Computers & Education 250 (2026) 105640

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Can asynchronous learning modules about generative AI increase students' knowledge, skills, and self-efficacy for genAI?

Allison E. Connell Pensky [a,*], Lydia E. Eckstein [a], Michael C. Melville [a], Laura O. Pottmeyer [a], Zach Mineroff [a,b], Avi Chawla [a], Judy Brooks [a,b], Chad Hershock [a], Marsha C. Lovett [b]

[a] Eberly Center for Teaching Excellence & Educational Innovation, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA
[b] Office of the Vice Provost for Teaching and Learning Innovation, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, USA

ARTICLE INFO

Keywords:
Generative AI competency
Online learning modules
SoTL
Assessment

ABSTRACT

The goal of the present large-scale experiment was to test if a 90-min, open access set of online learning modules with embedded practice and feedback would enhance students' generative artificial intelligence (genAI) competency. We also examined whether impacts were experienced equitably by all included student demographic groups. Participants were 1368 undergraduate and graduate students enrolled in 65 course sections. Sections were randomly assigned to treatment (engagement with the learning modules between pre- and post-tests) or control (no learning modules between assessments). Generative AI competency was defined as knowledge, skills, and attitudes and was measured at pre and post. The learning modules significantly improved treatment students' genAI knowledge, prompt engineering skills, fact- and source-checking skills, and self-efficacy. There was a null effect of the modules on the skill of critically evaluating potential bias in genAI's output. Improvements in genAI competency were equitable across all student groups (birth sex, race/ethnicity, student level, first-generation status, and academic discipline). Such readily-available competency trainings are important for supporting the emerging necessity of knowing when and how to work with genAI in one's education and future career. Limitations, implications, and future directions are discussed.

1. Introduction

The release of generative artificial intelligence (genAI) on the heels of the COVID-19 pandemic signified another momentous disruption for education. Unlike the unprecedented and uniformly challenging nature of teaching and learning in a pandemic, however, the arrival of large language models (LLMs), and genAI more broadly, was met by educators with a mixture of optimism and trepidation. Some recognized the potential of genAI, and especially chatbots like ChatGPT, highlighting their wide availability, scalability, and efficiency and argued that genAI tools could be powerfully leveraged by students to get additional customized explanations, feedback, and practice (e.g., Kalla et al., 2023; Kasneci et al., 2023; Lee et al., 2024; Sabzalieva & Valentini, 2023; Wu et al., 2025). Additionally, educators and scholars saw the potential of genAI to "level the playing field", especially for second-language learners (e.g., Young and Shishido, 2023), and students with heterogeneous backgrounds or traditionally underrepresented groups

* Corresponding author.
E-mail address: aconnel2@andrew.cmu.edu (A.E. Connell Pensky).

https://doi.org/10.1016/j.compedu.2026.105640
Received 11 November 2025; Received in revised form 6 March 2026; Accepted 12 April 2026
Available online 14 April 2026
0360-1315/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

(see James & Andrews, 2024 for more resources and details). Similarly, some scholars argued that genAI tools could offer a significant reduction in workload for administrators and instructors alike. Instructors could use genAI for anything from assisting with lesson plan design, the creation of exam questions, in-class activities, and rubrics, to assisting instructors with grading (e.g., Bowen & Watson, 2024). Others, however, raised alarms about the potentially negative effects of genAI use on the development of students' skills and cognitive abilities, especially writing and critical thinking (e.g., Warschauer et al., 2023). Recognizing both the affordances and limitations of genAI, nuanced discussions have begun to emerge about the potential for genAI to enhance learning, for instance through the design of thoughtful assignments (e.g., Mollick & Mollick, 2022).

Potential plagiarism and academic integrity violations are some of the most frequently vocalized concerns in the context of genAI use in education. Indeed, surveys reveal that students' potentially inappropriate use of genAI tools is widespread and that instructors perceive a significant uptick in academic dishonesty (e.g., Freeman, 2024; Watson & Rainie, 2025). The introduction of ChatGPT was soon followed by the release of ZeroGPT, a ChatGPT detector. However, such genAI detection tools have themselves proven problematic and controversial for inaccuracies, including false positives, claiming that a student has submitted AI-generated text when they have not, that may disproportionately affect second-language learners (Ardito, 2025).

Almost three years into the wide public release of OpenAI's ChatGPT, many institutions of higher education have embraced, leveraged, and even encouraged the use of genAI in education (Moorhouse et al., 2023). A recent analysis of 116 research-focused universities in the United States found that many of these institutions offer guidance to instructors on how to incorporate genAI in classrooms and to students by way of syllabus statements and institutional policies, especially around writing (McDonald et al., 2025; Watson & Rainie, 2025). In other words, early genAI resources have focused on guidance for faculty (e.g., Bowen & Watson, 2024) and anti-cheating and plagiarism policies for students (Watson & Rainie, 2025), but education around effective, critical, and ethical genAI use by students has largely been missing (also see Crawford et al., 2023).

1.1. The need for genAI competency development for college and beyond

Teaching students to effectively and ethically use genAI creates multiple challenges and needs for higher education. Although students who have entered college in the last 20 years are often referred to as "digital natives" (Prensky, 2001), assumed to have been raised in a digital environment and therefore inherently more familiar, comfortable, and adept with new technologies such as genAI, such sweeping proclamations have been challenged (e.g., Kennedy et al., 2008). Indeed, early research on genAI use among this generation (students and workers born after 1980) also indicates that generational assumptions about digital literacy automatically enhancing genAI literacy may be overly hasty and erroneous. For example, in a graduate employability report by the Cengage Group (2024), 61% of Generation Z students (i.e., students born between 1997 and 2012) said they felt unprepared to use genAI. Indeed, access to innovative technologies is often related to income level and socioeconomic status (Rogers, 1983) and some research has found income to predict genAI self-efficacy and intention to use genAI (Hong, 2022). Ultimately, the cost of genAI tools or their more powerful versions may widen the digital divide between high- and low-income students (Capraro et al., 2024; Davidoff, 2024). Gender biases and sociocultural norms further disadvantage women and girls and their access to the digital world, contributing to a digital gender divide (OECD, 2018). Depending on students' backgrounds and their exposure to both tools and training, they will therefore enter college with varying degrees of comfort and familiarity with different tools.

1.1.1. Defining genAI competency

Being prepared to use AI is often referred to as AI literacy, which Long and Magerko (2020) define as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (p.2). Ng and colleagues (2021) reviewed relevant literature and suggested four dimensions of AI literacy: (1) knowing and understanding AI, (2) using and applying AI, (3) critically evaluating AI technologies, and (4) awareness of ethical issues arising from AI use. In the context of generative AI, this may encompass skills like writing effective prompts ("prompt engineering"), critically evaluating AI-generated output, being aware of contextually-appropriate use and ethical implications, as well as limitations of genAI. For example, genAI has been known to make mistakes and produce biased or misleading content. GenAI systems may, for example, generate outputs that look factual, but are blatantly false, which are often referred to as "hallucinations". Students, and other informed users of genAI, need to be aware of these risks (e.g., Weidinger et al., 2022).

In the context of genAI skill building, genAI literacy and competency are frequently used terms. In this paper, we build on previous definitions of AI literacy and use *genAI competency* in line with the definition Chiu and colleagues (2024) have offered, namely that competency goes beyond literacy in that it encompasses knowledge, ability, and attitudes (also see Gimpel et al., 2024). Importantly, the idea of competency involving knowledge, skills, and attitude development is distinct from the knowledge-skills-attitudes framework literature in which researchers have examined the impact *of using* genAI tools on one or more educational outcomes (see Lee et al., 2025 for a review). One relevant attitude for competency development is confidence in applying one's knowledge and skills ethically and responsibly. This task-specific confidence in applying genAI knowledge and skills describes self-efficacy (Bandura, 1997) newly applied to the realm of genAI.

Self-efficacy refers to a person's belief in their ability to achieve a desired outcome or to perform a task (Bandura, 1997). Self-efficacy has been found to be an important construct in educational research with many scholars demonstrating its positive association with key academic outcomes such as greater engagement in learning (Bresó et al., 2011), better persistence (Torres & Solberg, 2001), and higher performance (Bresó et al., 2011). Talsma et al. (2018) conducted a systematic review and meta-analysis to better understand the relationship between self-efficacy and academic performance, specifically investigating whether the relationship is uni- or bi-directional in nature. Using a cross-lagged path analysis they found the relationship to be bi-directional with self-efficacy

having a positive effect on performance and vice versa. However, other research has suggested that there can be mismatches between self-efficacy and performance, i.e., students may be over- or under-confident relative to their performance (e.g., Talsma et al., 2019; Whitcomb et al., 2020).

Mastery experiences have often been found to be the most powerful tool for the development of self-efficacy (e.g., Usher & Pajares, 2009). In other words, by practicing a skill and, ideally, experiencing success, one increases their self-efficacy for that skill. However, in the context of genAI, it is unclear whether the experience of simply prompting and receiving responses from the tool (even erroneous or poor ones) may suffice to build self-efficacy. Studies measuring self-efficacy for genAI are sparse (though see Arslankara & Usta, 2024; Morales-García et al., 2024 for recent approaches). It is likely that, to build genAI self-efficacy, students will require more structured feedback on the effectiveness of their prompts, how to assess the tool's output, and considerations for appropriate use of genAI in order to experience positive impacts on their developing self-efficacy. It is well documented that such scaffolded instruction increases both efficacy and performance in other domains (e.g., Grothérus et al., 2019). The present study offers a way to measure self-efficacy for genAI knowledge and skills. To our knowledge, it is the first study of genAI competency to do so.

1.1.2. Why does genAI competency matter and where can students develop it?

Building genAI competency is not just an educational need, but has a real impact on students' future employability. The 2023 Future of Jobs report by the World Economic Forum (Di Battista et al., 2023) lists genAI as a technology most likely to drive industry transformation and Sue Duke of LinkedIn recently argued that AI literacy should be a major focus of workforce skill development (Duke, 2025). The Cengage Group (2024) also found that the majority of recent graduates surveyed (70%) believed that higher education has an obligation to teach genAI literacy while at the same time 55% said their degree programs did not do this, leaving them less prepared for the workforce. The same report states that 62% of employers now expect some degree of genAI literacy and lacking those skills risks putting employees and applicants at a disadvantage in the job market (Cengage Group, 2024). Given that most individuals entering the non-manual labor workforce will seek some amount of college education, institutes of higher education are not only an ideal place for developing these genAI skills and students' confidence in them, but arguably have the emerging responsibility to foster this genAI competency.

