



A Validation of AI-Enabled Discussion Platform Metrics and Relationships to Student Efforts

Audon Archibald¹ · Cassie Hudson² · Tania Heap¹ · Ruthanne “Rudi” Thompson¹ · Lin Lin¹ · Jaqueline DeMeritt¹ · Heather Lucke¹

Accepted: 18 December 2022 / Published online: 21 January 2023
© Association for Educational Communications & Technology 2023

Abstract

Asynchronous discussions are a popular feature in online higher education as they enable instructor-student and student-student interactions at the users’ own time and pace. AI-driven discussion platforms are designed to relieve instructors of automatable tasks, e.g., low-stakes grading and post moderation. Our study investigated the validity of an AI-generated score compared to human-driven methods of evaluating student effort and the impact of instructor interaction on students’ discussion post quality. A series of within-subjects MANOVAs was conducted on 14,599 discussion posts among over 800 students across four classes to measure post ‘curiosity score’ (i.e., an AI-generated metric of post quality) and word count. After checking assumptions, one MANOVA was run for each type of instructor interaction: private coaching, public praising, and public featuring. Instructor coaching appears to impact curiosity scores and word count, with later posts being an average of 40 words longer and scoring an average of 15 points higher than the original post that received instructor coaching. AI-driven tools appear to free up time for more creative human interventions, particularly among instructors teaching high-enrollment classes, where a traditional discussion forum is less scalable.

Keywords Discussion board · Feedback · Pedagogy · Artificial intelligence · Praising · Coaching

Introduction

Widely used in both online and in-person courses, asynchronous discussion boards have been touted as a key pedagogical tool in encouraging students to engage with, and develop, the critical thinking and writing skills required for university level success (Aloni & Harrington, 2018). Within these discussion platforms, students engage with one another and their instructor in text-based environments that ask students to respond to prompts or offer commentary on topics. Topics are often generated by instructors or other students and are left in open forums available for other students to review and comment on. Particularly for online courses, where it is more difficult for students to establish a social presence for themselves, discussion boards provide a platform for students to leave their ‘footprint’ upon the course in open

displays of content application in the form of posts. These posts are often used as graded assignments by instructors, with students being evaluated based on metrics such as participation frequency or contribution length (Dennen, 2008).

Asynchronous online discussions have been employed as a key role-player in facilitating successful student growth and engagement across disciplines (Dixson, 2010; Salter & Conneely, 2015). These discussion boards offer unmatched flexibility when compared to face-to-face discussions by giving students greater control over their schedules and environments (Dahlstrom-Hakki et al., 2020; Wang & Woo, 2007). Further, increased convenience of student engagement may ‘even the playing field’ between students who are predisposed to succeed in the more extroverted context of live discussion and those who prefer distance learning (Abe, 2020). Engagement with online discussion boards has also been linked to overall student performance in corresponding courses (Cheng et al., 2011; Lee & Recker, 2021), with the core of this improvement thought to occur in two ways: 1) the mechanical act of entering one’s thoughts into an online portal vs. voicing them in person, and 2) the greater opportunity

✉ Audon Archibald
Audon.Archibald@unt.edu

¹ University of North Texas, Denton, TX, USA

² Present Address: Global Giving, Washington, DC, USA

both students *and instructors* have to engage in considered, meaningful interactions with one another that are less feasible in a synchronous space. Instructors can require a minimum number of posts in asynchronous boards that are then kept as artifacts for review and commentary, allowing instructors to gain insight into the thought processes of each student.

Despite the advantages of asynchronous boards, instructors have become increasingly aware that they are not a panacea. Asynchronous discussion boards present several hurdles to successful implementation. Students can draw very different conclusions about what constitutes a ‘successful’ post depending on their personal understanding of how best to engage with their instructor and peers via online discussion boards. For instance, some students define success in giving their best effort, while others define success as meeting the minimum requirements set by the instructor (Knowlton, 2005). Furthermore, given their ability to engage with asynchronous discussions on their own terms, many students alter their approach to discussion content in favor of time- and labor-saving strategies like skimming posts rather than reading them, which undermines content retention (Peters & Hewitt, 2010).

The platform on which instructors choose to host their discussions also mediates how students experience and participate. Students express preference for platforms that emulate social media experiences they may already be familiar with as opposed to the explicitly academic (Hurt et al., 2012). Students may also experience more difficulty in the absence of explicit protocols created by instructors that guide students on how to interact with the discussion platform and each other (Zydney et al., 2012), demanding an instructor’s attention, where another method of assessment might not. Although asynchronous discussions afford the *possibility* for instructors to engage in more meaningful one-on-one commentary with each student who contributes, it is rare that instructors, especially those with a heavy course load, can actualize that potential for every student without considerable support from teaching assistants (Wuttikietpaiboon, 2013).

Examples of additional barriers to successful instructor intervention in asynchronous discussions include lack of instructor engagement, which is a risk factor for poor student performance, or, in contrast, too much oversight and/or extensive instructor protocols may be considered invasive. Furthermore, if there are too few protocols, students may struggle to stick to a regular posting schedule or fully commit to best practices for asynchronous discussions. Considerable work has tested the influence of contextual factors such as discussion protocols (Zydney et al., 2012), gamification (Ding et al., 2018), and even the discussion platform itself (Hurt et al., 2012) on student success and engagement patterns.

