Joseph, D. C., & Muthiah, R. (2026). A formative assessment of traditional undergraduate students' use of an AI learning platform in STEM courses. *Intersection: A Journal at the Intersection of Assessment and Learning, 7(1),* 158-173.

A Formative Assessment of Traditional Undergraduate Students' Use of an AI Learning Platform in STEM Courses

Dane C. Joseph, Ph.D., Rick Muthiah, Ph.D.

Author Note

"We have no conflicts of interest to disclose"

Intersection: A Journal at the Intersection of Assessment and Learning
Volume 7, Issue 1

Abstract: A small to midsized university supported science, technology, engineering, and mathematics (STEM) education by partnering with a reputable online learning company to provide students and faculty access to artificial intelligence powered textbook problem sets and chatbots. An in-house formative assessment study was conducted in Fall 2024 and Spring 2025 to understand how traditional undergraduate students engaged with the platform. Quantitative results suggests that the general population did not utilize it as much as initially expected. Qualitative findings imply that students might have preferred alternative platforms like ChatGPT, though this may have had nothing to do with the company itself. Overall, students sought superficial answers and neglected the conceptual foundations of assignments. They may have been overwhelmed when utilizing these tools given the lack of guidance from instructors on how, if at all, to use them. The same might be true of instructors. One specific recommendation is to increase training opportunities for how to interface with specific AI tools.

Keywords: *STEM, learning outcomes, formative assessment, choice, mixed methods*

Purpose of the Study

The purpose of this formative assessment study was to examine traditional undergraduate (TRAD) students' initial use of an artificial intelligence (AI) platform in science, technology, engineering, and mathematics (STEM) courses. Academic performance data was collected for the students who attended a small to midsized university in the Pacific Northwest (GFU). A handful of interviews were also conducted to provide context by investigating students' motivations for using the platform. While AI-embedded course experiences are quite nascent, higher education institutions have long contended with adapting instructional strategies to support students as new tools appeared and cultural changes occurred (Reigeluth, Beatty, & Myers, 2017; Richey, Klein, & Tracey, 2011).

Students bring a wide range of motivations, values, and abilities and evidence suggests that these differences are connected to poor or mediocre STEM performance in lower division courses (Botnaru et al., 2021; Perez, Cromley, & Kaplan, 2014; Renninger & Hidi, 2019). Low performance predicts course, major, and even institutional dropout as well as general academic struggle (Hecht et al., 2019; Young et al., 2018). As part of its continuous improvement mission and promise to support students, GFU partnered with an online learning company (OLC) to support its community of learners in their STEM academic pursuits. The OLC abbreviation is used in this paper to respect privacy and data concerns expressed by their chief executive officer (CEO), with whom we worked closely during the project.

Teachers and students received access to the OLC platform, which included textbook problem sets and an artificial intelligence (AI) chatbot. The chatbot is driven by large language models (LLMs) and content is validated by a national cadre of STEM instructors, content curators, instructional designers, technologists, and subject matter experts. There is large variation in LLMs, and the CEO, who chose to remain anonymous, agreed to use the following quote to explain their approach:

> Because we are on the application layer, our advantage is the ability to leverage the best models for our particular use cases combined with proprietary data. So, in these cases we are leveraging all the best of breed models, both closed and open source to provide the best output. Our ML team continually benchmarks each against each other and against our outputs, across subject and difficulty levels. And yes, the interactions and data across our entire platform would be used to train our own models. We have API Data Privacy agreements in place with each model provider. (personal communication, December 15, 2025)

Navigating the plethora of available AI-based education platforms is difficult in the current hype-saturated educational technology market. Although people often say they prefer more options, it often comes at a cost to making decisions (Thaler & Sunstein, 2009). Yet, AI tools will continue to proliferate the classroom for the near future (Khan, 2024; Mollick, 2024; Selwyn, 2019). Highly popular platforms like ChatGPT provide access to subject matter content with chatbots that can simulate interaction between teachers and learners. The likelihood is that they will only improve with time and that there is no end in sight as to which options students pick or how they utilize them. However, not all chatbots are specifically designed for the undergraduate STEM population in the way OLC is. Thus, we were thankful they came alongside us to provide expert support.

