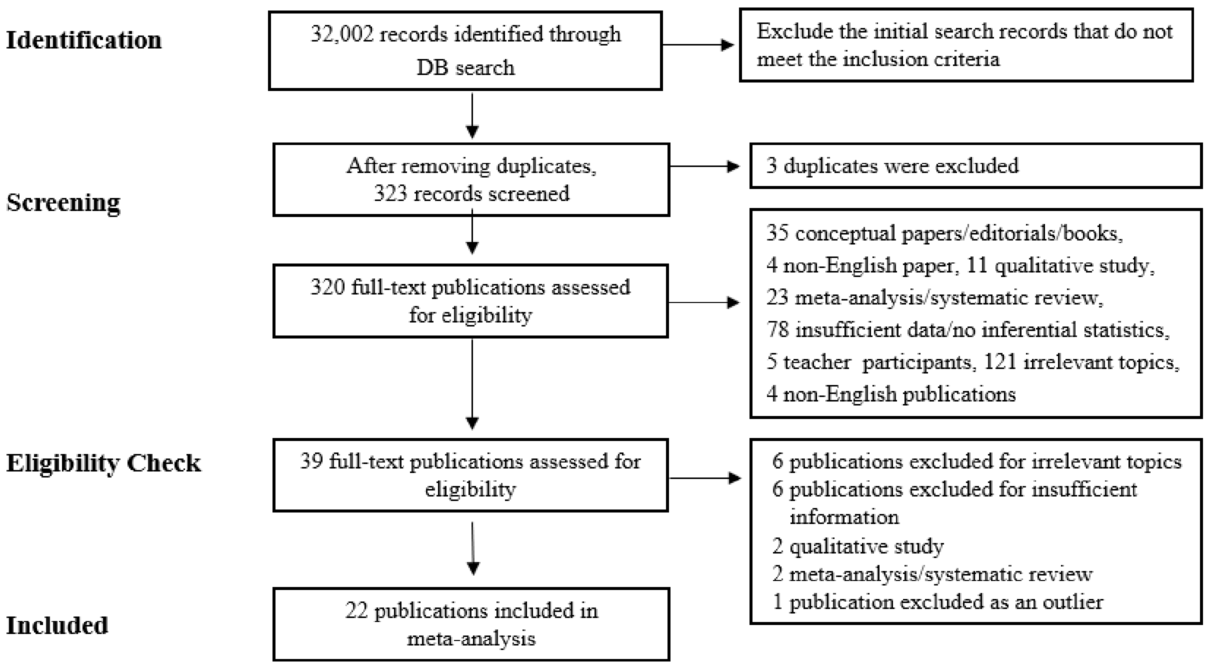
Literature Search

To search for eligible articles in an effective and efficient manner, we set five inclusion criteria satisfying the purpose of this research. The studies must have: (a) examined the effects of ChatGPT on learning achievement; (b) been published between November 2022 (the release of ChatGPT) and December 2024 (the time of the literature search); (c) adopted a rigorous research design; (d) reported quantitative learning outcomes (i.e., inferential statistics); and (e) been written and published in English. Exclusion criteria included qualitative studies (e.g., Garcia-Varela et al., 2025), conceptual papers, meta-analysis or systematic reviews (i.e., secondary data analysis), publications that reported insufficient data for calculating effect sizes, and studies published in languages other than English.

We conducted an electronic database search using Web of Science, Google Scholar, Education Source (EBSCOhost), ERIC (ProQuest), PsycINFO, and JSTOR for Dissertations & Theses. We chose these databases because they cover most publications in education and many previous meta-analysis studies have used these electronic databases to search for relevant literature (Coban et al., 2022; Schoenherr et al., 2024). The keywords used for the search included "ChatGPT," "generative AI," "AIED," "artificial intelligence in education," "learning achievement," "learning outcomes," and their combinations. Using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), the initial database search yielded 32,002 publications, which were reduced to 320 publications after excluding redundant or ineligible studies that did not satisfy the inclusion criteria (see Figure 1). We conducted a full-text review of 320 publications and examined the eligibility of each publication. Through this process, 39 relevant samples remained. In addition, in the full-text review process, six samples were not included due to irrelevance to the topic, two samples were excluded for qualitative research, two samples were systematic reviews or meta-analyses, and another seven samples were excluded for insufficient information. Thus, a total of 86 independent studies (i.e., k = 86) in 22 publications were included for analysis.

Figure 1

Search and Exclusion Process



Note. DB = database. Adapted from *PRISMA Flow Diagram*, by PRISMA, 2020 (https://www.prisma-statement.org/prisma-2020-flow-diagram). CC BY 4.0.

Information Gathering and Extracting Information From Studies

From the 22 publications, 86 independent studies were coded as we extracted information related to four types of variables: independent variables, moderating variables, dependent variables, and other variables (see Appendix). The extracted information was coded using a coding scheme in an Excel file. As illustrated in Table 1, we examined the effects of the moderating variables on the relationship between the independent and dependent variables, including the types of learning achievement, school levels of learners, learning disciplines, instructional approaches, and research methods.

Table 1

Coding Information for Meta-Analysis

Variable type	Categories	Coding information
Independent	Implementation of ChatGPT	ChatGPT versions
Moderating	School level	PreK–12, elementary, middle, and high school, higher education, adult learners
	Instructional approach	Direct instruction, PBL, flipped learning, team projects/collaborative learning, independent study
	Learning discipline	Languages, mathematics, STEM, humanities, business, social science, natural science, education, others, and unspecified
	Learning environment	Online learning, traditional classrooms, blended learning including flipped learning
Dependent	Learning outcomes	Affective, cognitive, metacognitive, behavioral (e.g., participation) Effect sizes: means and standard deviation, *t*-values, *F*-values, *r*, *p*-values, sample sizes
Other	Publication details	Title, author, year, journals, countries of publication
	Publication type	Journal articles, proceedings, unpublished dissertations or thesis, technical reports
	Research design	Experimental, quasi-experimental, one-group pretest-posttest design, and non-experimental (i.e., correlational studies)

Note. PBL = Problem-based learning; STEM = Science, technology, engineering, and mathematics

We coded the samples separately using the coding scheme detailed in Table 1. After discussing the coding scheme, we reached 91.7% intercoder agreement for the initial coding. Disagreements regarding coding were resolved through discussion of each case, after which we reached full consensus.