Most Learning Management Systems (LMSs), such as Canvas and Moodle, include native online discussion tools that allow instructors to oversee online asynchronous discussions on their own schedules. Some platforms are specifically tailored to facilitate asynchronous discussions by offering a more specialized set of tools than standard LMS environments. One such tool is an asynchronous discussion platform supported by the presence of an artificial intelligence (AI) alongside gamified features that moderate both student discussion quality and instructor-student interactions.

While there is no consensus regarding what exactly constitutes an AI, one useable definition includes any “computing system that is able to engage in human-like processes such as learning, adapting, synthesizing, self-correction, and use of data for complex processing tasks,” (Popenici & Kerr, 2017). Under this definition, simple tools such as spell checkers can be classified as AI. Several modern developers seeking to offer AI-supported learning tools have emerged in the past decade, such as Apple’s virtual-assistant Siri, or the Duolingo language learning companion in 2012. However, despite the best efforts of developers to stay current in response to more usage data becoming available, several concerns have arisen from instructors and administrators over the prospect of using AI in educational contexts. Chief among these are concerns about harmful algorithmic biases as well as student data privacy (Shum et al., 2017). However, research on the role of instructors in minimizing these concerns and using AI-supported tools to facilitate discussions in general is insufficient. There are considerable gaps in understanding what benefits these tools may offer the modern classroom (Zawacki-Richter et al., 2019).

Similarly, much is still unknown about the time and labor tradeoffs that AI-supported tools can offer instructors. For instance, while long-term use of an AI-supported tool may yield tangible benefits for a classroom, time-consuming training required to use a novel tool effectively may prove too great a disruption to education. However, some entities (Bryant et al., 2020) have reported that AI tools save instructor’s time in K-12 classrooms. Tools that save time for instructors on administrative tasks, e.g., grading or identifying specific students who may benefit from individual attention, are most beneficial. Moreover, AI tools that reduce administrative load (which may constitute up to approximately 40% of an instructor’s time) may prove useful in freeing up instructor time for more meaningful interventions, such as giving in-depth constructive feedback and or other student/instructor interactions.

One such tool that provides a potential boon to instructors is the AI-supported asynchronous discussion platform Packback, which approaches online discussions using protocols based on the Socratic method of questioning (Elder & Paul, 1998). The platform integrates principles of AI and

gamification to increase student engagement and reduce the administrative time-commitment required to oversee online discussions. The platform features tools such as auto-moderation to filter out low-effort or inappropriate posts, automated newsletters and leaderboards that highlight student achievement and, perhaps most strikingly, the AI generation of a numerical representation of a student's effort in their discussion post, referred to as a 'curiosity score'. The curiosity score is calculated using a proprietary algorithm that includes weighted measures for different quality factors: depth; credibility and presentation, as determined by word count; sentence structure; citations and formatting. The possibility of an AI-generated score that indicates to instructors which students may need feedback to curb difficulties in asynchronous online discussions, *without having the need to grade them first*, stands out as a potential powerful asset worthy of investigation. This is especially true for high-enrollment courses in which instructors may derive more benefit from the time-saving features of the AI (Lantz et al., 2022).

In addition to the AI-supported features, the platform offers a layered system for instructors to provide feedback to students, including one-button vs. more effortful methods, and overt coding of whether feedback should be considered primarily positive (praising) or actionable (coaching) in nature. Coaching is presented as private feedback between the instructor and the student, while praising is public, available for other students in the platform to see. Features of this system also include the ability to 'feature' student posts, and archiving posts in a highly identifiable place in the platform as an example to other students of quality work. Coaching can help instructors establish a sense of teaching presence within the classroom, an essential element of the Community of Inquiry Model (COI), promoting student cognitive presence (Park et al., 2015) as well as perceived learning and satisfaction (Arbaugh, 2010).

In line with COI, the public nature of 'praising' or 'featuring' a post gives opportunity to instructors to more easily enact the role of the facilitator or co-creator of a social environment that favors active and successful learning (social presence) (Garrison & Anderson, 2003). The combination of discussion prompt, AI-mediated feedback, and visible public praise provides students with the directions to facilitate themselves, reducing some of the burden on the instructor (Zydney et al., 2012). Instructor-initiated interaction that is highly visible but sporadic may offer students room to engage in behaviors typically associated with high teaching presence (Park et al., 2015). With curiosity scores acting, in theory, as a way for instructors to better identify which students' posts may deserve positive or actionable feedback, instructors may have additional time to capitalize on prior research that suggests discussion coaching helps students facilitate higher-order thinking skills (Stein et al., 2013).

Schartel (2012) argues that feedback or coaching is an essential element of learning, including in online discussions (Rochera et al., 2021). According to Schartel (2012) good quality feedback should be specific, focused on knowledge, behaviors, or actions and not on the person. The credibility of the feedback provider has an influence on students' acceptance. In giving critical feedback, guidance that encourages student self-reflection can mitigate the possible emotional response to criticism. In online discussions, reinforcement feedback that is regularly provided by instructors also appears to aid students' understanding of the topics (Rochera et al., 2021).