As such, it is incumbent on instructors and institutions of higher education to educate students on LLMs, their limitations, and ethical use in order to develop appropriate genAI literacy and competency (e.g., Gimpel et al., 2024; Kong et al., 2025; Yan et al., 2024; Zhang et al., 2024). Achieving this goal requires the ensuing and ongoing revisions of institutional policies, pedagogies, syllabi, lesson plans, assessments, and plagiarism detection. This has become a significant workload and curricular concern for administrators and faculty (Kohnke et al., 2024; Watson & Rainie, 2025). The implications for faculty are clear: After the rapid and unprecedented pivots to remote and hybrid instruction during the COVID years, they are now asked to pivot and adapt again to redesign programs, courses, and assignments to foster genAI literacy, facilitating "a fatigue from the continual adaptation to a new normal" (Concannon et al., 2023, p. 1). While many instructors have degrees of familiarity with genAI tools, in early 2024 only 14% of 2654 instructors surveyed felt confident in their ability to integrate genAI into their courses and most instructors indicated wanting some instructional support (Ruediger et al., 2024). The implications for students are clear, as well: Their genAI skill development may largely depend on the courses they choose, both in terms of disciplines and instructors' readiness (and willingness) to teach genAI skills. In other words, students completing courses with educators who are not prepared to teach them about genAI use do not have the educational opportunity to develop their genAI competency, even if they are using the tools on their own.

Realizing that instructors may feel inadequately prepared, some institutions have begun to invest significant funds into personnel, software, and hardware for genAI infrastructure and training (Watson & Rainie, 2025), including programming through university Centers for Teaching and Learning (CTLs, e.g., "The Generative AI Teaching As Research (GAITAR) Initiative" at Carnegie Mellon University, 2026, and "Generative AI in Teaching and Learning" at the University of Virginia, 2025). Another way to develop students' baseline genAI proficiencies is for institutions to prepare and deliver a targeted learning opportunity that is broadly disseminated to all students outside of their coursework.

1.2. Extant approaches to genAI competency development

Making genAI competency training as widely available to students as possible suggests a need for approaches that are portable, accessible, free, discipline-agnostic, and not reliant on the capacities of individual instructors to develop and include genAI-related education and assignments in their courses. Some approaches have been studied by educators, educational developers, and educational researchers affiliated with university CTLs and libraries, although not all have been focused on generative AI specifically, given that there is still a relative dearth of literature on genAI. For example, even before the broad release of *generative* AI, Kong et al. (2021) developed an AI literacy course for students of all disciplinary backgrounds. One hundred and twenty university students in Hong Kong enrolled in their three-week training. In addition to asynchronous readings about artificial intelligence and machine learning, students engaged with supervised and unsupervised learning, and attended two 90-minute workshops. Students' total time commitment was approximately seven hours across three weeks. In order to assess this training program, students completed AI concept tests, self-reported their AI literacy, and self-reported their level of AI empowerment at pre and at post. All three measures increased from pre to post, suggesting the course may have helped students' understanding of AI concepts. However, the study did not include a comparison group of students who did not complete the training, so any changes observed cannot conclusively be attributed to the literacy training. Additionally, students' skills navigating genAI (e.g., prompt engineering) were not assessed, so it is unclear if students' knowledge and belief in their skills also translated into observable behavior and therefore the kinds of skills employers ultimately seek.

In another study that also focused on AI literacy and directly built on Kong et al. (2021), two small samples of students in an

international university course (total n = 29 students from 13 countries across Europe and Asia) were exposed to five hours of content across two online workshops in 2021 and 2022 (Korte et al., 2024). The workshops were interactive and participatory in nature and focused on teaching students about machine learning. Students were expected to do some reading ahead of each workshop and they completed the same concept inventory and self-perceived AI literacy survey as in Kong et al. (2021) before the first (pre) and after the last (post) workshop. Although the authors replicated the promising increase in conceptual understanding and self-perceived literacy found by Kong et al. (2021), their study design also lacked a control group and direct measures of students' skills.

In a slightly different approach and one that explicitly focused on the development of *generative* AI literacy, a collaboration between the Center for Teaching and Learning and Library Services at a mid-sized Australian university offered a single 90-minute workshop for students who could choose between remote and in-person formats (Sullivan et al., 2024). Workshop content covered concepts of genAI, prompt engineering, and appropriate genAI use in the context of assignments. Data were collected from 171 students across four workshops and included pre-post measures of self-assessments of genAI use, genAI confidence, and a post-workshop skill assessment in the form of evaluation of genAI output. Results indicated that students' confidence in using genAI and understanding of the university's policies increased significantly from pre to post. Additionally, students' reported genAI use shifted from primarily use in non-academic settings (e.g., for recipes and advice) to more academic use (e.g., brainstorming, scheduling, research). While this study also lacked a control group for comparison, it did include a direct skill measure. The post-only assessment of students' skills evaluating genAI output suggested awareness by students to fact-check and source-check output.

Together, these studies suggest promising approaches to genAI literacy development using interventions that do not rely on individual instructors and do not require a full semester of effort. However, all three studies (some of which did not target *generative* AI specifically) had relatively small sample sizes of self-selected students, possibly because, to our best knowledge, all workshops required some level of synchronous attendance either remotely or in-person. Additionally, none of the above studies tracked changes in knowledge, skills, and attitudes among students *not* exposed to these workshops. Given the consistent lack of a comparison group, we cannot conclusively attribute changes in outcomes to the workshops. Lastly, student skill *development* was not assessed in any of the aforementioned papers. Given that this is a nascent area of inquiry, research for building genAI competency would benefit from a rigorous large-scale study that assesses knowledge, skills, and self-efficacy using an experimental research design with random assignment to comparison and treatment groups.

1.3. The present study

Our study builds on the early work on genAI competency development in a number of ways. We sought to extend (1) the breadth of genAI training content, (2) the assessment of learning of genAI competencies by including pre-post measures of knowledge, skills, self-efficacy, (3) the size and composition of the sample to include undergraduate and graduate students, and (4) the ability to draw causal conclusions by including a control group that completed the same pre-post measures but did not access the training in between. Additionally, access to student demographic data permitted us to disaggregate results across student populations and backgrounds.

More specifically, we developed a genAI training that students of any discipline could complete in a relatively short (< 2 hours) time in a fully asynchronous, online, self-paced learning experience. By enabling the pre-packaged training to be imported into existing courses and delivered through the university's course management system outside of class time, we mitigated concerns about instructor workload in content creation and delivery as well as use of class time for this non-disciplinary training.

We developed the training in collaboration with faculty and student genAI subject matter experts from across the university. The training modules included (a) a non-technical introduction to LLMs, (b) ethical implications of genAI in education, (c) opportunities and risks for students when using genAI, and (d) a detailed discussion of responsible use. We designed the training modules following established e-learning design principles and best practices of online education (Clark & Mayer, 2016). Specifically, we broke the overall content into several manageable units to manage cognitive load (Sweller, 2010). We introduced key concepts before exploring them more in depth ("pre-training principle"), utilized graphics in addition to text ("multimedia principle"), and connected new material to students' existing knowledge, e.g., on web searches (Clark & Mayer, 2016; Lovett et al., 2008; Schwartz et al., 2016). We further embedded several worked examples (e.g., on how to build an effective prompt) and active learning practices with immediate, explanatory feedback ("learn-by-doing exercises") into the content delivery (Clark & Mayer, 2016; Lovett et al., 2008). For a more detailed overview of the guiding design principles and example learn-by-doing exercises, please see Supplementary Materials.

In order to conduct this experiment at scale, study a broadly representative sample of the student population at our university, and prevent self-selection into the training at the individual student level, we recruited a large number of instructors (n = 46) of graduate and undergraduate courses (n = 53 unique courses across n = 65 sections) from all colleges at a midsize research-focused university in the Midwestern United States. Participating instructors agreed to assign the training modules (at various timepoints, depending on condition) to all students in their selected course(s) as part of their course's learning experience.

Further addressing a limitation of extant research, we randomly assigned courses to treatment or control conditions. Students in both conditions completed pre and post measures at the same timepoints, but only students assigned to the treatment courses had access to the genAI training modules in between those two measurements.

We further extended existing work by assessing not only genAI knowledge, but also by asking students to demonstrate prompt-engineering and evaluation-of-output skills and by assessing students' genAI self-efficacy at pre and post. By doing so, we were able to measure not only performance (a one-time assessment), but learning (Lovett et al., 2023) and how it was affected by exposure to the training.

In sum, our study design tested the primary research question: Can we develop genAI competency (as measured by knowledge, skills, and self-efficacy) at scale with a relatively short, asynchronous, self-paced, and discipline-agnostic online intervention? We

hypothesized that completing the genAI training modules would increase students' genAI competency, as measured by pre- and post-tests, compared to a control group. Additionally, we tested an exploratory research question regarding the extent to which different student populations experienced equitable genAI competency gains.

2. Method

2.1. Participants

Participants included students from courses taught by 46 unique instructors at a mid-sized private university in the Northeastern United States. To be considered for inclusion, instructors responded to an open call for this study and needed to meet certain eligibility criteria, like being the instructor of record, having a minimum enrollment of 15 students, being able to attend an onboarding seminar the week before the semester, and agreeing to maintain implementation fidelity. These instructors enlisted 53 unique courses. Some instructors taught multiple sections of the same course, for a total of 65 sections that were included in the study. Participating courses spanned small (fewer than 20 enrolled students), medium (20-50 students) and large courses (more than 50 enrolled students) at both the undergraduate and graduate levels. In total, we invited n = 2181 students across 38 undergraduate, 26 graduate, and 1 mixed undergraduate-graduate courses to participate in the study on the first day of classes. Of the students who were invited, n = 1964 provided at least one data point. Of these, 109 students who were enrolled in more than one involved course were removed from analysis. The final sample consisted of n = 1368 students[1] (control n = 610, treatment n = 758) who consented and completed both the pre-test and the post-test (see Table 1 for demographics). Students in the treatment condition needed to have accessed the final screen of the genAI learning modules (sections within the modules were unlocked sequentially) and to have completed at least one embedded learn-by-doing activity to demonstrate minimal engagement with the intervention.