Nevertheless, students might perceive that these platforms are their own, infallible one-on-one tutor. Kestin et al. (2025) reminded us that many AI tutors, even some used within higher education, were "generally designed to be helpful, not to promote learning. They are not trained to follow pedagogical best practices…" They further cite Kortemeyer (2023) when describing the overconfidence of some AI tutors "when giving an incorrect answer or when marking a correct reply as incorrect." Such perceptual errors are sometimes exacerbated for certain demographic groups, particularly at-risk ones such as financially poorer or first-gen status students. Compared with students who are more likely to have social networks with greater access to these systems, at-risk students might be more inclined to misunderstand the platforms' power and limitations (Qudrat-Ullah, 2025). Hence, we focused on examining differences between first generation and non-first-generation students.

Even when students have access, it is still unclear whether and how using AI platforms directly improve academic success. It is possible that it might mediate the burden on students' working memory capacity to recall everything that professors say in class. Students might struggle to comprehend in real time the conceptual or procedural connections that professors demonstrate in lectures, particularly those who ignore validated instructional design principles (Kirschner et al., 2018). A different theory is that such platforms create a false sense of reliability and deter students from attentively engaging and investing full cognitive and behavioral resources in class; knowing that they could rely on the tools for activities like homework assignments. Either way, AI platforms might arguably enable or hinder students' learning strategies to engage, comprehend, supplement, or reinforce course content.

Beyond access, evidence suggests that higher-achieving students have the self-regulation to plan, monitor, evaluate, and revise their learning strategies to attain their goals (Ertmer & Newby, 1996; Schunk & Greene, 2018). Examples include using generative learning strategies (GLS) such as summarizing content, mapping concepts, drawing diagrams, self-testing, and elaborating principles and theories in one's own words (Fiorella & Mayer, 2015). GLS can differ in their effectiveness by task and subject-matter demands. Expert consensus is that what matters overall is for learners to repeatedly and deliberately practice their subject to form stronger cognitive schemas and content mastery—assuming they have the appropriate motivations, quality and quantity of feedback, and a conducive environment within which to learn (Ambrose et al., 2010). Yet to what degree the environment, strategies, and motivations are effective when AI platforms are embedded remains to be understood.

Thus, we had no reason to think that OLC would or would not be utilized in evidence-based ways, or how it would impact strategies and motivations for learning STEM. So, we sought to conduct a formative assessment of student engagement with OLC. A formative approach is significant because it supplies GFU and its professors with evidence to make instructional improvements. It also aims to understand the situation without making summative judgments about OLC's merit or worth as a learning tool. It might help institutions like GFU form better partnerships with companies like OLC—which provided a temporary but extremely generous discount to GFU for using its platform—and for OLC to improve its services.

Overall, it is important to note that while research focuses on generalizing from samples to populations, assessment serves decision makers' practical needs. Evaluation passes a judgment on a program, person, policy, or process so that stakeholders can make higher-stakes decisions (Mertens, 2019)—such as whether to continue funding a program, shelf a policy, or terminate employment. But assessment adopts a continuous-improvement approach. It does not necessarily aim for judgment, generalizability, or reproducibility but instead allows users to efficiently make necessary adjustments along the way. In the end, we appreciate Suskie's (2018) note that "Assessment is simply deciding what we want students to learn and making sure they learn it (p.7)." Along with 'deciding what' we include 'deciding how' in the mix.

Methodology

We adopted a mixed methods approach to ascertain descriptive statistics and qualitative insights. This provided a more comprehensive and triangulated perspective (Chen, 2014). Participants were Fall 2024 TRAD students who also enrolled in the Spring 2025 semester, controlling for persistence. TRAD student ages typically range from 18 to 23. The sampling frame included those TRAD students who registered for a free OLC account and further utilized it at least once during the Fall 2024 semester.

This turned out to be 381 students and faculty, out of which 323 were TRAD students who used it at least once. We focused this study on mathematics courses as it is a perennially difficult subject for students, and the department members taught courses to most of these students. These faculty vary in academic status from Instructor to Full Professor. By focusing on mathematics courses and faculty, this helped to lower variability in instructional approaches across subject-matter domains.

Institutional review board (IRB) approval was obtained to collect and merge GFU data with OLC data. All protocols for using students' course data abided by FERPA laws and regulations. OLC provided GFU with use-statistics for their platform, including: the total number of pages viewed on the platform, total number of AI conversations between the student and the platform's LLM, the total number of AI messages sent, and the total number of AI questions asked. The difference between a conversation and a message is that for conversations the user engaged in multiple (at least two) responses to the LLM; while a message signifies that they sent a request, received a response, and were either satisfied with the answer or for other reasons did not follow up.

OLC provided this data linked to the students' email addresses that were used to sign up for their free account. A different unique identifier permanently replaced the email addresses within the working data set. Only the co-PIs had access to this key from then on so that OLC could not see how students performed within their GFU courses. We used this unique identifier to join OLC data with students' GFU records, including demographic variables [first-gen status, race, sex, major field of study, and term-to-term persistence] and achievement data [course grades, term grade-point average (GPA), and cumulative GPA].