A brief description of the studies included in this meta-analysis is as follows. Among the 22 publications, five papers were published in 2023 (22.7%), and 17 papers (77.3%) were published in 2024. In terms of publication types, 21 publications were peer-reviewed journal articles, and one was a conference proceeding (i.e., Xue et al., 2024). These publications were published in 12 countries. Taiwan was the most productive country (seven studies, 31.8%), followed by China (four studies, 18.2%) and Turkey (two studies, 9.1%). Several countries published one study, including Arab Emirates, Czech, Ghana, Hong Kong, Republic of Korea, Macau, Mexico, Norway, and the USA. Sample sizes in the publications ranged from 31 to 269, with an average sample size of 86.6.

Since different research designs were used in the publications, we examined the descriptive data for each effect size (i.e., k = 86), rather than by publication. We adopted Lo et al.'s (2024) category of research design, which is based on Creswell's (2012) classification. Among the research designs, the quasi-experimental design was the most frequently adopted (48 studies, 55.8%). The pretest-posttest experimental design with no control group was used in 18 studies (20.9%), while 10 studies (11.6%) employed a true-experimental design (i.e., pretest-posttest experimental design with random assignment). Additionally, two studies were correlational studies (2.3%), and eight studies (9.3%) used an experimental design without a pretest. Random assignment of participants was used in 38 studies (44.2%). There was a wide range of participants including elementary students (2 studies, 2.3%), middle and high school students (9 studies, 10.5%), undergraduates (61 studies, 70.9%), and adults including adult learners and clinical teachers (14 studies, 16.3%). Pre-K–12 participants (i.e., kindergarteners) were not included in this analysis. It is notable that more than half of the studies included undergraduates as participants in the analysis.

Learning outcomes as the dependent variable of this study were coded based on the four types of learning outcomes suggested by Bloom's taxonomy of learning objectives (Anderson, & Krathwohl, 2001) and Doo et al. (2023): (a) affective, (b) cognitive, (c) meta-cognitive, and (d) behavioral learning outcomes (i.e., participation). The affective domain included emotional engagement (Liang et al., 2024; Teng, 2024), learning motivation (Li, 2023; Ng et al., 2024; Teng, 2024; Yilmaz & Yilmaz, 2023), flow (Chen & Hou, 2024), attitudes toward learning, an intention for continuous learning, learning satisfaction, frustration, or anxiety (Chen & Hou, 2024; Liao et al., 2024). The cognitive domain of learning included mental activities, such as understanding and applying (Alneyadi & Wardat, 2023; Hsu, 2024; Huesca et al., 2024; Ng et al., 2024; Svendsen et al., 2024; Zhou & Kim, 2024), analyzing concepts and theories, generating knowledge (Guo et al., 2023; Xue et al., 2024) or problem solving (Chang et al., 2024; Lee et al., 2024; Urban et al., 2024). The behavioral domain included frequencies of logins, participation in learning activities, learning time, and completion of courses. The meta-cognitive domain of learning included higher-order thinking (Lee et al., 2024), planning, monitoring, critical thinking (Chang et al., 2024; Chang & G.-H. Hwang, 2024; Chang & G.-J. Hwang, 2024; Essel et al., 2024), collective efficacy (Urban et al., 2024), self-evaluation, and applications of self-regulatory strategies (Lee et al., 2024; Liao et al., 2024). In our research, nearly half the studies examined learning outcomes in the cognitive domain (44 studies, 51.2%) followed by the affective domain (25 studies, 29.0%) and the meta-cognitive domain (17 studies, 19.8%). No studies assessed learning outcomes in the psychomotor domain.

We also reviewed the instructional approaches used in the application of ChatGPT in each study. Of the 86 studies, 26 did not report the instructional approaches (30.2%). The most frequently used instructional approaches were lecture (19 studies, 22.1%) and self-regulated learning (nine studies, 10.5%) followed by problem-based learning (12 studies, 13.9%) and game-based learning (eight studies, 9.3%). Other instructional approaches included project-based learning (six studies, 7.0%), flipped learning (three studies, 3.5%), and case-based learning (three studies, 3.5%).