2.2. Materials & Measures

2.2.1. Condition assignment

We randomly assigned half of all course sections to treatment (n = 33 course sections) and half to control conditions (n = 32 course sections), and then made adjustments so that multiple sections taught by the same instructor received the same condition to minimize risks to implementation fidelity.

2.2.2. GenAI learning modules

The intervention[2] consisted of four locally-designed modules that were embedded in the university's learning management system (LMS) for asynchronous and self-guided learning that instructors incorporated into existing courses. All four modules consisted of a mix of text and video information, interspersed with learn-by-doing activities that provided instant feedback to learners. Learn-by-doing exercises were scaled to complexity such that lower-level Bloom's ("Remember & Understand"; Anderson & Krathwohl, 2001) were practiced with multiple-choice concept questions (see Fig. 1), while "Evaluate" was practiced with critique and judgement tasks, and "Create" was practiced with authentic tasks. Embedded practice also included metacognitive reflection questions.

The modules aimed to help students develop basic fluency with genAI tools and consider the opportunities and risks associated with genAI tools in the context of learning. Specifically, the four modules covered the basics of how genAI tools work (Module 1), ethical implications and concerns (Module 2), opportunities and risks for students (Module 3), and appropriate and responsible use (Module 4).

The sequencing of content was intentional, moving learners from foundational definitions and broad conceptual understanding to guided consideration, applied use in their role as learners, and ultimately to responsible use (e.g., after considering academic integrity). This approach was intended to ensure that learners not only know *what* and *how*, but also *when* and *why* to apply this knowledge (e.g., as communicated in their course syllabi).

Given the asynchronous format, we gated the content so that students needed to go through the modules in a predetermined order. They needed to complete one before the next one unlocked. Thoughtful completion of the modules and embedded learning activities was estimated to take 90-120 minutes in total, based on user testing.

2.2.3. Pre and post assessment

Learning objectives (LOs), which were shared with students, targeted knowledge about genAI and responsible use in an educational setting, prompt-engineering and critical-analysis-of-output skills, and self-efficacy (see Supplementary Material). The assessment evaluated genAI knowledge, genAI skills, and genAI self-efficacy. Pre and post assessments involved a combination of items (i.e., multiple-choice questions, matching tasks, open-ended authentic tasks, and quantitative self-efficacy items). We authored each

[1] There was no evidence of differential attrition across demographic variables, i.e., the percentage of the n = 596 students who were ultimately dropped from analyses was similar to the percentage of students who were retained across all sample demographic variables (sex, race and ethnicity, first-generation status, disciplinary background, first-year undergraduate vs. all other student levels).

[2] All modules used in this study have been updated based on the results of this study and other considerations. They have been formatted to be accessible to interested educators and institutions outside of our LMS. You can access these materials and request assistance here: https://proton.oli.cmu.edu/sections/join/rvnsr.

Table 1
Sample demographics based on university classifications within registrar data.

	Control (n = 610)	Treatment (n = 758)
Age (years)		
Mean (Standard Deviation)	21.4 (3.6 years)	20.7 (3.2 years)
Sex assigned at birth ('sex')		
Female	323 (53%)	329 (43.4%)
Male	285 (46.7%)	424 (55.9%)
No information	2 (.3%)	5 (.7%)
Race and Ethnicity		
Asian	307 (50.3%)	342 (45.1%)
Black	22 (3.6%)	21 (2.8%)
Hispanic	7 (1.1%)	13 (1.7%)
Pacific Islander	1 (.2%)	1 (.1%)
White	93 (15.2%)	133 (17.5%)
Bi-racial/multi-racial	58 (9.5%)	91 (12%)
No information	120 (19.7%)	152 (20.1%)
Student level		
First-year undergraduate	109 (17.9%)	151 (19.9%)
Second-, third-, or fourth-year undergraduate	284 (46.6%)	427 (56.6%)
Graduate student	213 (34.9%)	162 (21.4%)
No information	2 (.3%)	4 (.5%)
First-generation status		
First-generation student	54 (8.9%)	69 (9.1%)
Continuing-generation student	343 (56.2%)	515 (67.9%)
No information	213 (34.9%)	174 (23%)
Student enrollment by college		
Business	57 (9.3%)	110 (14.5%)
Computer Science	64 (10.5%)	34 (4.5%)
Engineering	185 (30.3%)	303 (40%)
Fine Arts	108 (17.7%)	43 (5.7%)
Humanities and Social Sciences	79 (13%)	97 (12.8%)
Information Systems and Public Policy	42 (6.9%)	51 (6.7%)
Science	39 (6.4%)	88 (11.6%)
No information	2 (.3%)	4 (.5%)

Note. Unless otherwise specified, cells display the number of students (%) within each condition.

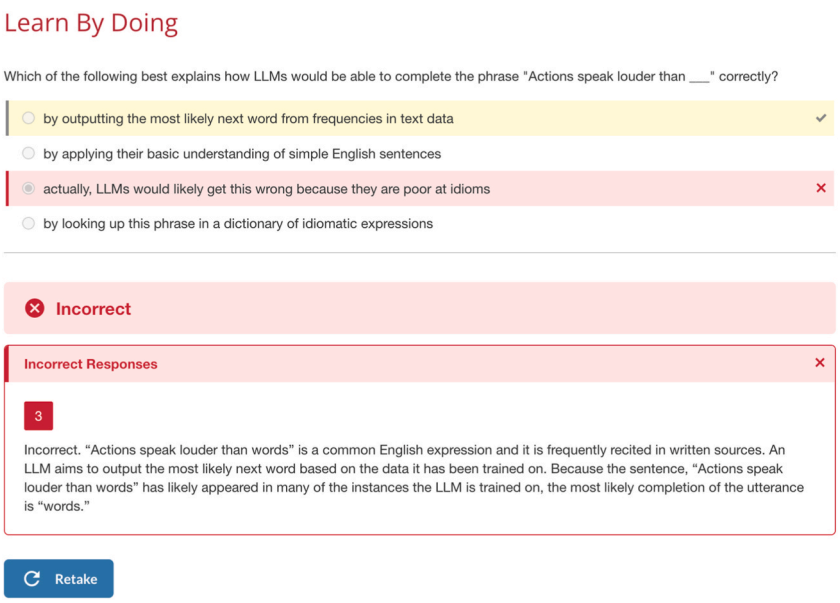


Fig. 1. Example embedded learn-by-doing practice activity with feedback related to learning objective1 (how LLMs work).

knowledge and skill assessment item as isomorphic pairs before randomly assigning each item in the pair to test Form A or test Form B. We counterbalanced presentation of test form order across both experimental conditions (i.e., half of the sample received test Form A at pre, and B at post and vice versa, regardless of experimental condition). Assessment of self-efficacy used identical items at pre and

post. We embedded pre- and post-tests in the LMS with programmed release and close dates to control appropriate access.

2.2.3.1. GenAI knowledge assessment. For each test form, eight multiple-choice questions (MCQs) tested students' knowledge about how large language models (LLMs) work, including hallucinations and how data are used to train LLMs (learning objective 1, LO1) and academic integrity, bias, and responsible use (learning objective 3, LO3). For research purposes, each correct response was awarded one point while any incorrect answer received 0 points.

Additionally, each test form included one item that asked students to match five scenarios to one main risk out of eight possible risks associated with genAI use. Risks included copyright infringement, violation of academic integrity, citing nonexistent references, amplification of harmful biases, privacy violations, ecological damage, human labor impacts, and physical damage to people. For each correctly matched pair, students received .2 points.

Four of the nine questions measured LO1 ("how LLMs work") and five questions measured LO3 ("responsible use"). We computed an overall (aggregate) knowledge score for all 9 items, as well as a score for LO1 and a score for LO3. Each knowledge score was converted to a percent correct score. Students did not receive feedback about their performance on either the MCQs or the matching. At the pre-test this was important to maintain the integrity of the learning intervention because feedback alone could have improved knowledge, independent of the modules.

2.2.3.2. GenAI skills assessment. Two authentic tasks assessed students' genAI usage skills.The first task provided students with a common academic scenario (prepare a presentation about a challenging topic in either an introductory Neuroscience or Statistics class, depending on the test form) and required them to write a prompt for a genAI chatbot to assist them with that task. With the exception of context-related cues, wording of the tasks was identical in both scenarios and presentation of scenarios at pre and post was counterbalanced. We scored student submissions on a rubric (see Supplementary Material) consisting of five categories aligned with the modules' instructional content, including the extent to which students (1) assigned the chatbot a role, (2) identified an audience, (3) defined the task, (4) articulated a purpose for the task, and (5) imposed constraints on the chatbot's response.

The second authentic task provided students with genAI output and asked them to identify two aspects that may require further investigation and to justify their response. Students provided their thoughts in two separate open-ended fields. Students were told that a fellow student had used ChatGPT to help them answer an essay prompt for a class. Specifically, that ChatGPT had been prompted as follows: "Write a paragraph arguing against gun control (Test Form A)/ legalization of marijuana (Test Form B). Make sure to include references." We scored submissions using a rubric (see Supplementary Material), evaluating discussions of: (1) the need to check the validity of facts or claims, (2) the need to check the credibility of the cited source, and (3) the awareness of potential bias in genAI's output. All criteria were coded as present if the student's response appropriately related the concept to the specific educational scenario (e.g., "I would need to track down the mentioned source to make sure it is real because AI might have made it up" would be coded as present for source-checking). Because the task asked students to identify two different aspects to investigate and provide reasoning, if a student provided reasoning about the same issue (e.g., source-checking) more than once, they received credit for that criterion only once. Students who received credit almost always exceeded simple recall by demonstrating comprehension of the principle in the context of the educational scenario.

For both authentic tasks, at least two researchers scored each response within a randomly-selected subset of submissions ($n = 174$), stripped of condition and pre/post submission identification. After norming on a smaller subsample of these responses (14% of the subset for the first, 8% for the second authentic task), a total sample of 174 students' responses were scored. Researchers discussed disagreements to consensus or, on rare occasions (2.7%), agreed to disagree on their ratings. For the first authentic task, an average of the scorers' ratings was used. For the second authentic task, disagreements on the binary rubric criteria were scored as the presence of that criterion as at least one rater detected evidence of the desired output analysis skill. For both tasks, rubric scores were converted to a percentage.