Quantitative data was primarily analyzed using the JASP statistics software program. Exploratory data analysis was performed, and while not inferential in intention, we used simple general linear modeling (GLM) to explore demographic differences in OLC use. The completely anonymized data and the JASP results were verified using one of the author's Gemini Pro accounts (Google, 2025). GFU has special accounts for faculty members that keeps the interactions private and do not train the Gemini LLMs. But the co-PI used prompt engineering to ascertain Gemini results based on the same analyses conducted in JASP. Both sets of analyses turned out to be almost identical, except for a few decimal places.

All mathematics faculty who taught during the Fall 2024 semester were invited along with a purposive set of students to participate in short, 30-minute semi structured interviews during the early Spring 2025 semester. This paper reports the interview results from three participants in total: two students and one mathematics faculty member. These interviews were solicited via email and interested participants met at a location of their choice limited to the GFU Newberg campus. At the session, the interviewer reviewed the informed consent form and all rights and responsibilities pertaining to participation in the study. The interview questions (see Appendix A) were largely inspired by the *Learning Strategies* scales on Pintrich's Motivated Strategies for Learning Questionnaire. Responses were recorded on paper so that the interviewer could quickly capture a general sense of participants' thoughts and motivations. Participants were assigned pseudonyms. OLC also provided a qualitative thematic analysis of students' AI conversations, but these are not included in this report because we could not guarantee maintaining their anonymity if we did. However, they are separate from our qualitative results.

Results

Quantitative Findings

The findings of our study are separated into quantitative and qualitative sections. However, we first provide a 'check-in' review of how students interacted with OLC roughly halfway through the Fall 2024 semester. By October 21st, 2024, there were 3,690 messages sent and 1,077 conversations with the LLM. While there were over 1,000 LLM conversations, with one user accumulating more than 150 of those, the average was just under four conversations per person. Twenty-eight people had 10 or more conversations at this point. On average, approximately 13 LLM messages were sent per person. Fifty-seven people sent more than 10 messages. Results showed 1,527 questions were asked with an average of around five questions asked per person, though only 19 people asked more than 10 questions. Information for content access by subject area is shown in Table 1.

Table 1.

OLC Content Accessed by Subject Area at Midpoint of Fall 2024 Semester

Subject Area	Total Pages Viewed	Unique Users	Average Pages per User
Calculus	1,325	55	24
Precalculus	171	14	12.2
Statistics	161	15	10.7
Algebra	58	31	1.9
Mathematics	5	1	5
Geometry	2	2	1
Pre-algebra	1	1	1

Note. Specific page view averages for subjects other than Calculus were calculated based on the provided total pages and user counts. Most users viewed between 15–45 pages total, with 27 users viewing 48 or more pages.

At the end of the Fall 2024 semester, data from roughly 323 students were collected with the following demographic representations: 93 (29.5%) were first-gen status; 192 (60%) identified as male, 126 (39%) as female, and fewer than 1% did not answer; 201 (63%) students identified as White, 25 (7.8%) as Asian, 6 (1.9%) as Black or African American, 48 (15%) as Hispanic, 3 (<1%) as Native Hawaiian or Pacific Islander, 24 (7.5%) as two or more races, and 15 (3.4%) reported they either did not know or did not respond. Students' self-identified academic majors were recoded to suit the following groups: math, engineering, and computer science (MECS); natural sciences such as biology, biochemistry, kinesiology, and nursing (NATS); and social sciences and humanities—such as social work, psychology, education, communication, and business administration—as well as undeclared majors (SOCS). They were split this way to reflect the depth of mathematics courses their respective majors typically encounter through the duration of their programs of study. The data set included roughly 151 (48%) MECS, 93 (29.5%) NATS, and 71 (22.5%) SOCS. See Table 2.

Table 2.

Demographic Characteristics of Student OLC Users at End of Fall 2024 Semester

Characteristic	n	%
First-Generation Status		
Non-First Generation	222	70.5
First-Generation	93	29.5
Ethnicity		
White	201	63
Hispanic	48	15.1
Asian	25	7.8
Two or more races	24	7.5
Unknown Race and Ethnicity	11	3.5
Black or African American	6	1.9
American Indian or Alaskan Native	3	0.9
Native Hawaiian or Other Pacific Islander	1	0.3
Sex		
Male	192	60.2
Female	126	39.5
Unspecified	1	0.3

Note. N = 323. Percentages may not total 100 due to rounding. Percentages are based on valid responses, excluding missing data (n=8 missing for First-Generation Status; n=4 missing for Ethnicity; n=4 missing for Sex).