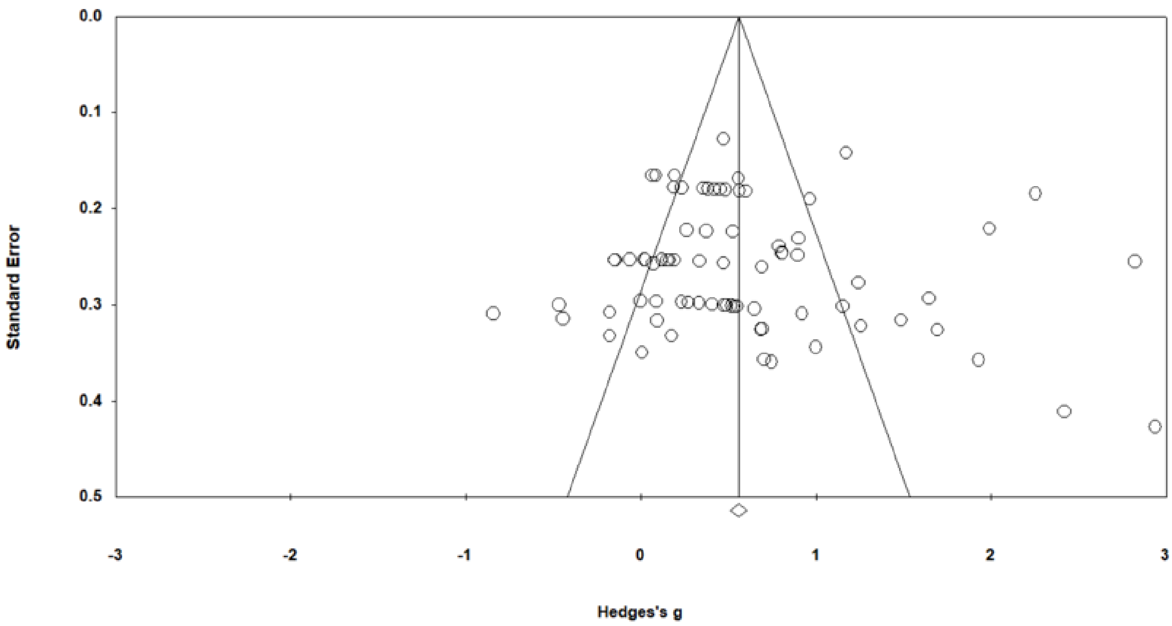
Analysis and Integration of Outcomes

We adopted Cohen's (1988) criteria to interpret the magnitude of effect sizes: 0.2 for a small effect size, 0.5 for a medium effect size, and 0.8 for a large effect size. We followed Higgins and Thompson's (2002) suggestions for interpreting I^2 scores: less than 30% would indicate mild heterogeneity, while more than 50% would indicate substantial heterogeneity. Depending on the heterogeneity of the effect sizes, we decided to use a random effects model. Among the different types of effect sizes, this study used Hedge's g as the type of effect size because more studies examined the group differences (e.g., pretest-posttest control group) than the relationship between the independent variables and dependent variables (e.g., correlation studies) in the analysis. In addition, because previous meta-analyses (e.g., Deng et al., 2025; Steenbergen-Hu & Cooper, 2014; Sun & Zhou, 2024) have used Hedge's g, it allowed us to easily compare our findings to their findings.

Publication bias was assessed using a funnel plot of the effect size and Egger's regression test. The funnel plot of effect size is slightly skewed to the right, as illustrated in Figure 2. Therefore, we double-checked for the absence of publication bias using Egger's regression tests. The results confirmed that publication bias was not detected (p = .59, t = .53; Egger et al., 1997). The meta-analysis was performed using Comprehensive Meta-Analysis (CMA Version 4.0) and IBM SPSS Statistics (Version 28.0).

Figure 2

Funnel Plot of Effect Size



Results

Overall Effects of ChatGPT on Learning Achievement

We synthesized the influences of ChatGPT on learning achievement using CMA software. The effect sizes ranged from g = -0.84 (Guo et al., 2023) to g = 2.94 (Teng, 2024), with the prediction interval [-.47, 1.61]. Statistics indicated substantial heterogeneity across the studies, Q = 511.1, df = 85, p < .001, I^2 = 83.37. Hence, we used a random-effects model to estimate the effect size and to examine the moderating effects of the variables (e.g., school level of learners, type of learning outcomes, type of instructional approaches, and research method) using sub-group analysis. The overall effect size of the influence of ChatGPT on learning achievement is shown in Table 2. According to Cohen's (1988) criteria for effect size, the overall effect size of this study was medium-sized.

Table 2

Estimation of the Overall Effect Size Using the Random-Effects Model

Model	*k*	*N*	ES(*g*)	*SE*	Variance	95% CI	*Z* (2-tailed)	*p*
Random	6,992	86	0.573	0.063	0.004	[0.451, 0.696]	9.170	< .001

Note. ES = effect size.

Sub-Group Analysis Based on School Levels of Participants

We conducted a sub-group analysis based on the school levels of students (i.e., elementary school, middle and high school, undergraduate, adult learner, and clinical teacher) in the influence of ChatGPT on learning achievement. Results are illustrated in Table 3. The effect size of middle and high school students was larger than other learner groups. Elementary school students showed the smallest effect size. However, these results should be interpreted with caution due to the small number of effect sizes for elementary school studies (k = 2). The sub-group analysis results showed no significant differences among the effect sizes of sub-groups based on school levels, $Q_B(4) = 4.281$, $p = .369$. The effect size of middle and high school students was larger than it was with undergraduates. However, there were no statistical differences between the two groups. There were also no significant differences in the effect size between adult learners and clinical teachers in the same age range, despite their different roles and responsibilities in using ChatGPT.

Table 3

Effect of Learner Type on Influence of ChatGPT on Learning Achievement

Learner type	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z* (2-tailed)	*p*
Adult learner	9	526	0.452	0.196	[0.067, 0.837]	2.301	.021
Clinical teacher	5	370	0.625	0.256	[0.123, 1.128]	2.439	.015
Elementary school	2	78	0.388	0.433	[-0.461, 1.237]	0.896	.370
Middle and high school	9	940	0.928	0.191	[0.554, 1.302]	4.861	.000
Undergraduate	61	5,088	0.538	0.075	[0.391, 0.684]	7.200	.000

Note. ES = effect size; CI = confidence interval.

Sub-Group Analysis Based on the Type of Learning Outcome

Next, we conducted a sub-group analysis based on the type of learning outcome (i.e., affective, cognitive, or meta-cognitive) in the influence of ChatGPT on learning achievement. Results are shown in Table 4. The effect size in the affective domain was smaller than the cognitive domain and the meta-cognitive domain. However, the results showed no significant differences among the effect sizes of the sub-groups, $Q_B(2) = 0.983$, $p = .612$. The 95% confidence intervals indicated that no type of learning outcome was statistically higher or lower than the others.

Table 4

Influence of ChatGPT on the Type of Learning Outcome

Learning Outcome	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z*	*p*
Affective	27	1,808	0.481	0.112	[0.262, 0.700]	4.303	.000
Cognitive	46	3,757	0.612	0.085	[0.446, 0.779]	7.205	.000
Meta-cognitive	13	1,427	0.619	0.153	[0.319, 0.919]	4.042	.000

Note. ES = effect size. CI = confidence interval.