Despite the existence of these tools, comparatively little research has been done in validating how different kinds of feedback do or do not effect student post patterns. Additionally, while the value of an AI-generated metric such as a curiosity score may have value to instructors as a quick, interpretable method of identifying student areas of success/improvement, little research has been done on whether a score appears to quantitatively represent what it intends to for real instructors in the field. The purpose of the present study is threefold, 1) validate if a third-party, AI-generated construct successfully corresponds to more traditional markers of student effort; 2) determine how different actions, both those that give positive vs. constructive feedback and have an instant vs. non-instant time commitment, taken by instructors toward student discussion posts affects the effort displayed in future student posts; and 3) identify if the patterns of action taken by instructors correspond in expected ways in *both* AI-generated and more traditional markers of student effort. Specifically, the present study seeks to answer the following research questions (RQs):

RQ1) Does a third-party AI-generated construct ('curiosity score') successfully correspond to an instructor-driven marker of student effort?

RQ2) How do patterns of instructor interaction impact students' effort in subsequent discussion posts?

Methods & Materials

Method & Participants

Conducted at a public research university in the Southern United States, this study investigated over 14,000 discussion posts on the discussion platform over 3 semesters, Fall 2019, Spring 2020, and Fall 2020. Over 800 students from 15 class sections (including, but not limited to, 1000 level biology, 2000 level research methods in psychology, 4000 level race psychology, and 4000 level investigations in human rights) consented to allowing their posts to be reflected in the 14,599 analyzed posts. While not reflected in the present

study, students in Fall 2019 and Spring 2020 semesters were also concurrent participants in another study directly comparing Packback to innate LMS supported discussion boards (Hudson et al., 2020). As participation in the present study involved only skimming existing data (discussion posts) from students and data extracted from the platform itself, students were not given formal surveys with demographic information.

The instructors involved in the research are three tenured professors, averaging two- decades of teaching experience. They have all used discussion forums extensively in online or in-person classes. The instructor of political science had prior experience teaching with AI-supported online discussions. The instructor teaching a high-enrollment biology course was supported by five teaching assistants (TAs), all of whom had previously completed an entirely online semester-long TA development program (Heap et al., 2020).

The platform creators provided the deidentified data packets for how students used the platform, which were compiled into a single dataset to analyze student posting patterns from week to week throughout the semester. Additionally, each post came individually flagged if the instructor/TA interacted with the post, and in what way. In doing so, students were evaluated at the individual level for how the effort metrics of their posts changed in the posts following the initial, instructor-tagged post. Specifically, researchers set flags for the next four posts following the initial post that were created by the student *after the instructor had given feedback*, marking the posts with a number from 0 (the initial) to 4. For example, if a student posted three times in one day and the instructor interacted with the post a week later, the next four posts the student made after the timestamp of the instructor would be considered the four posts of interest. In cases where instructors interacted with posts in a relatively short time, such as a student only having time to make two posts of interest before the instructor interacts with their posts again, the flag counter would reset. No post was counted more than once for purposes of evaluating changing effort metrics.

Markers of Student Effort

Within the discussion platform, students made posts in one of two ways. First, both instructors and students could pose open-ended questions, with supporting contextual information, which acted as the main ‘response chain’ of the asynchronous discussions. Typically, instructors posed a theme related to that week’s material and asked each student to come up with open-ended questions that explored that theme. Second, students posed their own answers to the questions of their peers. This pair of responses, a question and an answer, often fulfilled the minimum post requirement for students each week.

While providing each type of response, students were not held to strict posting standards. Evidence of effort, such as providing meaningful citations or a word count significant enough to properly elaborate upon an idea, were not required by instructors, but were encouraged by the platform itself through live estimations of the student’s potential curiosity scores. These curiosity scores are automatically publicly provided to each student’s question and answer, acting as an AI-generated representation of a student’s final effort of their discussion response. Scored on a scale from 1–100, a curiosity score was automatically calculated for each student post based on post length, sentence structure, lack of repetition, and successful use of citations. While this curiosity score is strictly not a grade (e.g., a student may receive a curiosity score of 70 but meet all the instructor-specific requirements for an A on that week’s posts), the score is intended to be a reliable method of reducing instructor workload by readily identifying students who may be devoting considerably more, or less, time to their discussion posts than their peers.

Though the usefulness of such a metric for instructors is apparent, some problems arise when considering that the score is based on the good faith of a third-party company’s AI. This is compounded by the fact that the formula for the generation of these scores is protected by the platform designers and not publicly available for review. Therefore, it is imperative that, if being used as a quantitative evaluation of student post effort, additional, more traditional metrics of effort are considered alongside curiosity scores. One metric of student effort that is quickly, quantitatively helpful is post word count, especially in absence of mandatory post-length minimums. While not an iron-clad representation of student effort, there is precedent to consider word count to be a useful heuristic representation of the amount of detail students devote to their posts (O’Brien & Baugh, 2013). Through checking how instructor input affects *both* metrics, not only do we arrive at a more holistic understanding of how input is related to student outcomes in their posts, but how the AI-generated curiosity score stands in contrast to other effort measurements.