To examine group differences, we used t-tests and ANOVAs. However, we were concerned with inflating the type-I error rate and finding false positives from repeatedly testing differences across multiple dependent variables between the same independent variable. Thus, we applied Bonferroni corrections. This correction divides the alpha value (0.05 in our case) by the number of comparisons; in this case, five of them: pages viewed, sessions, conversations, platform messages sent, and questions asked. Hence, the new, more conservative alpha (0.05/5 = 0.01) was used to interpret our results. Independent samples t-tests revealed the groups did not statistically differ in terms of their use of OLC's AI-platform (Table 3).

Table 3.

Independent Samples t-Test Results of OLC Engagement by First-Generation Student Status

Variable	Group	n	M	SD	t	df	p	Cohen's d
Pages Viewed	NFG	222	22.64	64.99	-0.07	313	0.94	-0.01
	FG	93	23.23	60.38				
Sessions	NFG	222	9.16	11.45	0.53	313	0.59	0.07
	FG	93	8.41	11.62				
Conversations	NFG	222	8.16	21.09	1.24	313	0.22	0.15
	FG	93	5.22	13.94				
Messages Sent	NFG	222	27.89	63.81	0.08	313	0.94	0.01
	FG	93	27.17	85.82				
Questions Asked	NFG	222	7.16	39.69	-0.61	313	0.54	-0.08
	FG	93	10.19	41.21				

Note. FG= First Gen and NFG = non-First Gen. N varies slightly due to missing academic data for some students. Statistical significance is indicated by $p < .01$. Positive t-values indicate the Non-First-Generation group had a higher mean; negative t-values indicate the First-Generation group had a higher mean. While the Brown-Forsythe test showed equal variances for the various dependent variables between the two groups, assumptions of normality were violated for all dependent variables.

Summary statistics for OLC engagement by general academic area is shown in Table 4. SOCS and MECS arguably had the lowest and highest overall engagement with OLC, respectively. One supposition for NATS falling in the middle is that they are more likely than SOCS to take Calculus and Precalculus though less likely than MECS to engage deeply with Mathematics content.

Table 4.

Descriptive Statistics of OLC Engagement by Students' Self-Reported Academic Plan

Academic Plan	Total Students	Total Pages Viewed	Active Users (n)	Questions Asked
MECS	151	5,210	109	2,292
NATS	93	1,808	63	154
SOCS	71	168	33	91

Note. "Active Users" are defined as students who viewed at least one page.

One-way ANOVAs were conducted to explore whether there were any statistically significant differences in OLC engagement across the academic groups. It is important to note that the Levene's test of equality of variances assumptions for the ANOVA tests was violated; thus, alternatives such as the Brown-Forsythe and Welch tests were used to investigate group differences. The Welch and Brown-Forsythe tests are robust to violations of the assumption of homogeneity of variance. These are reported in this paper using the conservative alpha of 0.01. Statistically significant differences from both the Brown-Forsythe and Welch tests were found for the number of OLC pages viewed and sessions. Questions asked were only statistically significantly observed for the Brown-Forsythe test (Table 5). Conversations and messages sent showed no statistically significant differences across academic groups for either the Welch or Brown-Forsythe tests.

Table 5.

Robust Tests of Equality of Means by Academic Groups

Variable	Group	M	SD	Welch	df1, df2	p	Brown	df1, df2	p
Pages Viewed	MECS	34.5	73.37	16.66	2, 159.59	< .001	8.62	2, 226.72	< .001
	NATS	19.44	66.31						
	SOCS	2.37	8.87						
Sessions	MECS	12.06	12.32	15.62	2, 189.73	< .001	14.59	2, 293.07	< .001
	NATS	7.61	11.05						
	SOCS	4.35	7.97						
Conversations	MECS	9.58	23.34	1.66	2, 185.19	0.194	2.1	2, 292.37	0.125
	NATS	6.01	13.95						
	SOCS	4.97	16.23						
Messages Sent	MECS	35.08	75.05	4.14	2, 190.13	0.017	2.54	2, 246.63	0.081
	NATS	28.67	82.06						
	SOCS	13.39	40.72						
Questions Asked	MECS	15.18	56.71	4.34	2, 180.20	0.014	8.41	2, 163.65	< .001
	NATS	1.66	5.93						
	SOCS	1.28	8.18						

Follow-up post-hoc tests revealed that the differences in the number of pages viewed occurred between MECS and SOCS students. MECS students differed from both SOCS and NATS students where the number of sessions was concerned. But questions asked showed no statistically significant differences across the post-hoc test groups using the 0.01 criterion. However, practically speaking, MECS students seemed to ask more questions than the other two groups (Table 6).