Sub-Group Analysis Based on Type of Instructional Approach

We compared the effect sizes of the influence of ChatGPT on learning achievement with different instructional approaches (i.e., case-based, flipped learning, game-based, lecture, problem-based, project-based, and self-regulated learning) using sub-group analysis. Table 5 illustrates that self-regulated learning yielded the largest effect size, followed by case-based learning. The smallest effect size was observed with game-based learning, which was not statistically significant. The results indicate significant differences between instructional approaches, $Q_B(7) = 20.01$, $p < .05$.

Table 5

Influence of ChatGPT on the Type of Instructional Approach

Instructional approach	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z*	*p*
Case-based	3	210	0.836	0.327	[0.196, 1.477]	2.560	.010
Flipped learning	3	416	0.626	0.315	[0.008, 1.244]	1.985	.047
Game-based	8	444	0.092	0.204	[-.0308, 0.491]	0.449	.653
Lecture	19	2,180	0.771	0.126	[0.524, 1.017]	6.127	.000
Not reported	26	1,699	0.343	0.114	[0.119, 0.567]	2.998	.003
Problem-based	12	1,325	0.589	0.160	[0.276, 0.902]	3.685	.000
Project-based	6	279	0.553	0.246	[0.071, 1.035]	2.247	.025
Self-regulated learning	9	439	1.115	0.201	[0.720, 1.510]	5.533	.000

Note. ES = effect size. CI = confidence interval.

Sub-Group Analysis Based on Research Method

Using sub-group analysis, we examined the moderating effect of research method (i.e., correlational studies, pretest-posttest control group design, mixed-design, quasi-experimental design, and posttest only design). As Table 6 illustrates, there were significant differences between sub-groups based on the research methods, $Q_B(4) = 18.474$, $p < .05$. An experimental design with posttest only was not statistically significant since its 95% confidence interval included zero. The effect size of correlational studies was significantly larger than mixed method and quasi-experimental studies. However, there was no significant difference between correlational studies and true experimental studies (i.e., pretest-posttest control group studies).

Table 6

Influence of ChatGPT on Research Method

Research method	*k*	*N*	ES(*g*)	*SE*	95% CI	*Z*	*p*
Correlational	2	538	1.698	0.353	[1.007, 2.389]	4.814	.000
True experimental	18	1,769	0.762	0.126	[0.516, 1.009]	6.067	.000
Mixed methods	10	1,200	0.520	0.160	[0.207, 0.833]	3.257	.001
Post-test only	8	469	0.113	0.194	[-0.268, 0.494]	0.580	.580
Quasi-experimental	48	3,016	0.529	0.079	[0.374, .0683]	6.704	.000

Note. ES = effect size. CI = confidence interval.

Discussion

The purpose of this meta-analysis was to synthesize current empirical studies examining the influence of the ChatGPT application on learning achievement. We found a steep increase in the number of publications investigating the influence of ChatGPT on learning achievement, increasing from five papers in 2023 (22.7%) to 17 papers in 2024 (77.3%), which represents a nearly threefold increase in publications over 2 years. This finding supports Deng et al.'s (2025) observation of the increasing trend in ChatGPT-related publications. It also reveals that researchers' interest in the influence of ChatGPT on learning achievement in academia is growing rapidly, and this trend is expected to continue with the extensive adoption of ChatGPT in schools. In terms of the location of the publications, we found that among the 12 countries represented, Taiwan was the most productive (31.8%), followed by mainland China (18.2%). Deng et al. (2025) analyzed the productivity of publications by continent, not by nation, and found also that Asia (approximately 71.0%) was the most productive continent in their research. In addition, Lo et al.'s (2024) study highlighted that more than half the studies on AI applications for educational purposes were published in Asia (58.3%). These findings and those of two other studies suggest that researchers in Asia are more actively exploring the educational applications of ChatGPT compared to those in other continents.

Regarding the first research question, the overall effect size of the influence of ChatGPT on learning achievement was medium (g = 0.573). See Table 2. This supports the findings of Sun and Zhou (2024; g = 0.533) and Deng et al. (2025; g = 0.712). Our results were smaller than those of Wu and Yu (2024), who reported a large effect size (d = 0.964). However, ours were larger than Steenbergen-Hu and Cooper's (2014) effect size (g = 0.32), which estimated the influence of ITS on undergraduates' academic learning. Nevertheless, the medium effect size of ChatGPT's impact on learning achievement in this study shows promise for use of the application. This confirms the findings of previous meta-analytic and many empirical studies indicating that ChatGPT has the potential to improve learning achievement. These results can be interpreted through the lens of cognitive load theory (Sweller, 1988). By reducing extraneous demands through immediate explanations, ChatGPT may enable learners to concentrate on germane processing, which helps explain the consistently positive effects observed in this and prior studies.

Related to the second research question, most participants in the studies we examined were undergraduates (70.9%), followed by middle and high school students (10.5%), and adult learners, including clinical teachers (16.3%). This distribution of was similar to that of Deng et al.'s (2025) research (undergraduates: 84.1%; K–12 students: 14.5%) and Lo et al.'s (2024) systematic review results (undergraduates: 79.2%; secondary school: 8.3%) These three studies indicate that the research on ChatGPT's influence on learning achievement has been more actively conducted in higher education than in K–12 settings.

This observation should be considered in light of the effect size of undergraduates versus K–12 students. Table 3 shows the results of the sub-group analysis based on participants' school level, indicating that the effect size of middle and high school students (g = 0.928) was larger than the effect size of undergraduates (g = 0.538) despite the non-statistical differences between the two groups. Despite the larger effect size for middle and high school students, few studies have examined the influence of ChatGPT on learning achievement in these settings. This discrepancy suggests that more research is needed in this area, and instructors and educational practitioners should work to further implement ChatGPT in middle and high school environments.