Patterns of Instructor Interaction

In interacting with student posts, instructors are afforded three main courses of action: coaching, praising, and featuring student posts. When coaching posts, the platform asks instructors to provide private feedback to students on how the targeted post could be improved through edits, and to also give students advice on what to keep in mind for future posts. Praising posts, which are public, likewise asks instructors to provide two main lines of feedback to a targeted post, why the post is important for other students, and what the student did well in the targeted post that should be emulated in future posts.

The final action, featuring posts, differs from praising and coaching in the amount of time instructors invest in the action. Unlike praising and coaching, where instructors are asked to provide specific feedback that can be up to a paragraph in length, featuring is a one-button action that informs the class that a post is noteworthy. Within praising posts, instructors are offered an option to praise the post on the same page but can also complete the action outside of the praise user interface (UI) with one click. In all three methods of interaction, students receive an email and within-platform notification of the instructor action.

Results

Research Question 1) Validation of the AI Generated ‘Curiosity Score’

We investigated patterns of how the AI-generated curiosity score was related to overall patterns of student word count at the post level. To validate whether the curiosity score was meaningfully related to this more traditional method of evaluating student effort, a linear regression was conducted using post curiosity score ($m = 72.08, SD = 17.07$) as the independent variable and post word count ($m = 119.48, SD = 59.21$) as the dependent variable. Evaluating over 14,500 student posts ($N = 14,599$), overall regression results suggest a statistically significant effect [$F(1, 14,597) = 4505.93, p < 0.001$] with a robust effect size ($Adj. R^2 = 0.24$), suggesting that a post’s curiosity score helped explain nearly a quarter of the variability in post word count. Moreover, a one-point change in curiosity score (scored on a scale of 0–100) roughly corresponded to about a two-word increase in post wordcount ($\beta = 1.78, p < 0.001$), implying that the higher the post’s curiosity score, the more detail students were adding to their posts.

Research Question 2) Student Post Outcomes and Instructor Feedback

We evaluated how each type of instructor interaction was related to patterns of student effort in our metrics of interest, by conducting a series of within-subjects MANOVAs using the order of the flagged posts as the independent variable, and post curiosity score and word count as the dependent variables. After checking assumptions, one MANOVA was run for each type of instructor interaction: private coaching, public praising, and public featuring.

Coaching

The first MANOVA was conducted to investigate the effects of coaching on post curiosity score and word count from an initial, interacted post to the next four posts. MANOVA results

revealed an overall significant effect of coaching as related to both curiosity score [$F(4, 1096) = 23.09, p < 0.001$] and post word count [$F(4, 1066) = 12.34, p < 0.001$], suggesting a difference exists between the count of both metrics between the five posts on average. When we investigated MANOVA descriptives (Table 1) and ran follow-up post-hoc tests (Tukey’s HSD) we found that across *both* curiosity score and word count, the original flagged post coached by the instructor showed a statistically significant difference with all four follow-up posts. Interestingly, among the four follow-up posts, none showed statistically significant differences from each other, implying that, though there was a jump in both curiosity score and wordcount in posts following the initial coached post, the effect was stable, with no decline over the next several posts.

While it may be tempting to attribute this effect to students’ early posts simply being less effective than their follow-up posts by virtue of being either unfamiliar with the platform or simply needing some time to learn about instructor expectations, evaluations of overall student trends suggest that this is not the case. To better contextualize the present results, we evaluated all students (not just those who had posts that were coached, $n = 811$), by conducting an exploratory MANOVA using the first, and last, post students made during the semester to test if students naturally improved in their curiosity score or post wordcount over the semester. Results revealed that there was not a statistically significant difference in either curiosity score [$F(1,1611) = 1.91, p = 0.17$] or word count [$F(1,1611) = 3.07, p = 0.08$]. Descriptive evaluations of the average difference in a student’s first and last post for the semester likewise revealed a difference in less than two points, from 68.70 to 70.38, in student curiosity score, and about four words, from 112.40 to 116.51, in student wordcount. Though both metrics showed slight improvement for students on average, neither result was statistically significant or of a particularly large effect size. Students who had posts coached, comparatively, showcased an average increase of nearly 15 points in curiosity score and almost 40 words on average in their following four posts.

Table 1 Coached Post Descriptive Statistics

Post flag	Curiosity Score (m)	Word Count (m)	Curiosity Score (SD)	Word Count (SD)
0 ($n = 235$)	58.50***a	86.20** a	25.29	61.28
1 ($n = 235$)	73.06***b	115.70***b	15.57	49.79
2 ($n = 226$)	70.04***b	108.48***b	17.19	43.21
3 ($n = 221$)	71.06***b	110.10***b	16.38	42.08
4 ($n = 191$)	71.35***b	106.03***b	15.86	32.83

** indicates a statistically significant between-groups test at $p < 0.001$
 a,b groups of posts that are significantly different from one another (a is significantly different from b)