Table 6.

Robust Post-Hoc Comparisons for Significant Engagement Dependent Variables

Dependent Variable	Comparison	Mean Difference	p (adj)
Pages Viewed	MECS vs. SOCS	32.14*	< .001
	NATS vs. SOCS	17.07	0.032
	MECS vs. NATS	15.06	> .05
Sessions	MECS vs. SOCS	7.71*	< .001
	MECS vs. NATS	4.45*	0.008
	NATS vs. SOCS	3.26	0.03
Questions Asked	MECS vs. NATS	13.52	0.011
	MECS vs. SOCS	13.90	0.011
	NATS vs. SOCS	0.37	> .05

Note. Post-hoc analysis was performed using pairwise Welch *t*-tests with Holm's correction for multiple comparisons to account for unequal variances. Significant differences ($p < .01$) are indicated with an asterisk (*). Mean difference is calculated as (Group 1 mean - Group 2 mean); a positive value indicates the first group listed had a higher mean.

Qualitative Findings

Four total interviews were conducted with three students and one mathematics faculty member. However, one student asked not to have their data included in the study after the interview was conducted. To honor their request, the analysis here reports the results from the two consenting students and faculty member. Amongst them were a male, senior engineering student (pseudonym Eddy), a female, sophomore mathematics student (pseudonym Maggie), and a mathematics faculty member (pseudonym Jordan). Because the mathematics department is small (fewer than six fulltime members), a gender-neutral name is used in this report to mask Jordan's identity. Both Eddy and Jordan discussed using several GLS in their learning and teaching activities, respectively.

Eddy described how his math classes were, to his pleasant surprise, very project based and believed that much of his success stemmed from his general love of the subject. He signed up for an OLC account because he saw the ad in the institution's newsletter. Although he initially used the platform very early on, he "frequently found several answers to be incorrect and quickly got turned off." Asked how he evaluated their correctness, he described that sometimes, "You could just tell…[and] I would usually if not always compare [OLC's] answers with ChatGPT answers since I also had an account for that one as well."

When asked about general comparisons between OLC and ChatGPT, Eddy perceived the former was less capable and more restrictive than the latter. He thinks that some of his peers may even have signed up for OLC because they did not have prior success with ChatGPT. He believes this may have been due to a lack of knowledge and ability to construct appropriate prompts, also known as prompt engineering. If he had to pick again, he would choose ChatGPT "99% of the time." At the time he had a paid subscription to its platform. It is important to note here Gertz's (2024) description of the vicious cycle between students and instructors as the former may often use tools like ChatGPT to complete tasks under the pressures of deadlines, as the latter increase the number of assignments and deadlines in an aim to tighten their grip on cheating. Another likely scenario for ChatGPT's popularity is its early adoption due to its more open access.

Nevertheless, Eddy strongly suggests the institution not subscribe to OLC in the future but instead use those funds to obtain paid ChatGPT subscriptions for students who want it. While he recognized that it has its own limitations—he even recently considered ending his subscription in favor of spending it on Deep Seek—he does not condemn platforms like OLC for their imperfections. He believes a good long-term strategy is training and exposing students to use them. Training could include workshops, Canvas tutorials, and live demos for prompt engineering, or even on how to use AI ethically. When asked what turned him off OLC, he responded, "I think [OLC] is like Chegg and has…huge cheating potential to game the system for a final answer without actually understanding how you arrived there." He recalled a few conversations with classmates while doing homework exercises that solidified his belief that people were "using it to get answers…not learn anything." But when asked whether students could do this with ChatGPT, he admitted he had no good answer. Admittedly, Eddy has not attended any of the student AI lunch sessions that the university had so far offered; but thought that would be different if he were a freshman or sophomore. He also questioned why some faculty embraced AI platforms while some did not.

One of Maggie's tutors recommended OLC to her. One of her professors also suggested using AI tools when they themselves were unavailable to explain concepts to her. However, Maggie thought OLC was only helpful when "you already knew the material." In a follow-up email I asked her to write her biggest concern for using OLC or similar tools as or along with other learning strategies. She penned:

> It's good for review for the most part, when you already have somewhat of a grasp on the information, because the explanations can be too complicated or have vocab words hyperlinked with "tell me more about…" but it will just explain them in the same wording as before. At times, it is very vague and doesn't respond well when you ask it to further explain concepts. The images also don't always make sense, which can make it more confusing when learning. Another bug is that it can quiz you, but it is inconsistent with keeping a flow of questions. I also tried to make a practice exam or two with my notes but it didn't quite pick up on the main focus details.