However, the research findings of this study are contradictory to that of Wu and Yu (2024), who reported a large effect size of undergraduates (d = 1.079) and a small effect size of secondary students (d = 0.214, *ns*). Deng et al. (2025) also reported that the effect size for undergraduates (g = 0.754) was larger than K–12 students (g = 0.547, p < .001), although there was no statistical difference between the two groups. Since they estimated the effect size for K–12 students, we estimated the effect size for K–12 in this study by combining elementary school students (k = 2) and middle and high school students (k = 9) for comparison purposes. The effect size for K–12 students (k =11) in this study was large, g = 0.84, 95% CI [0.498, 1.182], p < .001. Although it is smaller than the effect size for middle and high school students (g = 0.928), it was still larger than the effect size for undergraduates (g = 0.538). However, no statistical difference was found between K–12 students and undergraduates in the study. This finding supports Deng et al.'s (2025) finding of no moderating effects of school level on learning outcomes.

In terms of the type of learning outcome related to the third research question, we found that nearly half the studies focused on the cognitive domain (51.2%) followed by the affective domain (29.0%) and the meta-cognitive domain (19.8%). These results indicate that ChatGPT has been used extensively to help learners in activities such as memorizing terminology or knowledge (Hsu, 2024; Ng et al., 2024), problem-solving (Chang et al., 2024; Lee et al., 2024; Urban et al., 2024), computer programming (Xue et al., 2024), and argumentation skills (Guo et al., 2023). However, few studies examined learning outcomes in the meta-cognitive domain, such as higher-order thinking skills (Lee et al., 2024) and critical thinking or reflection skills (Chang et al., 2024; Chang & G.-H. Hwang, 2024; Chang & G.-J. Hwang, 2024; Essel et al., 2024). Thus, more ChatGPT applications should be developed that focus on learners' meta-cognitive skills as well as in the affective domain.

As shown in Table 4, the lack of a statistical difference among the three learning domains—cognitive (g = 0.612), affective (g = 0.481) and meta-cognitive (g = 0.619)—implies that students can benefit from using ChatGPT to achieve learning outcomes regardless of the learning domain. The large number of

ChatGPT applications focusing on the cognitive domain may be due to the heavy emphasis on the cognitive domain in the curriculum (e.g., knowledge, comprehension, application, analysis, synthesis, and evaluation; Bloom, 1956). Smith and Ragan (2005) pointed out that attitudinal learning (affective domain) and psychomotor skills (psychomotor learning domain) also include substantial cognitive components, such as the knowledge structure of learning content or know-how. Notably, no publication in this meta-analysis addressed the psychomotor skill domain. Educational researchers and practitioners should pay more attention to ChatGPT applications or other generative AI for a variety of subjects in different learning domains beyond the cognitive. This balanced influence across domains resonates with Anderson and Krathwohl's (2001) taxonomy, which emphasized that achievement encompasses not only cognitive skills but also affective dispositions and metacognitive strategies.

To answer the fourth research question, we also compared the effect sizes of ChatGPT's influence on learning achievement across different instructional approaches (case-based, flipped learning, game-based, lecture, problem-based, project-based, and self-regulated learning) using subgroup analysis. The most frequently used instructional approaches with ChatGPT applications were lecture (22.1%) and self-regulated learning (16.3%). The finding that lecture was the most frequently used instructional approach indirectly supports Deng et al.'s (2025) observation that most ChatGPT intervention settings were classrooms (86.96%). The seven instructional approaches coded for analysis were grouped into instructor-led approaches (e.g., lecture) and learner-centered approaches (e.g., case-based, flipped learning, game-based, problem-based, project-based, and self-regulated learning). Excluding studies that did not indicate instructional approaches (N = 26), we found that ChatGPT was used substantially more in learner-centered approaches (68.3%) than in traditional instructor-led approaches (31.7%). Given that ChatGPT is being adopted across various instructional approaches, it is important to investigate how ChatGPT is used in each approach and to develop customized instructional strategies to maximize learning outcomes.

In terms of the moderating effect of instructional approach, there were statistical differences among the various instructional approaches in terms of the influence of ChatGPT on learning achievement. These are shown in Table 5. The effect size for self-regulated learning was the largest effect size (g = 1.115) followed by case-based learning (g = 0.836), while the smallest effect size was game-based learning (g = .092, *ns*). This research finding supports Sun and Zhou's (2024) and Steenbergen-Hu and Cooper's (2014) results. Sun and Zhou (2024) reported that generative AI's influence on college students' academic achievement is most effective in the independent learning setting (g = 0.600) compared to cooperative learning (g = 0.328). Steenbergen-Hu and Cooper (2014) reported that ITS have a large effect size for self-regulated learning (self-reliant learning or no instruction; g = 0.86), but have a small effect size for classroom instruction (instructional learning; g = 0.37) where lectures are frequently used. The plausible reasons for the statistical differences in ChatGPT applications across different instructional approaches include whether the instructional methods give learners autonomy or the amount of structure for the learning activities (e.g., well-structured problems vs. poorly-structured problems). As Rospigliosi (2023) pointed out, learners benefit from ChatGPT when they ask questions using the application or interact with it as they progress in their learning. However, if students are expected to follow predefined learning paths, they may not feel the need to use ChatGPT, and we cannot expect it to improve learning outcomes.

This is why the effect size for self-regulated learning and problem-based learning was large, while the effect size for game-based learning was small. This finding implies that instructors need to understand how students use ChatGPT in the learning process with different instructional methods and how learning activities should be designed to leverage the strengths of generative AI. These differences also resonate with self-regulated learning theory and technology acceptance perspectives (Davis, 1989; Venkatesh et al., 2003), which stress that learners' autonomy and their perceptions of usefulness critically mediate whether a technology such as ChatGPT could enhance achievement.