Table 2 Praised Post Descriptive Statistics

Post flag	Curiosity Score (m)	Word Count (m)	Curiosity Score (SD)	Word Count (SD)
0 (<i>n</i> = 168)	79.86	128.39	14.19	56.19
1 (<i>n</i> = 168)	78.55	138.51	11.92	65.45
2 (<i>n</i> = 166)	77.27	130.81	14.14	60.63
3 (<i>n</i> = 162)	77.50	130.27	13.36	64.72
4 (<i>n</i> = 142)	77.11	126.91	13.57	50.05

Praising

We conducted a second MANOVA on instructor praising posts using a similar approach to our coaching analysis. Post flags were used as the grouping variable and curiosity score and post wordcount as the dependent variables. MANOVA results suggested no overall effect in either curiosity score [$F(4,800)=0.91$, $p=0.56$] or wordcount [$F(4,800)=1.20$, $p=0.31$] for posts that followed an instructor's praise. In investigations of descriptives (Table 2), no two sets of posts revealed themselves to have noteworthy differences across either curiosity score or wordcount, with post-hoc tests returning no statistically significant result. Overall, results suggest that praising did *not* influence the amount of effort students put into their following posts. However, when evaluated in contrast to coaching results, we do see that posts instructors identified as worthy of praising scored considerably higher than initial posts instructors elected to coach, featuring a difference of almost 20 points in the curiosity score average and over 40 words in the average wordcount. In addition, while coached posts did show a jump in scores after instructor intervention, on average, these posts were not brought in line with the standards of students with praised posts. When considering the curiosity score as a marker, present results suggest that this score was ultimately related to the choice of effortful action (praising vs. coaching) instructors decided to take, with posts that were coached scoring considerably lower than posts that were praised, adding to the strength of the measurement as a tool for aiding instructors in identifying relevant post patterns.

Featuring

The final MANOVA on featuring was conducted using an identical analysis to praising and coaching. Here, there was an overall significant MANOVA effect for featured posts across both curiosity score [$F(4, 2062)=22.79$, $p<0.001$] and wordcount [$F(4, 2062)=8.95$, $p<0.001$]. However, when we examined descriptives for student post patterns following an instructor featuring their post, we saw statistically significant downticks in both curiosity score and wordcount (see Table 3). Post-hoc tests revealed that the initial, featured post came in at about 7 curiosity score points higher and 20

words longer than the following four posts (at $p<0.001$), with this initial post registering a statistically significant difference than each of the following four posts. The following four posts, in turn, were not statistically different from each other.

Although an initial read might suggest that perhaps featuring a post may inspire students to devote less effort on following posts, we see that posts that were featured as well as the posts that followed them typically earned higher curiosity scores and wordcounts than general student posting averages. When compared directly to praised posts, featured posts appear to be of even higher quality, with the average featured post showcasing similar amounts of improvement to praised posts that followed the feature. Important contextual information is also relevant to note, though instructors had the option to feature praised posts, more than twice the number of posts were featured ($n=447$) than were praised ($n=168$). This suggests that instructors were more likely to make use of the ability to feature quality posts than they were to praise them, perhaps due to the ease of the one-button validation of the post's effort. However, though featured posts appeared to be of particularly high quality on average, even more so than praised posts, results suggest that this course of action did not inspire students to maintain that level of effort at a significant level in their following posts. While students with featured posts did maintain a level of post quality that appeared to be better than average, featuring posts likely served more as an indicator of an especially high-level post from already high-achieving student, who appeared to revert to a lower, but still relatively high achieving mean after their post was featured.

Discussion and Future Directions

The purpose of this study was to determine the influence of an artificial intelligence (AI)-driven discussion platform on student discussion efforts and quality of different types of instructor-student interaction. More specifically, with concerns over AI replacing human action (Nedelkoska &

Table 3 Featured Post Descriptive Statistics

Post flag	Curiosity Score (m)	Word Count (m)	Curiosity Score (SD)	Word Count (SD)
0 (<i>n</i> = 447)	84.79 ^{**a}	155.50 ^{**a}	11.41	77.69
1 (<i>n</i> = 447)	78.02 ^{**b}	138.26 ^{**b}	14.54	84.55
2 (<i>n</i> = 431)	77.64 ^{**b}	136.02 ^{**b}	14.88	60.53
3 (<i>n</i> = 402)	77.99 ^{**b}	129.17 ^{**b}	14.17	64.72
4 (<i>n</i> = 341)	77.23 ^{**b}	133.42 ^{**b}	14.18	50.05

** indicates a statistically significant between-groups test at $p<0.001$
^a_b groups of posts that are significantly different from one another (a is significantly different from b)

Quintini, 2018), this study intended to investigate whether an AI-driven discussion platform provides the opportunity for effective human intervention. Additionally, we sought to understand how an AI-generated score compares to more traditional metrics of post effort.

Our findings indicate that instructor coaching appears to positively impact both curiosity scores (i.e., the platforms' AI-generated metric of post quality) and word count, with later posts being longer and scoring higher than the original post that received instructor coaching. This effect was not limited to only students who posted early in the semester, with instructors coaching all throughout a semester showing similar levels of improvement in posts that followed. Instructor praising and featuring, on the other hand, did not appear to have an overall positive effect on the amount of effort students put into their following posts. However, posts instructors identified as worthy of praising scored significantly higher than initial posts instructors elected to coach. This suggests some validity to the AI-driven measurement as a tool in assisting instructors to assess post quality (i.e., there was a high overlap between posts the AI flagged as high quality and posts the instructors elected to publicly praise).