Maggie still saw OLC's supplementary potential but preferred to stick to its mastery study tools when using it as it aligned more with what she believed she was learning in class. She described that, "There are a lot of tools out there; I pick and choose when to use some and what to use when." This sentiment reinforces notions of the constant tension between choice, learning, and performance. It isn't clear whether Jordan directly taught Eddy or Maggie as we purposefully sought to refrain from asking such questions; both to protect anonymity and out of a concern that students might use it as a window into venting about faculty members. Furthermore, as an in-house assessment, the interviewer is a professional colleague with many of these faculty.

Jordan described the importance of teaching students to use GLS in courses, such as drawing with pictures, imagining plausible but alternative or hypothetical worlds, engaging in thought experiments, and mapping concepts. For example, Jordan would have students imagine walking on a long curve to describe certain parts of Calculus or functions; they believed that this sort of engagement was more motivational and interesting. Jordan believed that their teaching experience led to a hunch that students—typically those in Calculus and first-year mathematics courses—engaged with AI platforms at the surface level. When asked to explain what that meant, they elaborated that they found students sought answers to the homework assignments but did not take the time to process how problems could be better framed and approached.

Jordan used OLC for a little while but did not like the features of other professors explaining topics. In one situation Jordan found an exact matching problem that was both used in class and on the OLC platform itself, but only after the fact. So now Jordan does extra work to find ways to challenge students beyond what is regularly called for in terms of subject-matter learning. For instance, Jordan might ask for a 'direct proof' instead of a proof by contradiction as conceptual problems are approached quite differently to procedural ones. But Jordan highly encouraged students to use elaboration and explanations during class time to observe their understanding and help them improve this aspect so that they might replicate that process with AI tools if they did end up using it.

To be fair, Jordan is not opposed to educational technology and AI but has a deep concern for whether students are actively choosing to use it in responsible and effective ways—although sometimes Jordan perceived a sort of personal teaching competition with AI. Overall, like Eddy, Jordan loves mathematics and has observed quite a lot of variability across mathematics instructors' classroom practices and expectations for students' AI use, remarking, "Some instructors I know won't even allow calculators in class!" This striking comment echoes Eddy's questions on mixed faculty adoption. Could faculty be afraid of something personal, such as being outwitted or replaced by AI? Or was it solely a matter of caring that their students learned the material on their own and engaged their own faculties of say, imagination? It is difficult to say.

Discussion

Two main themes arose from the qualitative findings: engagement and preparedness. Users seemed to struggle with how to engage with the platforms themselves. In an educational technology world of many options, students seem to relish the availability of different platforms to find answers. Video-based platforms like YouTube provide one level of access to knowledge. But unless there is a specific video on the exact STEM problem a professor has assigned in class or for homework, students must still 'think for themselves' to connect the example content to their understanding of the assigned problem. Or perhaps they might encounter roadblocks when framing the problem in order to even know what to look for. Thus, as Eddy and Jordan alluded, the problem may not have been with OLC at all, but rather with something much simpler: mathematical comprehension.

While educational technology has been around for quite some time, there is still a dearth of research on how students' preferences, choices, and behaviors with various platforms impacts their learning strategies and achievement. The lack of replication studies hinders our ability to make better inferences, as different tools have different capabilities, and one tool's hype seems to replace another before it has enough time to be adequately studied and understood. It is arguably reasonable that students could employ GLS approaches like mapping or self-testing *via* platforms such as OLC. In fact, Maggie seemed to have been using the OLC self-testing function to some benefit. This could well change how we model training students to use such systems by viewing them not only as strategic mediators to learning outcomes but also as collaborators.

While not intended to be a deep-dive analysis, this study echoes Stolk, Gross, and Zastavker's (2021) call for more investigation into how motivation and pedagogy interact to influence STEM learning outcomes. Wang, Peng, and Wang (2022) piloted a promising approach to helping 7th grade students regulate their learning through directed scaffolding training. Scaling this approach at GFU could improve faculty and students' abilities to reach their course achievement goals. A major part of the scaffolding approach could include faculty actively demonstrating to students how to prompt AI tools to improve their conceptual and procedural knowledge. But this assumes that the students and instructors are consistently using the same tool to replicate best practices for engaging with it, as different tools have different purposes.