2.2.3.3. Self-efficacy for genAI skills assessment. To assess students' self-efficacy for generative AI skills, we administered the same 11 self-efficacy items (see Supplementary Material) at both the pre- and post-test. In line with recommendations for assessing self-efficacy (Bandura, 1997) students rated their confidence in their ability to do each of the 11 skills on a scale from 0 to 100% confident. Sample items included "I am __ % confident that I can explain the difference between a search engine result and generative AI output" and "I am _% confident that I can identify when using generative AI would be inappropriate in my courses." Internal consistency was high at the pre-test (McDonald's $\omega = .88$, see Hayes & Coutts, 2020) and the post-test (McDonald's $\omega = .94$). We calculated a self-efficacy composite for each time point to reflect the overall average self-efficacy at pre and post (see all items in the Supplementary Material).

2.3. Procedure

Prior to the fall 2024 semester, a university-wide call invited instructors to enroll in the present study, provided students' learning about genAI fit into the goals of their course. Interested instructors participated in an onboarding meeting prior to the semester, and committed to include the pre-test, post-test, and learning modules as required course assignments outside of instructional contact hours within the first two weeks of the semester. Instructors each received a small honorarium for complying with study parameters. Students received course participation credit for completing all required assignments. Student data was included based on student consent and age (18 years or older).

Instructors communicated the deadlines for each of these three assignments (pre-test, modules, and post-test). Students in the

control condition completed the pre-test, then the post-test before the learning modules became available. Students in the treatment condition completed the pre-test, then the learning modules, and then the post-test. This allowed us to compare the change from pre to post to measure the impact of the modules while not disadvantaging the control students by withholding the learning content. Both groups completed the pre- and post-tests within the exact same time windows (i.e., all students had to complete the pretest during a fixed 48-hour window and then again for the post-test during a fixed 48-hour window). The groups only differed in whether or not they had access to the genAI learning modules content prior to the post-test (treatment) or not (control). Treatment students had two days between the pre-test closing and the post-test opening to complete the learning modules, whereas control students were given access to the learning modules 24 hours after the post-test closed. As such, data collection for the entire experiment was completed within a six-day timeframe during the first week of classes (for an overview of the study design, please see Fig. 2.).

Instructions informed students that: the pre- and post-test would each take between 10 and 15 minutes to complete and that each had to be completed in a single session, the genAI learning modules would take approximately 90 minutes to complete, and they could stop and start as often as they wanted during the time window in which the modules were available. For consistency across courses and conditions, researchers programmed assignments (and their opening and closing times) in the LMS.

3. Results

To test the effects of the modules on students' genAI competency development, we analyzed the impact of the intervention on students' knowledge, skills, and self-efficacy. For the analyses below, sample size varies slightly based on completeness of data.

To account for the nesting of students within instructors (random assignment happened at the level of the instructor, i.e., course section), we tested our primary research question using multi-level models, treating students (level 1 units) as nested in instructors (level 2 units). We fit mixed-effects models using restricted maximum-likelihood (REML) in SPSS 30.0 to analyze the impact of condition (treatment, control), time (pre, post), and their interaction on student outcome variables. We modeled repeated measures over time within students and specified random intercepts at the instructor level. The F-tests reported here correspond to tests of the fixed effects in this model (with Satterthwaite approximation for degrees of freedom).

3.1. Impact of learning modules on genAI knowledge

Students' aggregated knowledge score (% correct) across the nine items and their scores disaggregated by relevant learning objectives were analyzed. Test form was included as a covariate.[3]

3.1.1. Aggregate genAI knowledge score

Results revealed a significant effect of time, suggesting students performed better at post than pre, $F(1,1366.00) = 399.07$, $p < .001$. There was a significant effect of condition, indicating students in the treatment condition outperformed students in the control condition, $F(1,34.08) = 12.05$, $p = .001$. The predicted time × condition interaction did not emerge, $F(1,1366.00) = .00$, $p = .99$. See Fig. 3.

3.1.2. Learning objective 1 (how LLMs work) knowledge score

As with the aggregate knowledge score, results for LO1 revealed a significant effect of time, $F(1,1364.58) = 257.70$, $p < .001$, such that students performed better at post than pre. There was a significant effect of condition, indicating that students in the treatment condition outperformed students in the control condition, $F(1,27.99) = 6.27$, $p = .02$. A significant time × condition interaction did emerge, $F(1,1364.58) = 45.87$, $p < .001$. Simple effects tests indicated that the two groups did not differ at pre, $F(1,38.73) = .00$, $p = .97$, but they did at post, $F(1,33.07) = 22.69$, $p < .001$. Although both groups improved from pre to post, the treatment group improved more, $F(1,1363.24) = 292.22$, $p < .001$, than the control group, $F(1,1365.66) = 38.85$, $p < .001$. See Fig. 4.

3.1.3. Learning objective 3 (responsible use) knowledge score

Results for LO3 again revealed a significant effect of time, $F(1,1365.45) = 239.28$, $p < .001$, such that students performed better at post than pre. There was a significant effect of condition, indicating that students in the treatment condition again outperformed students in the control condition, $F(1,47.99) = 13.20$, $p < .001$. A significant time × condition interaction emerged, $F(1,1365.45) = 34.38$, $p < .001$. Simple effects tests indicated that while both treatment, $F(1,1363.52) = 51.81$, $p < .001$, and control, $F(1,1367.01) = 205.05$, $p < .001$, improved over time, the treatment group improved less. The two conditions differed at pre, $F(1,93.70) = 34.40$, $p < .001$, but not at post, $F(1,72.43) = .07$, $p = .80$. See Fig. 5.

3.2. Impact of learning modules on genAI self-efficacy

Self-efficacy was assessed with identical items at pre and post, so test form was no longer a covariate. Results revealed a significant effect of time, such that students performed better at post than pre, $F(1,1348.43) = 993.23$, $p < .001$. There was a significant effect of condition, indicating students in the treatment condition outperformed students in the control condition, $F(1,22.28) = 27.42$, $p < .001$.

[3] Test Form A was found to be easier than test form B at the pre-test, and test form was not evenly distributed across groups. By chance, 66.6% of students in the treatment condition, but only 25.6% of students in the control condition, were given Form A at pre.

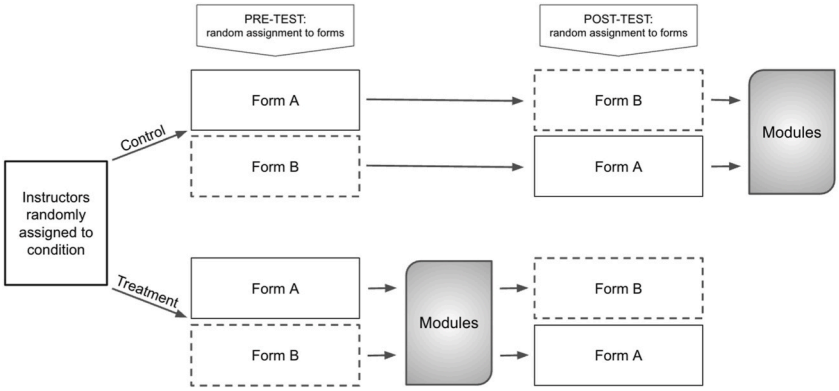


Fig. 2. Overview of study design and materials in order of presentation. *Note*: Students had 48 hours each time to complete the pre-test, modules, and post-test.

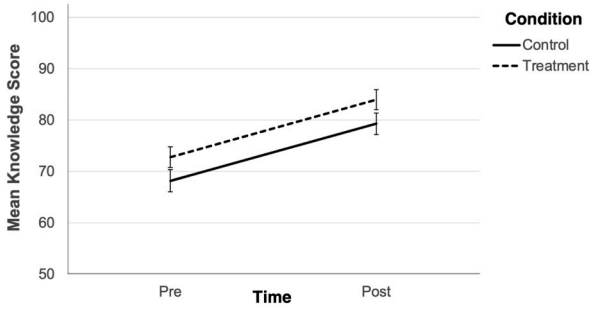


Fig. 3. Mean aggregate genAI knowledge score (% correct) by time and condition, after adjusting for test form as a covariate. Error bars represent 95% confidence intervals.

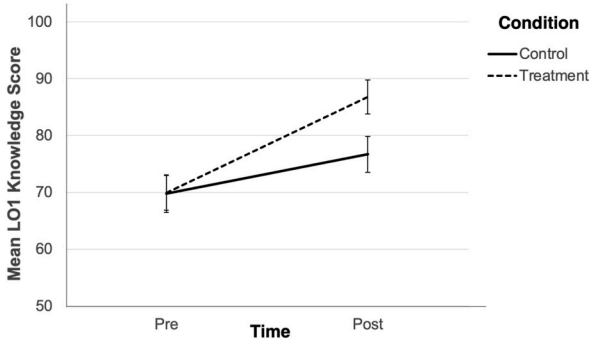


Fig. 4. Mean LO1 knowledge score (% correct) by time and condition, after adjusting for test form as a covariate. Error bars represent 95% confidence intervals.

The predicted time × condition interaction emerged, $F(1,1348.43) = 300.62$, $p < .001$. Simple effects tests showed that self-efficacy increased in both conditions, but the increase was substantially greater in the treatment, $F(1,1347.57) = 1340.90$, $p < .001$, than in the control condition, $F(1,1349.13) = 90.53$, $p < .001$. See Fig. 6.

3.3. Impact of learning modules on genAI skills

To test for the impact of the modules on genAI skills, we randomly selected a subsample of 174 students whose responses were evaluated using a rubric. Selections were made to ensure that this subsample was perfectly balanced across conditions and test forms,

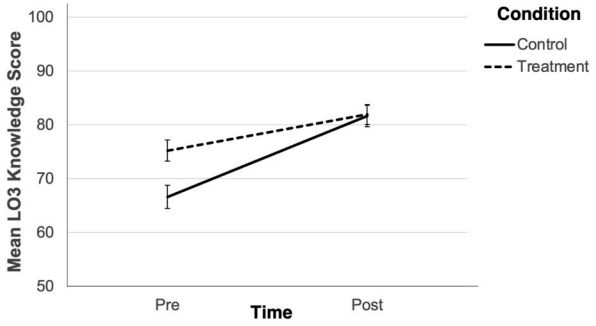


Fig. 5. Mean LO3 knowledge score (% correct) by time and condition, after adjusting for test form as a covariate. Error bars represent 95% confidence intervals.