One possible explanation for this effect is that posts coached by instructors had more room to grow than posts that were already high-quality. Students at the highest-level curiosity score may not have much room for improvement in their already high-level posts and were simply encouraged to keep up the effort rather than improve on it. Posts elected by instructors as worthy of being featured appeared to be of even higher quality than praised posts, coming in at about seven curiosity score points higher and 20 words longer than following posts. Yet, our findings suggest that featuring did not inspire students to maintain that significant level of effort over the next several weeks. Again, this may be due to the students already providing an unusually high level of effort in featured posts. Given that following posts often actually fell significantly in terms of both curiosity score and word count, it is also possible that these featured posts captured students at the outliers of their effort, with many regressing to a more reasonable effort 'mean' in the following weeks.

Formalizing our results into instructor recommendations, beyond the studied platform, the actions of praising, coaching, and featuring can be essentialized into their core components. First, in terms of coaching, our results suggest that targeted, one-on-one digital feedback with students can drive students who may be struggling in their online discussion to achieve at least an adequate (passing) level of effort in their posts, with this effect persisting over the student's next several submissions. Compare that to praising, or the act of providing public feedback that is not necessarily constructive, but congratulatory

in nature. While students may appreciate the gesture, the present results suggest that instructor time may be better served elsewhere (such as with the low-scoring students) if the sole goal of instructor feedback is to improve student performance. Finally, in featuring, or the act of publicly indicating a post is worthwhile but NOT accompanying it with any sort of feedback, may also not be a high priority for instructors who wish to devote their time to students who may benefit more from intervention. Unlike praising, though, these acts that give kudos to a student via a simple 'button click' may still be undertaken by a busy instructor, thus interacting with students who may not need intervention to improve their performance but may appreciate the sentiment, with minimal investment of instructor time due to the immediacy of the effect.

The use of AI can further time saved for instructors in other ways, such as using curiosity scores as a means of quickly flagging posts in need of intervention and auto-moderating posts that would have needed intervention. Prior studies have demonstrated that feedback in the form of coaching can help students increase higher-order thinking skills (Stein & Wanstreet, 2020; Stein et al., 2013). Coaching and feedback also help establish a sense of teaching presence, a core element of the Community of Inquiry model (COI) (Garrison & Anderson, 2003) which is associated with improved cognitive presence (Park et al., 2015) and students' perceived learning and satisfaction (Arbaugh, 2010). Part of what drove improvement for coached students may be the feedback integrating the instructor into the online discussion community. Additionally, in using social media-like tools and learning technologies, knowledge is individually generated but socially mediated. With the present platform, students formulate their own questions, moving a step beyond traditional discussion protocols, enabling students to discuss personally meaningful topics with one another, under the guidance of an instructor. Looking back at the COI model (Garrison & Anderson, 2003), the AI appears to be designed to partly replace a few more mundane aspects of human intervention such as grading (Teaching Presence) and moderating (Social Presence), which in turn plays a role in influencing student learning and writing skills (Cognitive Presence), efforts which are reflected in the displays of effort by students when coached by the instructor.

The ability for an AI-driven tool to enhance human intervention and free up time for more creative activities appears to be well received particularly among instructors teaching high-enrollment classes, for whom a traditional discussion forum is challenging to manage at scale (Smith, 2019). For one of our current instructors, a biology professor, the use of AI-driven assistance was a noted positive compared to previous, LMS based discussions. Online Contemporary Biology for non-majors is a required core course. Often students are not enthusiastic about participating in the topic.

The AI-driven platform created a patterned change not only in the quantity and quality of the discussion posts, which were vast improvements from previous semesters which used a Canvas based platform, but also in the quality of their writing assignments. Whereas students in previous semesters struggled to include meaningful citations in their assignments, improvement was noted in the semesters which included the AI-driven discussions.

A limitation to consider is that two of the three studied semesters were in the middle of the COVID-19 pandemic. Results show very similar trends across semesters on the impact of individual coaching and public featuring and praising of posts. However, students with pre-existing housing insecurity and poorer health reported greater disruption to their learning (Bartolic et al., 2022), which also reflected into students' reading (Domingue et al., 2021), writing, and general digital literacy skills (Indrajit & Wibawa, 2020). A follow-up study could look into subsequent semesters post-pandemic and lockdown measures and draw comparisons with AI-driven interventions pre-pandemic, to investigate the extent to which AI may or may not assist instructors and students in developing digital literacy skills efficiently and in disruptive contexts.

The AI takes the burden of grading and moderating off the instructors, freeing up time for human-driven interventions such as providing feedback, which this study indicates being impactful for students' writing skills. However, this study did not compare time spent engaging in AI-supported platforms vs. traditional platforms. A follow up mixed-methods study could observe time spent by instructors in traditional vs. AI-supported discussion boards, as well as collect instructors' self-assessment of time spent, to determine if AI-supported initiatives do indeed save instructors' time compared to non-AI-supported.