One of the study co-PIs is the institution's Director of Learning Support Services and has offered short informational sessions at lunchtimes to help university members learn about and engage with different AI tools. This might be a boon, increasing students' autonomy and competence as recommended by Deci and Ryan's (2017) self-determination theory. Or it might be deflating in the long run, increasing cognitive load on which tools are available and how to navigate them. Furthermore, navigating an AI platform can often have little to do with learning a specific subject. But the flip side, a wild wild west of 'do as you please', sounds worse. Hence, regardless of whether GFU continues with OLC, engaging and preparing students and faculty through continuing education should become a staple and consistent practice. Over time, the institution might take a train-the-trainer approach to help faculty members learn and adapt how they embed two or three AI tools into their courses.

Limitations to this study included its observational nature, due to the self-selection students exhibited in deciding to sign up for and use OLC. Future studies might employ more rigorous quasi-experimental approaches. They might restrict access to a certain tool and compare students using one platform with (matched pairs of) students in those same STEM courses (but different sections) who do not use the platform. This might be more difficult to accomplish since it is challenging to restrict and track students who nonetheless access a tool outside of class on their own. So, it might help if the instructional design is tailored to force the student to demonstrate their work within a specific AI platform. This does not alleviate the concerns from Eddy and Maggie around incorrect AI responses. But perhaps one of the things that students might be missing, to which Jordan hinted, is that understanding mistakes—whether one's own or someone else's, including AI's—can be a valuable part of the learning process.

That can only come from being adequately prepared and engaging deeply with subject matter regardless of the platform.

Another limitation pertains to who responded for interviews. Eddy and Maggie are MECS students. After multiple attempts, we failed to get any SOCS or NATS majors to agree to an interview. So, it is difficult if not impossible to say why SOCS or NATS students' OLC usage was significantly lower in pages viewed and sessions without any contextual evidence. The same can be said for the lack of statistically significant observed differences by first-gen status. Neither Eddy nor Maggie self-identified as first-gen students. Follow up studies require more intense recruitment of students from such underrepresented groups, though we suspect historical issues with equitable access and opportunity lurk behind these results.

Future phases might also include faculty from non-Math or Engineering departments, including but not limited to natural and social sciences. Although mathematics and science have a longstanding partnership, the abstract nature of mathematics content might be different from, and hence more difficult, than some of the more concrete concepts encountered in other science courses that merely incorporate mathematics. Self-selection bias is arguably the biggest obstacle to fairly interpreting these results. Not only do students self-select into GFU, but for this study they willingly chose to register for and continue using OLC. Thus, they might have had an existing bias or natural curiosity towards AI platforms, and will likely continue to, even if professors mandate or restrict their use.

Overall, the findings indicate that students did not use OLC to their advantage. Instead, it was accessed by a relatively small proportion of users (both faculty and students) and for many of them usage consistently declined over the course of the Fall 2024 semester. If the institution desires to partner with AI education companies in the future, it might be helpful to first conduct a needs assessment and perceptions of competing platforms' pros and cons. AI may be here to stay but we suspect that until much of the hype around it calms down, or there is a sustained and effective cultural habitus for how to use them to mediate the learning process, students will continue to jump from one platform to another. Motivations for choosing and using AI platforms can often be misguided, and instructors might be neglecting or underestimating how to responsibly teach with AI. We sought to simply ascertain some observations of how TRAD students use just one of a plethora of available tools. In a fast and frugal way, we got a rough picture of the landscape, which is often exactly what formative assessment provides.

Acknowledgments

We sincerely thank the editors and reviewers at *Intersection* for their invaluable feedback and considerate advice. We also thank Tyler Susmilch, GFU's Executive Director of Analytics & Data, for providing merged data sets and answering methodological questions along the way.

References

Ambrose, S., Lovett, M., Bridges, M. W., DiPietro, M., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.

Botnaru, D., Orvis, J., Langdon, J., Niemiec, C. P., & Landge, S. M. (2021). Predicting final grades in STEM courses: A path analysis of academic motivation and course-related behavior using self-determination theory. *Learning and Motivation*, *74*, 101723-. https://doi.org/10.1016/j.lmot.2021.101723.

Chen, H. T. (2014). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2nd ed.). Sage.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1–24. https://doi.org/10.1007/bf00156001.

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding*. Cambridge University Press.

Google (2025). *Gemini AI (PRO)* [Large Language Model].

Gertz, N. (2024). *Nihilism and technology*. Bloomsbury Publishing.