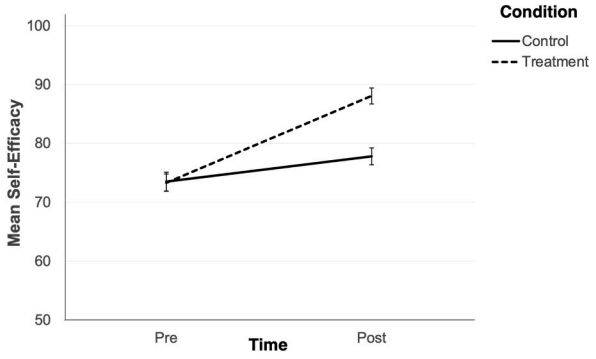


Fig. 6. Mean genAI self-efficacy score (% confidence) by time and condition. Error bars represent 95% confidence intervals.

though not across instructors.[4] Because of this, test form no longer needed to be considered as a covariate and data were treated as independent observations not nested in instructors.

3.3.1. Authentic task 1 - genAI-prompt-engineering skill

Rubric-scored performance on the first authentic task was calculated as the percentage of points earned out of 12 possible points across 5 component skills (see Supplementary Material). To examine whether the prompt engineering rubric scores improved following completion of the modules in the treatment condition, but not the control, we conducted a 2 (time: pre, post) × 2 (condition: treatment, control) mixed-model ANOVA. This revealed a significant main effect of time ($F(1,172) = 31.57$, $p < .001$, $\eta_p^2 = .16$), a significant main effect of condition ($F(1,172) = 4.88$, $p = .03$, $\eta_p^2 = .03$), and the predicted interaction ($F(1,172) = 39.60$, $p < .001$, $\eta_p^2 = .19$). Analysis of simple main effects revealed that scores improved in the treatment, $F(1,172) = 70.94$, $p < .001$, $\eta_p^2 = .29$, but not the control condition, $F(1,172) = .23$, $p = .63$, $\eta_p^2 = .001$. See Fig. 7.

We further investigated the five individual rubric criteria separately. As is evident in Table 2, students at the pre-test were best able to identify the task ($M = 62.2\%$) and the purpose ($M = 47.9\%$). However, they struggled with recognizing the need for assigning the chatbot a role ($M = 6.1\%$), and imposing constraints ($M = 5.3\%$). To test whether the pattern detected in the overall rubric scores held across all rubric criteria, a series of 2 x 2 mixed-model ANOVAs were run. The predicted time × condition interaction emerged for all criteria, except for specifying the task ("what"), for which a marginal interaction was found. Simple main effects revealed that for the other four rubric criteria the treatment group experienced a significant increase in scores from pre to post, whereas the control group did not (see Table 2). Students in the treatment group showed the largest improvements for identifying an audience (31.4% improvement), assigning the chatbot a role (25.5% improvement), and imposing constraints on the chatbot's response (20.4% improvement).

3.3.2. Authentic task 2 - evaluating-output skill

Rubric-scored performance on the second authentic task, in which students were asked to evaluate output produced by genAI, was calculated as the percentage of points earned out of 3 possible points. One point each was awarded for discussing fact-checking, source-checking, and the potential of bias. This task was scored using the same randomly selected sub-sample of 174 students that was used for Authentic Task 1. However, one treatment student of this subsample did not complete the second authentic task, leaving a final sample

[4] Our random subsample contained an average of 3 students per each of the 56 included sections out of the 65 total possible sections.

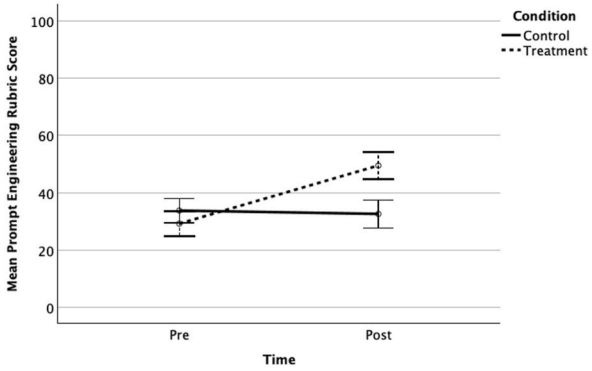


Fig. 7. Mean authentic task 1 (prompt engineering) rubric score (%) by time and condition. Error bars represent 95% confidence intervals.

Table 2
Descriptive statistics for earned rubric score (%) for authentic task 1 (prompt-engineering skill) by rubric criterion, time, and condition.

	Control (n = 87) Mean % (SD)		Treatment (n = 87) Mean % (SD)	
	Pre	Post	Pre	Post
Assigned role ("who")[a,b,c]	5.9 (23.4)	5.4 (20.7)	6.3 (23.9)	31.8 (45.1)
Audience ("for whom")[a,c]	29.7 (37.3)	29.6 (37.3)	18.7 (31.6)	50.1 (42.7)
Task ("what")	63.2 (33.3)	61.7 (34.4)	61.2 (33.9)	69.8 (34.1)
Purpose ("why")[c]	49.4 (37.2)	44.6 (39.7)	46.3 (37.4)	61.8 (35.8)
Imposed constraints[a,c]	7.8 (24.8)	8.7 (24.8)	2.9 (14.0)	23.3(39.2)

Note.
[a] Main effect of time, $p < .05$.
[b] Main effect of condition, $p < .05$.
[c] Time × condition interaction, $p < .05$.

of n = 173 for this analysis.

To test if students' skill improved from pre to post in the treatment, but not the control condition, we again ran a 2 (time: pre, post) x 2 (condition: treatment, control) mixed-model ANOVA. While there was no significant main effect of time, $F(1,171) = .23$, $p = .63$, $\eta_p^2 = .00$, there was a significant main effect of condition, $F(1,171) = 6.76$, $p = .01$, $\eta_p^2 = .04$, indicating that students in the treatment condition performed better overall than students in the control condition. However, the predicted time × condition interaction did not emerge, $F(1,171) = 1.53$, $p = .22$, $\eta_p^2 = .01$. See Fig. 8.

We again investigated individual rubric criteria. A series of Pearson's chi-square tests was run to see if the proportion of students who satisfied each criterion at pre and at post differed across conditions. At pre, there were no significant differences for any of the three criteria across conditions. At post, however, significantly more treatment students argued the need to fact-check and to source-check than did control students (see Table 3). There was no difference across conditions at post for satisfying the bias criterion.

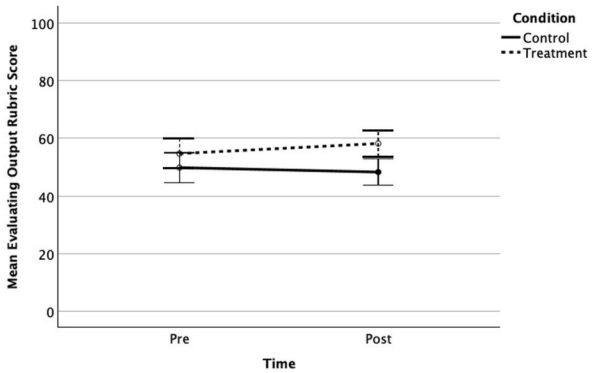


Fig. 8. Mean authentic task 2 (output evaluation) rubric score (%) by time and condition. Error bars represent 95% confidence intervals.

3.4. Equity of impact of learning modules on genAI knowledge, skills, and self-efficacy across student demographics

To test if treatment gains were equitable, a third factor of student demographic variable (i.e., sex, race, first-generation status, student level, academic discipline) was added to each of the above analyses that previously showed a positive impact of the modules (i.e., for LO1 knowledge, self-efficacy, and authentic task 1). Across all analyses, the two-way time × condition interaction did not differ for any of the outcome variables across levels of any of our demographic variables (i.e., there was no three-way interaction with sex, race, first-generation status, student level, or academic discipline), suggesting equitable effects of the modules across the board. That said, one three-way interaction (time x condition x sex) emerged, suggesting the modules leveled a previously detectable inequity in self-efficacy, $F(1,1338.50) = 5.91, p = .02$. Simple main effects showed that within the control condition, males had marginally higher self-efficacy than females at pre ($p = .053$) and this difference became significant at post ($p = .01$). Within the treatment condition, males had significantly higher self-efficacy than females at pre ($p < .001$), and this difference disappeared at post after engaging with the learning modules ($p = .20$). See Fig. 9. Engaging with the learning modules helped female students catch up to their male peers in their self-efficacy.

4. Discussion

We experimentally evaluated the effectiveness of a package of genAI competency training modules in a large and diverse student sample. Specifically, we sought to enhance students' knowledge, skills, and self-efficacy around how genAI works and how to use it responsibly in a higher education setting. Effective interventions at scale cannot be dependent on student background, prior knowledge, financial means, or individual instructors' capacity to develop these as part of their courses. Our modules meet all of these criteria and intentionally leverage evidence-based practices in online education (e.g., practice with immediate feedback, multi-modal engagement; Clark & Mayer, 2016; Clark & Mayer, 2023).

4.1. Development of genAI knowledge and self-efficacy

Consistent with previous research, our learning modules improved students' genAI knowledge of how LLMs work (LO1) and overall self-efficacy. Kong et al. (2021) and Korte et al. (2024) both found that students performed better on an AI concept inventory and reported higher self-perceived AI literacy after several hours of a learning intervention in the form of a workshop. Likewise, Sullivan et al. (2024) found that students reported increased genAI confidence after a 90-minute workshop (online or in-person). While previous studies suggested that genAI trainings held promise, our study substantially extended this work by including a control group, recruiting a large representative sample of students, and incorporating pre and post measurements of genAI knowledge and self-efficacy. Crucially, we also used authentic tasks to evaluate students' genAI prompt-engineering and output-analysis skills at pre and post, which we will discuss in more detail below. While our learning intervention was similar to Sullivan et al.'s (2024) training in length, unlike the three previous studies, students completed our learning modules asynchronously as a scaffolded, independent learning experience. As such, we demonstrate that a set of relatively brief online learning modules significantly improved knowledge of how LLMs works and overall self-efficacy in a representative sample *at scale*, obviating the dependency on in-person workshops or instructor capacity to achieve these outcomes.