Further exploration of the impact of AI-driven interventions on student learning and quality of instructor-student interactions could also include case studies from other subject disciplines that we have not tapped into, as well as narrowing efforts onto unrepresented or more diverse student demographics, given the dearth of evidence-based recommendations in AI-driven STEM education applicable to a diverse student body (Skowronek et al., 2022) and concerns over bias and equity in AI implementations (E. Gilbert, 2021). Furthermore, while no direct effect of praising or featuring was observed on the students who submitted the posts, the present study does not account for the public nature of praising or featuring on the posts of other students. While the present data suggests that praising and featuring are ineffective ways of increasing an individual student's effort, the public nature of these interactions may provide a blueprint for quality that other students may follow. Future studies may do well to account for the public vs. private nature of feedback on student post efforts. Student posting

effort was also only tracked for four follow-up posts to the initial post. Four posts may have been as little as one week of work. As such, the effect of instructor coaching beyond an arbitrary post cut-off, or perhaps indeed outside the discussion platform, would be a prudent next direction. Other directions include investigating the impact of instructor-driven interventions that include a wider instructor demographic, from full-time instructors to adjuncts and teaching assistants (TAs).

Data Availability The data that support the findings of this study are available from the corresponding author, A.A., upon reasonable request.

Declarations

Conflicts of Interests The authors declare no affiliation or involvement with any organization or party of financial interest that presents a conflict of interest with the present research.

References

- Abe, J. A. A. (2020). Big five, linguistic styles, and successful online learning. *The Internet and Higher Education*, 45. <https://doi.org/10.1016/j.iheduc.2019.100724>
- Aloni, M., & Harrington, C. (2018). Research based practices for improving the effectiveness of asynchronous online discussion boards. *Scholarship of Teaching and Learning in Psychology*, 4(4), 271–289. <https://doi.org/10.1037/stl0000121>
- Arbaugh, J. B. (2010). Sage, guide, both, or even more? An examination of instructor activity in online MBA courses. *Computers & Education*, 55(3), 1234–1244. <https://doi.org/10.1016/j.compedu.2010.05.020>
- Bryant, J., Heitz, C., Sanghvi, S., & Waple, D. (2020). *Artificial intelligence in education: How will it impact K-12 teachers* | McKinsey. Retrieved April 10, 2020 from <https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers>
- Bartolic, S., Matzat, U., Tai, J., Burgess, J. L., Boud, D., Craig, H., Archibald, A., De Jaeger, A., Kaplan-Rakowski, R., Lutze-Mann, L., Polly, P., Roth, M., Heap, T., Agapito, J., & Guppy, N. (2022). Student vulnerabilities and confidence in learning in the context of the COVID-19 pandemic. *Studies in Higher Education*, 47(12), 2460–2472. <https://doi.org/10.1080/03075079.2022.2081679>
- Cheng, C. K., Paré, D. E., Collimore, L.-M., & Joordens, S. (2011). Assessing the effectiveness of a voluntary online discussion forum on improving students' course performance. *Computers & Education*, 56(1), 253–261. <https://doi.org/10.1016/j.compedu.2010.07.024>
- Dahlstrom-Hakki, I., Alstad, Z., & Banerjee, M. (2020). Comparing synchronous and asynchronous online discussions for students with disabilities: The impact of social presence. *Computers & Education*, 150. <https://doi.org/10.1016/j.compedu.2020.103842>
- Dennen, V. P. (2008). Looking for evidence of learning: Assessment and analysis methods for online discourse. *Computers in Human Behavior*, 24(2), 205–219. <https://doi.org/10.1016/j.chb.2007.01.010>
- Ding, L., Er, E., & Orey, M. (2018). An exploratory study of student engagement in gamified online discussions. *Computers &*