Hecht, C. A., Harackiewicz, J. M., Priniski, S. J., Canning, E. A., Tibbetts, Y., & Hyde, J. S. (2019). Promoting persistence in the biological and medical sciences: An expectancy-value approach to intervention. *Journal of Educational Psychology*, *111*(8), 1462–1477. https://doi.org/10.1037/edu0000356.

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, *15*(1), Article 17458. https://doi.org/10.1038/s41598-025-97652-6.

Khan, S. (2024). *Brave new worlds: How AI will revolutionize education (and why that's a good thing)*. Viking Press.

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R. J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233. https://doi.org/10.1007/s11412-018-9277-y.

Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory physics course? *Physical Review Physics Education Research*, *19*(1), Article 010132. https://doi.org/10.1103/PhysRevPhysEducRes.19.010132.

Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. SAGE Publications.

Mollick, E. (2024). *Co-intelligence: living and working with AI*. Portfolio/Penguin.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, *106*(1), 315–329. https://doi.org/10.1037/a0034027.

Qudrat-Ullah, H. (2025). *Navigating complexity: AI and systems thinking for smarter decisions*. Springer Nature.

Reigeluth, C., Beatty, B. J., & Myers, R. D. (2017). *Instructional-design theories and models: The learner-centered paradigm of education (Volume IV)*. Routledge.

Renninger, K. A., & Hidi, S. E. (2019). *The Cambridge handbook of motivation and learning*. Cambridge University Press.

Richey, R., Klein, J. D., & Tracey, M. W. (2011). *The instructional design knowledge base: Theory, research, and practice*. Routledge.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. The Guilford Press.

Schunk, D., & Greene, J. A. (2018). *Handbook of self-regulation of learning and performance* (2nd ed.). Routledge / Taylor & Francis Group.

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

Stolk, J. D., Gross, M. D., & Zastavker, Y. V. (2021). Motivation, pedagogy, and gender: examining the multifaceted and dynamic situational responses of women and men in college STEM courses. *International Journal of STEM Education*, *8*(1), 1–19. https://doi.org/10.1186/s40594-021-00283-2.

Suskie, L. (2018). *Assessing student learning: A common sense guide* (3rd Ed.). Wiley.

Thaler, R., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness* (Revised and expanded ed.). Penguin Books.

Wang, Q., Peng, Y., & Wang, H. (2022). A curation activity-based self-regulated learning promotion approach as scaffolding to improving learners' performance in STEM courses. *Journal of Educational Computing Research*, *60*(4), 843–876. https://doi.org/10.1177/07356331211056532.

Young, A. M., Wendel, P. J., Esson, J. M., & Plank, K. M. (2018). Motivational decline and recovery in higher education STEM courses. *International Journal of Science Education*, *40*(9), 1016–1033. https://doi.org/10.1080/09500693.2018.1460773.

Appendix A

General Information

1. Please tell me a little bit about your major field of study and why you chose that one.
2. What STEM class did you take last semester? What was your motivation for taking this class?
3. What's your experience with STEM classes? Do you usually succeed/struggle at STEM? Why?

Learning Strategies Focus

1. Generative learning strategies include things like mapping concepts, drawing, using your imagination or thought experiments, self-testing, retrieval practice, elaboration, and explaining things to others. Do you use any of these in class or perhaps while studying for class? If so, can you describe your process using them?
2. Do or did you ask questions in class, whether to the instructor or peers? If so, how often do you ask questions? What are they typically about? Does it lead to meaningful conversations?
3. Do or did you set goals in class for yourself—whether performance-based or motivational? What tends to be successful or unsuccessful?

OLC Use Focus

1. Can you tell me why you decided to register for an account with OLC?
2. Do you recall any specific reasons that prompted you to use OLC at a specific time?
3. When did you typically use OLC, e.g., before a quiz/test, after class, during moments of confusion, or otherwise? Can you tell me why?
4. What are your general perceptions of OLC? Anything specific stand out?
5. Which functions/aspects of OLC did you find most and least useful? Why?
6. Can you recall and share a time when you solved a problem, completed a class assignment, or achieved a certain score on a quiz/test and thought, "OLC made the difference!"?
7. Can you share any experience or reasons you might have to believe that your classmates also used/liked/disliked OLC?
8. Do you think OLC contributed to your success/struggle in your STEM course? Why or why not? Since that is the case, would you continue to use it, or do/would you use other sources?

About the Authors

Dane C. Joseph, Professor of Education, George Fox University, djoseph@georgefox.edu
Rick Muthiah, Director of Learning Support Services, George Fox University, rmuthiah@georgefox.edu