In addition to the predicted knowledge (LO1) and overall self-efficacy gains in the treatment group, we also observed gains in the control group on these outcomes, though much less-pronounced. Significant increases in the control group may reflect assessment effects, as retrieval practice itself (even without feedback) can enhance recall (Roediger & Butler, 2011), and repeated exposure to the same assessment content can improve performance (Duff et al., 2012). Control group gains in knowledge and self-efficacy may also have been facilitated by incidental exposure to relevant material on genAI outside of this study (e.g., in the context of other instruction or self-driven exploration), or by social desirability or expectancy effects, given participants' self-selection into a university with a focus on AI research and innovation. Even though students were not made aware of their condition assignment, these factors may limit the generalizability of results from our student population to others. Nonetheless, gains in the control group were modest, and the significantly greater gains in the treatment group indicate that the modules provided value above and beyond other influences on students' performance for LO1 knowledge and self-efficacy.

However, we did not see evidence of an effect of the modules on students' knowledge of ethical and responsible use applications (LO3). Although students improved from pre to post regardless of condition, gains in the control group were more pronounced than in the treatment group, so that students in both conditions achieved equivalent knowledge scores for LO3 at post. Why might that be?

Table 3
Summary of authentic task 2 (evaluating-genAI-output skill) by rubric criterion, time, and condition. Cells report the proportion of students who earned the point for each criterion.

	Control (n = 87)		Treatment (n = 86)	
	Pre	Post	Pre	Post
Fact-checking[a]	80.5%	80.5%	87.2%	91.9%
Source-checking[a]	48.3%	41.4%	55.8%	67.4%
Identifying bias	20.7%	23.0%	20.9%	15.1%

Note. [a] χ^2 at post $p < .05$.

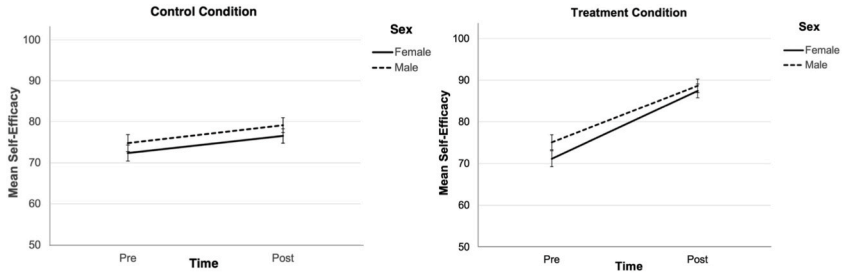


Fig. 9. Mean genAI self-efficacy (% confidence) by time and sex visualizing condition separately. Error bars represent 95% confidence intervals.

Students in the control group started significantly lower than their peers in the treatment group. The fact that they caught up to their peers in the treatment group without any exposure to the training modules suggests that both groups grew for reasons unrelated to the intervention. Given that the content of this LO is ethical and responsible use and that the timing of this study was during the first week of classes, one possibility is that students in both conditions were exposed to instructional reminders of what constitutes appropriate use as part of their courses. This null result suggests that a stronger intervention may be needed to impact this knowledge above and beyond what is already achieved by what instructors cover in their courses or that a more sensitive instrument may be needed to capture changes in this knowledge.

4.2. Development of genAI skills

Ultimately, knowledge and self-efficacy are in service to skills and to students being able to actually use the genAI tools effectively and appropriately. A distinctive contribution of the present study is the inclusion of authentic tasks asking students to demonstrate their skills in practice, rather than relying on self-report metrics or knowledge gains alone. We targeted and assessed two critical genAI skills: 1) How to effectively prompt a genAI agent to accomplish a goal, and 2) how to critically evaluate genAI output for accuracy and bias.

While engaging with the modules significantly improved overall prompt-engineering skills in the treatment group, analysis of individual rubric criteria revealed more nuanced effects. Initially, students were most likely to identify the task ("what" in Table 2) and the purpose of a genAI prompt ("why"). Arguably, these closely match the keyword-based web searches that students commonly employ. However, the unique affordances of genAI go beyond the ways in which one might interact with a search engine to produce more individualized results. This requires assigning the chatbot a specific role, including mention of the audience for whom the output is meant, and imposing constraints, i.e., clearly instructing the chatbot what *not* to do. These subskills were notably absent in students' initial prompts across treatment and control conditions. After completing the intervention, students in the treatment group experienced the largest learning gains for these subskills (performance increased by 20-30% compared to −.5% to .9% change in the control group). These findings suggest that web search experience does not automatically translate to effective prompt engineering, and that students require more scaffolded instruction to develop this new skill. Indeed, without such training, our control students did not naturally improve over time.

The second skill that we assessed was the ability to critically evaluate genAI-produced output. We did not find that treatment students grew in their aggregate rubric score for this skill more than the control group. However, when disaggregating overall rubric scores to look at more nuanced patterns, we found that the modules increased the frequency with which treatment students described the need for checking both sources and facts. These findings align with the learning objective and content of the specific module that covered lateral reading as a strategy to check the validity of genAI output.

The modules did not, however, change the frequency with which students discussed the potentially biased nature of genAI output. Coupled with the null effect for knowledge on appropriate use (LO3), which is related in content to understanding the ethical implications of genAI use, these findings suggest several opportunities for strengthening the training content of the modules. Students may need more practice in the various aspects of becoming responsible users of genAI chatbots, including prompting to mitigate bias (e.g., Bura et al., 2025; Morales and Raman, 2025) and critically evaluating genAI output for the potential of bias. Biases often feel intuitive, are automatic, and difficult to recognize or challenge unless people are explicitly encouraged and motivated to do so (Kahneman, 2011). Given this, bias detection may be particularly challenging to train, and it is possible that students require additional active learning examples, case studies, and reflective prompts to better understand the risks, challenges, and ethical implications of chatbot use in personal and professional settings. Similarly, faculty are often at odds with each other about what constitutes (in) appropriate use of genAI in a classroom setting (Watson & Rainie, 2026), complicating coherent messaging to students about ethical use. The increase in academic dishonesty on campuses (Watson & Rainie, 2026) and the potential of genAI output to increase inequities, reinforce stereotypes, and propagate bias (e.g., Rani & Bhosale, 2025) make training students in these areas particularly critical.

Taken all together, effectively interacting with genAI presents unique and novel challenges and requires a distinct skillset, especially given genAI's ability to take on assigned roles and its propensity to hallucinate and produce biased results. Most previous genAI training programs focused on knowledge development. Our results reinforce the value of providing targeted instructional support and authentic practice and feedback regarding the emergent skills required for efficacious genAI interactions, and of measuring their

growth. Since this kind of skill development is ultimately the goal of training, knowledge and self-efficacy are necessary, but not sufficient components. We observed a positive impact of our modules on treatment students' self-efficacy, indicating that they felt more confident to work with genAI. However, the training did not significantly improve knowledge for ethical use nor the singular skill of detecting potential bias in genAI output. This further reinforces the importance of strengthening the training of these key challenging outcomes.

4.3. Testing for equity in baseline genAI competency and learning outcomes

Consistent with literature reviewed in the introduction suggesting genAI instruction, access, and exposure vary by sociodemographic factors (e.g., Capraro et al., 2024; Davidoff, 2024; OECD, 2018), we observed baseline differences in genAI knowledge at the pre-test across birth sex, student level, first-generation status, and academic discipline (see Supplementary Material for all baseline analyses). Students' knowledge at the outset did not vary by race in our study (nor did skills or self-efficacy). Students' self-efficacy at baseline differed by their birth sex, academic discipline, and student level. For example, first-year undergraduate students had significantly lower knowledge and prompt-engineering skills at the pre-test than graduate students. These differences highlight the importance of not assuming that students enter higher education equally well-prepared to use this new technology. Instead they point to the need for genAI competency training, especially for incoming undergraduate students, to level the playing field.

Indeed, one of our goals was to test if the genAI training would lead to equitable learning outcomes. In line with this, we found that when there were greater learning gains in the treatment group, these were equitably experienced by all demographic groups. This suggests that all student groups in the treatment condition benefited from the training, regardless of their background. Given that genAI competency is an increasingly relevant workforce skill, addressing underlying inequities early is particularly important so that students can spend their time in college and graduate school effectively utilizing and building on this shared foundation.

4.4. Limitations and future directions

The present study demonstrates that the genAI training modules were an effective intervention in our large sample, even at a university known to attract students with a strong background in technology, science, and engineering. Promising as our findings are, several questions and needs remain.

For example, the present intervention consisted of the completion of the entire set of four learning modules. Our results speak to the impact of the holistic package of instruction on students' genAI competency, and cannot speak to the effects of isolated modules. It is possible that students experienced additive gains across the four modules, or that individual modules had differential effects. If future researchers or educators wish to target specific genAI knowledge or skills from individual modules, the impact of these should be assessed.

Additionally, our training was a one-time, 90-minute intervention. Although this efficient training was effective in the ways we have demonstrated in this paper, it was necessarily limited in content and scope. Still, we show that even a relatively small treatment "dose" was largely effective in the short term. Testing the robustness of the effects over time is a worthwhile area of inquiry for future research, as is whether booster trainings are necessary to maintain long-term competency in a rapidly evolving field. As with almost any knowledge and skill acquisition, learners will likely need repeated exposure to content and practice opportunities before reaching proficiency and fluency, even for foundational skills. It remains an open question whether the observed improvement in foundational genAI competency is sufficient to equip students for more advanced exploration and use of genAI, whether disciplinary differences appear at the more advanced level of competency, and whether additional training is necessary to address inequities at that level.

Similarly, students at other institutions may have different competency levels at baseline, and inequities across different student demographics may be more or less pronounced. Data collection for the present study happened in August 2024, at which time it was not reasonable to assume that all students had received appropriate exposure and training on genAI. It is feasible that students entering college going forward will evidence more genAI competencies as high schools increasingly offer students guidance and scaffolded practice on genAI use and output evaluation. Regardless, educators who wish to use these training modules to improve their students' genAI competency would benefit from confirming the positive treatment effects in their student population.