- Education*, 120, 213–226. <https://doi.org/10.1016/j.compedu.2018.02.007>
- Dixon, M. D. (2010). Creating effective student engagement in online courses: What do students find engaging? *M. D.*, 10(2), 13.
- Domingue, B. W., Hough, H. J., Lang, D., & Yeatman, J. (2021). Changing Patterns of Growth in Oral Reading Fluency during the COVID-19 Pandemic. *Policy Analysis for California Education*, PACE.
- E. Gilbert, J. (2021). Equitable AI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–2). Association for Computing Machinery. <https://doi.org/10.1145/3411763.3457780>
- Elder, L., & Paul, R. (1998). The role of socratic questioning in thinking, teaching, and learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 71(5), 297–301. <https://doi.org/10.1080/00098659809602729>
- Garrison, D. R., & Anderson, T. (2003). *E-Learning in the 21st Century: A Framework for Research and Practice* (0 ed.). Routledge. <https://doi.org/10.4324/9780203166093>
- Heap, T., Hudson, C., & Archibald, A. (2020). Investigating the impact of an AI-driven discussion platform on educator perceptions and feedback. *EDEN Conference Proceedings*, 1, 127–136. <https://doi.org/10.38069/edenconf-2020-ac0010>
- Hudson, C., Archibald, A., & Heap, T. (2020). Integrating an ai-driven discussion platform: The impact of platform on engagement and quality. *EDEN Conference Proceedings*, 1, 117–126. <https://doi.org/10.38069/edenconf-2020-ac0009>
- Hurt, N. E., Moss, G., Bradley, C., Larson, L., Lovelace, M., Prevost, L., Riley, N., Domizi, D., & Camus, M. (2012). The ‘facebook’ effect: College students’ perceptions of online discussions in the age of social networking. *International Journal for the Scholarship of Teaching and Learning*, 6(2). <https://doi.org/10.20429/ijstol.2012.060210>
- Indrajit, R. E., & Wibawa, B. (2020). Portrait of higher education in the covid-19 period in a digital literacy perspective: a reflection on the online lecture process experience. In *2020 Fifth International Conference on Informatics and Computing (ICIC)* (pp. 1–5). IEEE.
- Knowlton, D. S. (2005). A taxonomy of learning through asynchronous discussion. *Journal of Interactive Learning Research*, 16(2), 155–177.
- Lantz, J. L., Liu, J. C., & Basnyat, I. (2022). *Piloting Artificial Intelligence (AI) to Facilitate Online Discussion in Large Online Classes: A Case Study* [Chapter]. Cases on Innovative and Successful Uses of Digital Resources for Online Learning; IGI Global. <https://doi.org/10.4018/978-1-7998-9004-1.ch009>
- Lee, J.-E., & Recker, M. (2021). The effects of instructors’ use of online discussions strategies on student participation and performance in university online introductory mathematics courses. *Computers & Education*, 162, 104084. <https://doi.org/10.1016/j.compedu.2020.104084>
- Nedelkoska, L., & Quintini, G. (2018). Automation, skills use and training. *OECD*. <https://doi.org/10.1787/2e2f4eea-en>
- O’Brien, S., & Baugh, V. P. H. (2013). Effects of minimum word counts on writing tasks. *Modern Psychological Studies* 7, 11–15.
- Park, J. H., Schallert, D. L., Sanders, A. J. Z., Williams, K. M., Seo, E., Yu, L.-T., Vogler, J. S., Song, K., Williamson, Z. H., & Knox, M. C. (2015). Does it matter if the teacher is there?: A teacher’s contribution to emerging patterns of interactions in online classroom discussions. *Computers & Education*, 82, 315–328. <https://doi.org/10.1016/j.compedu.2014.11.019>
- Peters, V. L., & Hewitt, J. (2010). An investigation of student practices in asynchronous computer conferencing courses. *Computers & Education*, 54(4), 951–961.
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1). <https://doi.org/10.1186/s41039-017-0062-8>
- Rochera, M. J., Engel, A., & Coll, C. (2021). The effects of teacher feedback: a case study of an online discussion forum in Higher Education. *Revista de Educación a Distancia (RED)*, 21(67). <https://doi.org/10.6018/red.476901>
- Salter, N. P., & Conneely, M. R. (2015). Structured and unstructured discussion forums as tools for student engagement. *Computers in Human Behavior*, 46, 18–25. <https://doi.org/10.1016/j.chb.2014.12.037>
- Schartel, S. A. (2012). Giving feedback—An integral part of education. *Best Practice & Research Clinical Anaesthesiology*, 26(1), 77–87.
- Shum, S. B., Sándor, Á., Goldsmith, R., Bass, R., & McWilliams, M. (2017). Towards reflective writing analytics: rationale, methodology and preliminary results. *Journal of Learning Analytics*, 4(1), 58–84. <https://doi.org/10.18608/jla.2017.41.5>
- Skowronek, M., Gilberti, R. M., Petro, M., Sancomb, C., Maddern, S., & Jankovic, J. (2022). Inclusive STEAM education in diverse disciplines of sustainable energy and AI. *Energy and AI*, 7, 100124. <https://doi.org/10.1016/j.egyai.2021.100124>
- Smith, T.W. (2019). Making the Most of Online Discussion: A Retrospective Analysis. *The International Journal of Teaching and Learning in Higher Education*, 31, 21–31.
- Stein, D. S., & Wanstreet, C. E. (2020). E-coaching success strategies for synchronous discussions. *Distance Learning*, 17(4), 113–118.
- Stein, D. S., Wanstreet, C. E., Slagle, P., Trinko, L. A., & Lutz, M. (2013). From ‘hello’ to higher-order thinking: The effect of coaching and feedback on online chats. *The Internet and Higher Education*, 16, 78–84. <https://doi.org/10.1016/j.iheduc.2012.03.001>
- Wang, Q., & Woo, H. L. (2007). Comparing asynchronous online discussions and face-to-face discussions in a classroom setting. *British Journal of Educational Technology*, 38(2), 272–286. <https://doi.org/10.1111/j.1467-8535.2006.00621.x>
- Wuttikietpaiboon, K. (2013). *Engaging graduate students in rich asynchronous online discussions* (2013–99090–524; Issues 11-A(E)) [ProQuest Information & Learning]. Retrieved January, 2021 from <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-99090-524&site=ehost-live>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zydney, J. M., deNoyelles, A., & Kyeong-Ju Seo, K. (2012). Creating a community of inquiry in online environments: An exploratory study on the effect of a protocol on interactions within asynchronous discussions. *Computers & Education*, 58(1), 77–87. <https://doi.org/10.1016/j.compedu.2011.07.009>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.