Regarding the last research question, most publications in this analysis adopted an experimental research design, including quasi-experimental and pretest-posttest control group studies, except for one publication (i.e., Dalgıç et al., 2024). It is promising that the research findings were derived from experimental research because experimental design studies provide stronger internal validity than correlational studies (Campbell & Stanley, 1963). We confirmed the moderating effect of research method (i.e., correlational studies, pretest-posttest control group design, mixed-design, quasi-experimental design, and posttest only design) on the influence of ChatGPT on learning achievement. See Table 6. The effect size for correlational studies (g = 1.698, p < .001) was larger than experimental design, such as pretest-posttest control group studies (g = 0.762, p < .001), and quasi-experimental studies (g = 0.529, p < .001). However, despite the large effect size for correlational studies, it should be noted that these studies, which are typically measured using self-reports such as surveys or questionnaires, carry the risk of common-method variance (Doo et al., 2023; Garger et al., 2019).

The limitations of this study have the potential to help readers better understand and interpret the results of this meta-analysis. First, this study analyzed only 22 publications, as ChatGPT was launched only 2 years before conducting this study. Given the growing academic interest in ChatGPT for educational purposes, future research will likely include a larger number of empirical studies. Second, since a meta-analysis is a secondary examination of individual experimental studies, our analysis was restricted to the data and information provided by the authors of each publication. This is an inherent limitation of a meta-analysis. Third, this study included only empirical research published in English. Our findings indicated that Asia was more productive in this area, and thus numerous studies have likely been published in local journals in Asian languages such as Chinese, Korean, or other regional languages. Future research extending this study may want to include publications in languages other than English. Finally, since the number of studies that met our inclusion criteria was limited, we did not exclude any based on quality. Future researchers are encouraged to conduct quality assessments to enhance validity.

Generative AI was not originally developed for teaching and learning. Over the past 2 years, innovators and early adopters in education have made bold attempts to integrate ChatGPT into teaching and learning with scholarly curiosity. This mirrors a recurrent pattern in the emergence of new instructional media throughout the history of educational technology (Reiser, 2001). The results of this study are promising, as the effect size for ChatGPT's influence on learning achievement was medium or moderate. The research findings give educational practitioners and researchers a solid rationale for implementing ChatGPT for educational purposes and exploring how students can learn with and from generative AI to achieve desirable learning outcomes across different instructional methods and learning environments. However, the application of ChatGPT itself does not guarantee successful learning outcomes. To fully leverage ChatGPT, educators need clear guidelines on how to integrate it as a scaffolding tool rather than a mere content provider, aligning with instructional objectives and learners' needs. Policymakers should also consider differentiated strategies across contexts, given the varying effects across educational levels and instructional approaches. To develop practical and effective guidelines for ChatGPT applications in education, more empirical studies are needed to investigate the influence of ChatGPT on learning achievement in a variety of educational settings.

Declarations

- Funding: The authors declare that they have no funding source for this research.
- Ethical approval and informed consent statements: Not applicable.
- Conflict of interest: There were no conflicts of interest in this research.
- Data availability: The data used and/or analyzed in the current study are available from the corresponding author upon reasonable request.

References

Alneyadi, S., & Wardat, Y. (2023). ChatGPT: Revolutionizing student achievement in the electronic magnetism unit for eleventh-grade students in Emirates schools. *Contemporary Educational Technology, 15*(4), Article ep448. https://doi.org/10.30935/cedtech/13417

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman, Inc.

Becker, E., Wünsche, J., Veith, J. M., Schrader, J., Bitzenbauer, P. (2025). *From cognitive relief to affective engagement: An empirical comparison of AI chatbots and instructional scaffolding in physics education*. arXiv. https://doi.org/10.48550/arXiv.2508.06254

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: David McKay.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education*, 195, Article 104711. https://doi.org/10.1016/j.compedu.2022.104711

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand-McNally.

Chang, C.-C., & Hwang, G.-H. (2024). Promoting professional trainers' teaching and feedback competences with ChatGPT: A question-exploration-evaluation training mode. *Educational Technology & Society*, *27*(2), 405–421. https://doi.org/10.30191/ETS.202404_27(2).TP06

Chang, C.-C., & Hwang, G.-J. (2024). ChatGPT facilitated professional development: Evidence from professional trainers' learning achievements, self-worth, and self-confidence. *Interactive Learning Environments*, *33*(1), 883–900. https://doi.org/10.1080/10494820.2024.2362798

Chang, C.-Y., Yang, C.-L., Jen, H.-J., Ogata, H., & Hwang, G.-H. (2024). Facilitating nursing and health education by incorporating ChatGPT into learning designs. *Educational Technology & Society*, *27*(1), 215–230. https://doi.org/10.30191/ETS.202401_27(1).TP02

Chen, Y.-C., & Hou, H.-T. (2024). A mobile contextualized educational game framework with ChatGPT interactive scaffolding for employee ethics training. *Journal of Educational Computing Research*, *62*(7), 1737–1762. https://doi.org/10.1177/07356331241268505

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review, 36*, Article 100452. https://doi.org/10.1016/j.edurev.2022.100452

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Council of Europe. (2022). *Artificial intelligence and education.* https://www.coe.int/en/web/education/artificial-intelligence

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education, 20,* Article 22. https://doi.org/10.1186/s41239-023-00392-8

Dalgıç, A., Yaşar, E., & Demir, M. (2024). ChatGPT and learning outcomes in tourism education: The role of digital literacy and individualized learning. *Journal of Hospitality, Leisure, Sport & Tourism Education, 34,* Article 100481. https://doi.org/10.1016/j.jhlste.2024.100481

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education, 227,* Article 105224. https://doi.org/10.1016/j.compedu.2024.105224.