The rapid evolution of LLM chatbots will necessitate frequent updates to the content of any genAI training, including our modules, to stay current. In this continually evolving space, it will be important to build an iterative practice of content refinement, extension, and learning assessment. Monitoring the emerging essential skills for students and users of genAI can help guide future competency training efforts. For example, newer chatbot models can increasingly process multimodal input like handwritten notes, graphs, figures, tables, images, or voice requests, and produce audiovisual output that students may increasingly be asked to evaluate and work with. Our training focused on written prompt engineering and evaluation of text outputs, whereas future genAI interactions may increasingly include spoken input and output. As learning objectives and training content evolve with the changing capabilities and use of genAI, assessments of students' learning outcomes will likewise need to be updated in an ongoing iterative cycle. Any updates should use rubrics that are as sensitive as possible to the skills targeted by the training.

4.5. Implications for higher education

As the literature has documented, genAI competency is already emerging as a necessary workforce skill expected of new graduates (e.g., Cengage Group, 2024). Much like critical thinking and effective communication, this will likely be a transferable asset and therefore an enduring requirement for anyone in higher education and beyond. The present study, to our knowledge, is the first of its

kind to provide compelling evidence that a short, asynchronous, online intervention that is scalable and independent of academic discipline and instructors' capacity can be deployed to effectively develop students' foundational genAI competency, regardless of their position within a curricular trajectory.

Importantly, our results showed positive impacts of the training modules in correcting pre-existing inequities in knowledge, skills, and self-efficacy in different demographic groups. Beyond the demographics included in our study, extant research has also shown that genAI's affordances may be used to benefit second-language learning students, or those who have been traditionally under-represented in higher education, if careful guidance and instruction are provided (e.g., James & Andrews, 2024; Warschauer et al., 2023). The modules we present here are one such touchpoint for instruction.

Given that our learning modules are relatively brief and can be completed asynchronously, colleges and universities may embed training modules like these in onboarding requirements for incoming students, or in campus-wide first-year courses. Mirroring other required foundational courses, first-year seminars may be a particularly good opportunity for genAI competency development as students are likely to enter institutions of higher education with different levels of experience and exposure. Additionally, individual instructors of students of any level could require these modules in their courses. As with any learning experience, such mandatory training would benefit from giving students sufficient context for why the training provides low-stakes practice of important college readiness skills.

As genAI skills are increasingly treated as foundational for students, it is equally imperative for instructors to stay abreast of genAI developments, best practices, and how students are using these tools. In fact, instructors may benefit from taking trainings like the one presented here themselves, or institutions could modify student modules and provide versions better tailored to instructors' needs. However, trainings for instructors do not have to be limited to online learning tools, but may also consist of workshops, communities of practice, reading groups, one-on-one consultations, or institutional guidance. Such support is critical to alleviating pressure on individual instructors to find and review latest research, and to develop or refine practices accordingly.

While genAI carries a lot of potential for supporting various aspects of teaching, learning, and campus life, indiscriminate use of the technology also carries risks (see for example Bastani et al., 2025; Brewster et al., 2025; De Freitas, Uğuralp, Oğuz-Uğuralp, & Puntoni, 2024; Klimova & Pikhart, 2025). GenAI competency more broadly therefore also means that students, faculty, and administrators accept the ethical and pedagogical responsibility to keep negotiating bounds on its use. The time and resources that this ongoing work requires cannot be understated. Administrators can, and must, take an active role in supporting their campus communities. This includes connecting and investing in relevant initiatives and offices that are equipped to develop instructional content and policies. With administrative support, Centers for Teaching and Learning could be uniquely positioned to assist with, if not take a leading role in, providing support that aligns with their campus's needs. Campus leadership also has an obligation to support students, faculty, and staff in judiciously adapting to the quickly evolving uses of genAI. For example, student use of genAI goes beyond academic purposes, frequently extending to emotional and mental health support (e.g., Brewster et al., 2025; Pratt et al., 2024). Administrators can support employees' professional development through investing in internal trainings, funding for external trainings, or for conference attendance. Additionally, administrators might offset the expanding workload with efforts to increase compensation and/or reduce other job responsibilities.

5. Conclusion

Effectively and prudently using genAI is arguably among the most emergent job-readiness skills and as such it is incumbent on educational institutions to equip students with these skills. We have demonstrated the effectiveness of a brief, portable, asynchronous training for improving genAI competency at scale, across diverse populations of students at a highly selective research intensive university, as defined by an improvement in knowledge, self-efficacy, and, importantly, skills. The effectiveness of the training likely stems from its design, which emphasizes active practice with feedback rather than relying on passive content delivery.

Still, it is important to remind ourselves and our students that chatbots are tools to use intentionally and critically, not capriciously. Learning is a process that benefits from productive struggle (e.g., Steenhof et al., 2020), not the passive reception of knowledge. The affordances of genAI and the potential for innovative classroom use not withstanding, some of the most powerful predictors of educational success and retention are, indeed, rooted in what makes us uniquely human: the potential for interactions with mentors and peers to stimulate positive learning emotions, social connection and support, and a sense of belonging and meaning. As such, a crucial part of genAI competency may be to not lose sight of the fact that the "human-in-the-loop" is still very much a necessity.

CRediT authorship contribution statement

Allison E. Connell Pensky: Writing – review & editing, Writing – original draft, Visualization, Resources, Project administration, Methodology, Investigation, Formal analysis. **Lydia E. Eckstein:** Writing – review & editing, Writing – original draft, Visualization, Resources, Methodology, Formal analysis. **Michael C. Melville:** Writing – review & editing, Resources, Methodology, Formal analysis, Data curation. **Laura O. Pottmeyer:** Writing – review & editing, Resources, Methodology, Formal analysis. **Zach Mineroff:** Writing – review & editing, Software, Project administration, Investigation, Data curation. **Avi Chawla:** Writing – review & editing, Software. **Judy Brooks:** Writing – review & editing, Supervision, Software, Conceptualization. **Chad Hershock:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Marsha C. Lovett:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Data statement

The data are not publicly available due to the IRB protocol used to collect these data not permitting this. Aggregate data are available upon request from the corresponding author.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

On rare occasions, the authors used ChatGPT for minor linguistic and concision recommendations. All content was reviewed, edited, and approved by all authors.

Declaration of competing interest

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

Acknowledgements

We thank the faculty, student, and staff generative AI subject matter experts who assisted in identifying appropriate learning content of the modules. We also thank the faculty and student pilot testers. This work was supported by the Generative Artificial Intelligence Teaching as Research Initiative funded by the Provost of Carnegie Mellon University. Open Access funding was provided by Carnegie Mellon University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2026.105640.

Data availability

The authors do not have permission to share data.

References

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*. Longman.

Ardito, C. G. (2025). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning, 2025*(182), 11–28. https://doi.org/10.1002/tl.20624

Arslankara, V. B., & Usta, E. (2024). Generative artificial intelligence as a lifelong learning self efficacy: Usage and competence scale. *Journal of Teacher Education and Lifelong Learning, 6*(2), 288–302. https://doi.org/10.51535/tell.1489304

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Bowen, J. A., & Watson, C. E. (2024). *Teaching with AI: A practical guide to a new era of human learning*. Johns Hopkins University Press.

Bresó, E., Schaufeli, W. B., & Salanova, M. (2011). Can a self-efficacy-based intervention decrease burnout, increase engagement, and enhance performance? A quasi-experimental study. *Higher Education, 61*(4), 339–355. https://doi.org/10.1007/s10734-010-9334-6

Brewster, R. C., Zahedivash, A., Tse, G., Bourgeois, F., & Hadland, S. E. (2025). Characteristics and safety of consumer chatbots for emergent adolescent health concerns. *JAMA Network Open, 8*(10), e2539021–e2539022. https://doi.org/10.1001/jamanetworkopen.2025.39022

Bura, C., Myakala, P. K., & Jonnalagadda, A. K. (2025). Ethical prompt engineering: Addressing bias, transparency, and fairness. *International Journal of Research and Analytical Reviews, 12*(1), 145–152. https://doi.org/10.5281/zenodo.15124973

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., & Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus, 3*(6), 191. https://doi.org/10.1093/pnasnexus/pgae191

Cengage Group. (2024). *2024 graduate employability report*. https://cengage.widen.net/s/bmjxxjx9mm/cg-2024-employability-survey-report.

Chiu, T. K., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open, 6*, Article 100171. https://doi.org/10.1016/j.caeo.2024.100171

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). John Wiley & Sons. https://doi.org/10.1002/9781119239086

Clark, R. C., & Mayer, R. E. (2023). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (5th ed.). John Wiley & Sons.

Concannon, F., Costello, E., Farrell, O., Farrelly, T., & Wolf, L. G. (2023). There's an AI for that: Rhetoric, reality, and reflections on EdTech in the dawn of genAI. *Irish Journal of Technology Enhanced Learning, 7*(1). https://doi.org/10.22554/ijtel.v7i1.116

Crawford, J., Vallis, C., Yang, J., Fitzgerald, R., & O'Dea, C. (Eds.). (2023). Artificial intelligence is awesome, but good teaching should always come first. *Journal of University Teaching and Learning Practice, 20*(7), 1–12. https://doi.org/10.53761/1.20.7.01

Davidoff, M. (2024). How broadening AI access can help bridge the digital divide. *Educause Review*. https://www.proquest.com/openview/441b867e034de0f458f016a1bfc5add3.

De Freitas, J., Uğuralp, A. K., Oğuz-Uğuralp, Z., & Puntoni, S. (2024). Chatbots and mental health: Insights into the safety of generative AI. *Journal of Consumer Psychology, 34*(3), 481–491. https://doi.org/10.1002/jcpy.1393

Di Battista, A., Grayling, S., Hasselaar, E., Leopold, T., Li, R., Rayner, M., & Zahidi, S. (2023). *Future of jobs reports 2023*. World Economic Forum. https://www.weforum.org/publications/the-future-of-jobs-report-2023/.

Duff, K., Callister, C., Dennett, K., & Tometich, D. (2012). Practice effects: A unique cognitive variable. *The Clinical Neuropsychologist, 26*(7), 1117–1127. https://doi.org/10.1080/13854046.2012.722685

Duke, S. (2025). *AI is changing work - The time is now for strategic upskilling.* World Economic Forum. https://www.weforum.org/stories/2025/04/linkedin-strategic-upskilling-ai-workplace-changes/.

Eberly Center for Teaching Excellence & Educational Innovation. (2026). *The generative AI teaching as research (GAITAR) initiative*. https://www.cmu.edu/teaching/gaitar/index.html.