Díaz, B., & Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education, 217,* Article 105071. https://doi.org/10.1016/j.compedu.2024.105071

Dikilitaş, K., Klippen, M. I. F., & Keles, S. (2024). A systematic rapid review of empirical research on students' use of ChatGPT in higher education. *Nordic Journal of Systematic Reviews in Education, 2,* 103–125.

Doo, M. Y., Zhu, M., & Bonk, C. J. (2023). MOOC learners' self-directed learning and learning outcomes: A meta-analysis. *Distance Education, 44*(1), 86-105. https://doi.org/10.1080/01587919.2022.2155618

Driscoll, M. P. (1993). *Psychology of learning for instruction* (2nd ed.). Allyn & Bacon.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). *Bias in meta-analysis detected by a simple, graphical test.* BMJ (Clinical Research Ed.), *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

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 and Education: Artificial Intelligence*, *6*, Article 100198. https://doi.org/10.1016/j.caeai.2023.100198

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Gagne, R. M. & Briggs, L. J. (1984). *Principles of instructional design*. Holt, Rinehart, and Winston.

Garcia-Varela, F., Beckerman, Z., Nussbuam, M., & Mendoza, M. (2025). Reducing interpretative ambiguity in an educational environment with ChatGPT. *Computers & Education, 225,* Article 105182. https://doi.org/10.1016/j.compedu.2024.105182

Garger, J., Jacques, P. H., Gastle, B. W., & Connolly, C. M. (2019). Threats of common method variance in student assessment of instruction instruments. *Higher Education Evaluation and Development, 13*(1), 2–17. https://doi.org/10.1108/HEED-05-2018-0012

Gunawardena, M., Bishop, P., & Aviruppola, K. (2024). Personalized learning: The simple, the complicated, the complex and the chaotic. *Teaching and Teacher Education, 139*, Article 104429. https://doi.org/10.1016/j.tate.2023.104429

Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2023). Effects of chatbot-assisted in-class debates on students' argumentation skills and task motivation. *Computers & Education*, *203,* Article 104862. https://doi.org/10.1016/j.compedu.2023.104862

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Hsu, M.-H. (2024). Mastering medical terminology with ChatGPT and Termbot. *Health Education Journal*, *83*(4), 352–358. https://doi.org/10.1177/00178969231197371

Huesca, G., Martínez-Trevino, Y., Molina-Espinosa, J. M., Sanromán-Calleros, A. R., Martínez-Román, R., Cendejas-Castro, E., & Bustos, R. (2024). Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. *Educational Science, 14,* Article 660. https://doi.org/10.3390/educsci14060660

Jeon, J., Lee, S. & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers and Education, 206,* Article 104898. https://doi.org/10.1016/j.compedu.2023.104898

Lee, H.-Y., Chen, P.-H., Wang, W.-S., Huang, Y.-M., & Wu, T.-T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education, 21,* Article 16. https://doi.org/10.1186/s41239-024-00447-4

Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning, 28*(3), 82–106. https://doi.org/10.24059/olj.v28i3.4593

Li, H.-F. (2023). Effects of a ChatGPT-based flipped learning guiding approach on learners' courseware project performances and perceptions. *Australasian Journal of Educational Technology*, *39*(5), 40–58. https://doi.org/10.14742/ajet.8923

Liang, H.-Y., Hwang, G.-J., Hsu, T.-Y., & Yeh, J.-Y. (2024). Effect of an AI-based chatbot on students' learning performance in alternate reality game-based museum learning. *British Journal of Educational Technology, 55*(5). 2315–2338. https://doi.org/10.1111/bjet.13448

Liao, X., Zhang, X., Wang, Z., & Luo, H. (2024). Design and implementation of an AI-enabled visual report tool as formative assessment to promote learning achievement and self-regulated learning: An experimental study. *British Journal of Educational Technology, 55*(3), 1253–1276. https://doi.org/10.1111/bjet.13424

Lo, C. K., Hew, K. F., & Jong, M. S.-y. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education, 219*, Article 105100. https://doi.org/10.1016/j.compedu.2024.105100

Martin, A. J., Collie, R. J., Kennett, R., Liu, D., Ginns, P., Sudimantara, L. B., ... & Rüschenpöhler, L. G. (2025). Integrating generative AI and load reduction instruction to individualize and optimize students' learning. *Learning and Individual Differences*, 121, 102723. https://doi.org/10.1016/j.lindif.2025.102723

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*(4), 1328–1353. https://doi.org/10.1111/bjet.13454

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review, 4,* 251–277. https://doi.org/10.1007/BF00168958

Patac, L. P., & Patac Jr, A. V. (2025). Using ChatGPT for academic support: Managing cognitive load and enhancing learning efficiency–A phenomenological approach. *Social Sciences & Humanities Open*, 11, 101301. https://doi.org/10.1016/j.ssaho.2025.101301

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*, Article 22. https://doi.org/10.1186/s41039-017-0062-8

PRISMA. (2020). *PRISMA flow diagram*. https://www.prisma-statement.org/prisma-2020-flow-diagram

Reiser, R. A. (2001). A history of instructional design and technology: Part I: A history of instructional media. *Educational Technology Research and Development, 49*(1), 53–64. https://doi.org/10.1007/BF02504506

Reuters. (2024, August 30). *OpenAI says ChatGPT's weekly users have grown to 200 million*. https://www.reuters.com/technology/artificial-intelligence/openai-says-chatgpts-weekly-users-have-grown-200-million-2024-08-29/

Rospigliosi, P. (2023). Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT? *Interactive Learning Environments*, *31*(1), 1–3. https://doi.org/10.1080/10494820.2023.2180191

Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and artificial intelligence in higher education: Quick start guide*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000385146

Saettler, P. (1990). *The evolution of American educational technology*. Libraries Unlimited.

Schoenherr, J., Strohmaier, A. R, & Schukajlow, S. (2024). Learning with visualizations helps: A meta-analysis of visualization intervention in mathematics education. *Educational Research Review, 45*, Article 100639. https://doi.org/10.1016/j.edurev.2024.100639

Smith, P. L., & Ragan, T. J. (2005). *Instructional design* (3rd ed.). Wiley.