Freeman, J. (2024). New HEPI Policy Note finds more than half of students have used generative AI for help on assessments – But only 5% likely to be using AI to cheat. *HEPI*. https://www.hepi.ac.uk/2024/02/01/new-hepi-policy-note-finds-more-than-half-of-students-have-used-generative-ai-for-help-on-assessments-but-only-5-likely-to-be-using-ai-to-cheat/.

Generative AI. (2025). In *Teaching and learning*. University of Virginia. Retrieved October 31, 2025 https://genai.provost.virginia.edu/.

Gimpel, H., Gutheil, N., Mayer, V., Bandtel, M., Büttgen, M., Decker, S., Eymann, T., Feulner, S., Kaya, M. F., Kufner, M., Kühl, N., Lämmermann, L., Mädche, A., Ruiner, C., Schoop, M., & Urbach, N. (2024). *(Generative) AI competencies for future-proof graduates. Inspiration for higher education institutions*. Stuttgart, Germany: University of Hohenheim. https://doi.org/10.5281/zenodo.10680210

Grothérus, A., Jeppsson, F., & Samuelsson, J. (2019). Formative scaffolding: How to alter the level and strength of self-efficacy and foster self-regulation in a mathematics test situation. *Educational Action Research, 27*(5), 667–690. https://doi.org/10.1080/09650792.2018.1538893

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hong, J. W. (2022). I was born to love AI: The influence of social status on AI self-efficacy and intentions to use AI. *International Journal of Communication, 16*, 20. https://ijoc.org/index.php/ijoc/article/view/17728/3632.

James, T., & Andrews, G. (2024). Levelling the playing field through genAI: Harnessing artificial intelligence to bridge educational gaps for equity and disadvantaged students. *Widening Participation and Lifelong Learning, 26*(3), 250–260. https://doi.org/10.5456/WPLL.26.3.250

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kalla, D., Smith, N., Samaah, F., & Kuraku, S. (2023). Study and analysis of chat GPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology, 8*(3), 827–833. https://doi.org/10.1016/j.chbah.2025.100121

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

Kennedy, G. E., Judd, T. S., Churchward, A., Gray, K., & Krause, K. L. (2008). First year students' experiences with technology: Are they really digital natives? *Australasian Journal of Educational Technology, 24*(1). https://doi.org/10.14742/ajet.1233

Klimova, B., & Pikhart, M. (2025). Exploring the effects of artificial intelligence on student and academic well-being in higher education: A mini-review. *Frontiers in Psychology, 16*, Article 1498132. https://doi.org/10.3389/fpsyg.2025.1498132

Kohnke, L., Zou, D., & Moorhouse, B. L. (2024). Technostress and English language teaching in the age of generative AI. *Educational Technology & Society, 27*(2), 306–320. https://www.jstor.org/stable/48766177.

Kong, S. C., Cheung, W. M. Y., & Zhang, G. (2021). Evaluation of an artificial intelligenceliteracy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence, 2*, Article 100026. https://doi.org/10.1016/j.caeai.2021.100026

Kong, S. C., Korte, S.-M., Burton, S., Keskitalo, P., Turunen, T., Smith, D., Wang, L., Lee, J. C. K., & Beaton, M. C. (2025). Artificial intelligence (AI) literacy–an argument for AI literacy in education. *Innovations in Education & Teaching International, 62*(2), 477–483. https://doi.org/10.1080/14703297.2024.2332744

Korte, S. M., Cheung, W. M. Y., Maasilta, M., Kong, S. C., Keskitalo, P., Wang, L., … Gu, M. M. (2024). Enhancing artificial intelligence literacy through cross-cultural onlineworkshops. *Computers and Education Open, 6*, Article 100164. https://doi.org/10.1016/j.caeo.2024.100164

Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Empowering ChatGPTwith 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*, 16. https://doi.org/10.1186/s41239-024-00447-4

Lee, H. Y., Huang, Y. M., & Wu, T. T. (2025). ChatGPT in education: A systematic review of current landscape, limitations and future directions through general system theory lens. *European Journal of Education, 60*(4), Article e70262. https://doi.org/10.1111/ejed.70262

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and designconsiderations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). https://doi.org/10.1145/3313831.3376727

Lovett, M. C., Bridges, M. W., DiPietro, M., Ambrose, S. A., & Norman, M. K. (2023). *How learning works: Eight research-based principles for smart teaching* (2nd ed.). John Wiley & Sons.

Lovett, M., Meyer, O., & Thille, C. (2008). The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education,* (1), 1–16. https://doi.org/10.5334/2008-14

McDonald, N., Johri, A., Ali, A., & Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in Human Behavior: Artificial Humans, 3*, Article 100121. https://doi.org/10.1016/j.chbah.2025.100121

Mollick, E. R., & Mollick, L. (2022). New modes of learning enabled by AI chatbots: Three methods and assignments. https://dx.doi.org/10.2139/ssrn.4300783.

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open, 5*, Article 100151. https://doi.org/10.1016/j.caeo.2023.100151

Morales, J., & Raman, P. (2025). Prompt-engineering strategies for minimizing bias in large language model outputs: Applications in computing education. *Proceedings of the 56th ACM Technical Symposium on Computer Science Education, 2*, 1743. https://doi.org/10.1145/3641555.3705080

Morales-García, W. C., Sairitupa-Sanchez, L. Z., Morales-García, S. B., & Morales-García, M. (2024). Development and validation of a scale for dependence on artificial intelligence in university students. *Frontiers in Education, 9*, Article 1323898. https://doi.org/10.3389/feduc.2024.1323898

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy:an exploratory review. *Computers and Education: Artificial Intelligence, 2*, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041

Organisation for Economic Co-operation and Development. (2018). Bridging the digital gender divide: Include, upskill, innovate. https://www.enterprise-development.org/bridging-the-digital-gender-divide-include-upskill-innovate-oecd-2018/.

Pratt, N., Madhavan, R., & Weleff, J. (2024). Digital dialogue—how youth are interacting with chatbots. *JAMA Pediatrics, 178*(5), 429–430. https://doi.org/10.1001/jamapediatrics.2024.0084

Prensky, M. (2001). Digital natives, digital immigrants part 2: Do they really think differently? *On the Horizon, 9*(6), 1–6. https://doi.org/10.1108/10748120110424843

Rani, S., & Bhosale, A. (2025). Bias propagation in generative AI: Risk and mitigation strategies. *Applied Engineering Solutions and Technologies, 1*(1), 4. https://doi.org/10.70844/aest.2025.1.4

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Rogers, E. M. (1983). *Diffusion of innovations*. Collier Macmillan.

Ruediger, D., Blankstein, M., & Love, S. (2024). Generative AI and postsecondary instructional practices: Findings from a national survey of instructors. https://doi.org/10.18665/sr.320892.

Sabzalieva, E., & Valentini, A. (2023). ChatGPT and artificial intelligence in higher education:quick start guide. United Nations Educational, Scientific and Cultural Organization; International Institute for Higher Education in Latin America and the Caribbean. https://eduq.info/xmlui/handle/11515/38828.

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. W. W. Norton & Company.

Steenhof, N., Woods, N. N., & Mylopoulos, M. (2020). Exploring why we learn from productive failure: Insights from the cognitive and learning sciences. *Advances in Health Sciences Education, 25*(5), 1099–1106. https://doi.org/10.1007/s10459-020-10013-y

Sullivan, M., McAuley, M., Degiorgio, D., & McLaughlan, P. (2024). Improving students' generative AI literacy: A single workshop can improve confidence and understanding. *Journal of Applied Learning and Teaching, 7*(2), 88–97. https://doi.org/10.37074/jalt.2024.7.2.7

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138.

Talsma, K., Schüz, B., & Norris, K. (2019). Miscalibration of self-efficacy and academic performance: Self-efficacy ≠ self-fulfilling prophecy. *Learning and Individual Differences, 69*, 182–195. https://doi.org/10.1016/j.lindif.2018.11.002

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Torres, J. B., & Solberg, V. S. (2001). Role of self-efficacy, stress, social integration, and family support in Latino college student persistence and health. *Journal of Vocational Behavior, 59*(1), 53–63. https://doi.org/10.1006/jvbe.2000.1785

Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*(1), 89–101. https://doi.org/10.1016/j.cedpsych.2008.09.002

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*, Article 101071. https://doi.org/10.1016/j.jslw.2023.101071

Watson, C. E., & Rainie, L. (2025). Leading through disruption: Higher education executives assess AI's impacts on teaching and learning. *American Association of Colleges and Universities & Elon University's Imagining the Digital Future Center*. https://files.eric.ed.gov/fulltext/ED671878.pdf.

Watson, C. E., & Rainie, L. (2026). The AI challenge: How college faculty assess the present and future of higher education in the age of AI. In *American Association of colleges and universities & Elon University's imagining the digital future Center*. https://dgmg81phhvh63.cloudfront.net/content/user-photos/Research/PDFs/AI_Challenge.pdf.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., & Gabriel, I. (2022). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 214–229). https://doi.org/10.1145/3531146.3533088

Whitcomb, K. M., Kalender, Z. Y., Nokes-Malach, T. J., Schunn, C. D., & Singh, C. (2020). Comparison of self-efficacy and performance of engineering undergraduate women and men. *International Journal of Engineering Education, 36*(6), 1996–2014. https://www.ijee.ie/1atestissues/Vol36-6/24_ijee4004.pdf.

Wu, T. T., Lee, H. Y., Chen, P. H., Lin, C. J., & Huang, Y. M. (2025). Integrating peer assessment cycle into ChatGPT for STEM education: A randomised controlled trial on knowledge, skills, and attitudes enhancement. *Journal of Computer Assisted Learning, 41*(1), Article e13085. https://doi.org/10.1111/jcal.13085

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour, 8*(10), 1839–1850. https://doi.org/10.1038/s41562-024-02004-5

Young, J. C., & Shishido, M. (2023). Evaluation of the potential usage of ChatGPT for providing easier reading materials for ESL students. In *Proceedings of EdMedia+ Innovate Learning* (pp. 155–162). https://www.learntechlib.org/p/222496.

Zhang, X., Zhang, P., Shen, Y., Liu, M., Wang, Q., Gašević, D., & Fan, Y. (2024). A systematic literature review of empirical research on applying generative artificial intelligence in education. *Frontiers of Digital Education, 1*(3), 223–245. https://doi.org/10.1007/s44366-024-0028-5