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331–347. https://doi.org/10.1037/a0034752

Sun, L., & Zhou, L. (2024). Does generative artificial intelligence improve the academic achievement of colleges? A meta-analysis. *Journal of Educational Computing Research*, *62*(7), 1673-1713. https://doi.org/10.1177/07356331241277937

Svendsen, K., Askar, M., Umer, D., & Halvorsen, K. H. (2024). Short-term learning effect of ChatGPT on pharmacy students' learning. *Exploratory Research in Clinical and Social Pharmacy*, *15,* Article 100478. https://doi.org/10.1016/j.rcsop.2024.100478

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

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers & Education: Artificial Intelligence, 7*, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270

Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education, 215,* Article 105031. https://doi.org/10.1016/j.compedu.2024.105031

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

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

Warnick, B. R., & Burbules, N. C. (2007). Media comparison studies: Problems and possibilities. *Teachers College Record*, *109*(11), 2483–2510. https://eric.ed.gov/?id=EJ820500

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology, 55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Xue, Y., Chen, H., Bai, G. R., Tairas, R., & Huang, Y. (2024). Does ChatGPT help with introductory programming? An experiment of using students ChatGPT in CS1. In A. Paiva, R. Abreu, K. Gama, & J. Siegmund (Chairs), *ICSE-SEET '24: Proceedings of the 46th international conference on software engineering: Software engineering education and training* (pp. 331–341). ACM. https://doi.org/10.1145/3639474.3640076

Yilmaz, R., & Yilmaz. F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers & Education: Artificial Intelligence, 4*, Article 100147. https://doi.org/10.1016/j.caeai.2023.100147

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research and future directions. *Computers and Education: Artificial Intelligence, 2,* Article 100025. https://doi.org/10.1016/j.caeai.2021.100025

Zhou, W., & Kim, Y. (2024). Innovative music education: An empirical assessment of ChatGPT-4's impact on student learning experiences. *Education and Information Technology, 29,* 20855–2088. https://doi.org/10.1007/s10639-024-12705-z

Appendix

Overview of Selected Papers

Author(s) & Year	Participants	Instructional approach	Learning discipline	Domain of learning outcomes		Research design
Alneyadi & Wardat (2023)	K-12 students (Middle and high)	Lecture	Science	C	Cognitive achievement (knowing, applying, reasoning)	Mixed-methods
Chang & H. Hwang (2024)	Clinical teacher	Problem based learning	Health	C M	Learning achievement Critical thinking	Quasi-experimental
Chang & J. Hwang (2024)	Clinical teachers	Case-based learning	Education	C M A	Test scores, Critical thinking Self-worth, self-confidence	Quasi-experimental
Chang et al. (2024)	Adult learners	Problem based learning	Health	C M A	Problem solving Critical thinking Learning enjoyment	Quasi-experimental
Chen & Hou (2024)	Adult learners	Game-based learning	Ethics	C A	Learning achievement Flow antecedent, motivation, anxiety, flow experience, flow	Quasi-experimental
Dalgıç et al. (2024)	Undergraduates	Not reported	Tourism	C	Digital literacy, individualized learning	Correlational
Essel et al. (2024)	Undergraduates	Lecture	Research	M	Critical thinking skills, creative thinking, reflection, thinking skills	Mixed-methods
Fan et al. (2024)	Undergraduates	Not reported	Language	A	Interests/enjoyment, perceived competence, efforts, pressure/tension	Post only
Guo et al. (2023)	Undergraduates	Not reported	Language	C A	Argumentation skills Task motivation	Quasi-experimental
Huesca et al. (2024)	Undergraduates	Flipped learning	Computer science	C	Learning gain in computer science	Experimental (Pre-Post test)
Hsu (2024)	Undergraduates	Lecture	Medical	C	Learning achievement	Quasi-experimental

Author(s) & Year	Participants	Instructional approach	Learning discipline	Domain of learning outcomes		Research design
Lee et al. (2024)	Undergraduates	Self-regulated learning	Science	C	Knowledge construction Problem-solving	Experimental (Pre-Post test)
				M	Critical thinking, creativity	
Li (2023)	Undergraduates	Flipped learning	Education	C	Project performance	Quasi-experimental
				A	Learning motivation	
Liang et al. (2024)	K-12 students (Elementary)	Game-based learning	Interdisciplinary learning	C	Learning achievement	Quasi-experimental
				A	Emotional engagement	
Liao et al. (2024)	K-12 students (Middle and high)	Lecture	Biology	C	Test-scores	Quasi-experimental
				M	Cognitive strategies	
				A	Test anxiety, self-efficacy	
Ng et al. (2024	K-12 students (Middle and high)	Self-regulated learning	Science	C	Science knowledge	Quasi-experimental
				A	Motivation	
Svendsen et al. (2024)	Undergraduates	Self-regulated learning	Pharmacy	C	Knowledge test	Experimental (Pre-Post test)
Teng (2024)	Undergraduates	Self-regulated learning	Language	A	Learning motivation, learning engagement	Experimental (Pre-Post test)
Urban et al. (2024)	Undergraduates	Problem based learning	Creative thinking	C	Problem-solving	Experimental (Pre-Post test)
				M	Self-evaluation, mental effort	
				A	Self-efficacy, making the task interesting	
Xue et al. (2024)	Undergraduates	Project based learning	Computer science	C	UML diagram, Java Programing, Post-evaluation scores	Post test only
Yilmaz & Yilmaz (2023)	Undergraduates	Project based learning	Computer science	C	Computational thinking	Experimental (Pre-Post test)
				A	Self-efficacy, motivation	
Zhou & Kim (2024)	Undergraduates	Lecture	Music	C	Music knowledge	Quasi-experimental

Note. C = cognitive; A = affective; M = meta-cognitive; UML = unified modeling language