
INVESTIGATING THE IMPACT OF AN AI-DRIVEN DISCUSSION PLATFORM ON EDUCATOR PERCEPTIONS AND FEEDBACK

*Tania Heap, Cassie Hudson, Audon Archibald, University of North Texas,
United States of America*

Abstract

Asynchronous discussions are a popular element in online education, often used to replace the student-student and educator-student interactions that occur in physical classrooms. Most Learning Management Systems (LMSs), such as Canvas and Moodle, include native discussion tools that offer students and instructors the ability to interact and communicate at their own time and pace. Some specialized products offer a wider variety of discussion features that standard LMS tools may not have. In this study, we discuss the impact of an online discussion platform powered by artificial intelligence (AI) and gamified features. We conducted a qualitative study examining how AI elements affect the workflow, grading, and feedback experience of educators. Preliminary findings offer insights into the impact of the AI elements on graders' perceptions, in particular on their perception of giving feedback and the perceived ambiguity of the way the AI assesses students' post quality.

Perspectives of Artificial Intelligence in Education

Educators are in the driving seat when it comes to determining which technology to use to facilitate a learning experience. When the learning experience is online discussions, there is a wealth of options to choose from. The purpose of this study is to document the educators' experience using an artificial intelligence (AI) driven discussion tool, as opposed to a native discussion tool in Learning Management Systems (LMS) like Canvas or Blackboard. In this section, we are discussing what we know about applications of AI in education and its benefits and challenges to students and educators.

We can define AI broadly as "computing system that are 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; p.2). This is an inclusive definition, which allows less complex AI-driven tools to be classified as "artificial intelligence". AI is not a new concept, having been first coined and defined by John McCarthy in 1956 (Russell

& Norvig, 2003), but AI-driven applications have become more popular in education only in the latter part of the last decade. Several companies offering AI-driven learning applications were established in the last decade (e.g., Duolingo in 2012, Packback in 2014) and they have been refining their products as more and more data becomes available about usage and learning analytics.

Despite the growth in popularity, there is still a dearth of empirical studies on AI in education (AIEd) applications, particularly in higher education and in high-enrolment classes, and on their impact on student learning and educator workflow.

The successful integration of educational technology in courses is wholly dependent on the educators of these courses who choose when and how to integrate tools into a course. AI-driven tools, such as Packback, are no exception. Despite this, the educator perspective of AIEd is understudied—to the point where the title of a recent article by Zawacki-Richter et al. (2019; p.1) asks “...where are the educators?” while lamenting the “lack of critical reflection of challenges and risks” of AIEd. According to the authors, this oversight is likely due to the lack of research conducted by education faculty – of 146 articles reviewed, only nine had first-authors from Education departments. The result is that the vast majority of study designs are quasi-experimental, testing the impact of an intervention on students, with little to no exploration of the challenges of adoption, implementation, or impact on educators. More in-depth research in this area would likely illuminate some of the concerns associated with the implementation of AIEd, such as the educator’s fear of replacement, bias, privacy concerns and data protection, each of which could contribute toward a resistance to adopt an AI-driven technology by both educators and students.

There is also a general ignorance of what constitutes AI in general and AIEd in particular. This causes some to recoil at the term, despite using both simple (e.g., spell checkers) and complex (e.g., virtual assistants) AI tools on a day-to-day basis. An alternative is viewing AI as an enhancement of human intervention. A January 2020 report by McKinsey & Co., focused on the impact of AI in K12 education, claims that AIEd saves teachers’ time. According to the report, only half of the teachers’ workload is actually spent interacting with students, while the rest is spent on repetitive administrative tasks, lesson preparation, and grading. Their research suggests that 20 to 40 percent of current teacher workload could be automatized, enabling teachers to focus more on activities that are more creative and socially engaging for them and their students.

Human intervention still occurs in AIEd but may just shift in type and timing. Educators may have to learn how to integrate and use the tool in their course, either independently or with support from vendors or support staff. In the long-term, an AI tool may save time, but instructors must invest time up front to use the tool successfully.

Using AIEd in Online Discussions

Our research investigates the impact of an AI-driven discussion platform, Packback, on student learning and educator workflow in online courses. Packback approaches online discussions using a protocol based on the Socratic method of questioning, assigning students a “curiosity score”, and creating a newsletter of featured posts. It includes automated moderation of inappropriate posts. The AI, based on length of posts, sentence structure, and presence of a source citation, amongst other variables, calculates grading and curiosity scores.

The platform has other features that mimic a social networking site and gamify the experience. Students can “spark” other students’ post, mimicking a “like” and they are ranked on a Learner Leaderboard, which relies on the curiosity score. Students can view their own ranking in the class on the Learner Leaderboard and can access the profiles of other students, which show a breakdown of this students’ participation and each of their contributions to the discussion throughout the semester.

With AI taking away the time spent grading and moderating posts, educators can focus on more meaningful interactions with students, like providing public praise or private coaching. This helps establish a sense of teaching presence, a core element of the Community of Inquiry model (COI) (Garrison & Anderson, 2003) associated with improved cognitive presence (Park et al., 2015) and students’ perceived learning and satisfaction (Arbaugh, 2010).

Fundamentally, asynchronous discussions derive meaning as a cognitive constructivist tool to promote learning. Cognitive constructivism emphasizes both the impact of the environment and cultural context in shaping knowledge, as well as valuing the learner’s individual characteristics. Using social media tools and learning technologies, including online discussions, knowledge is individually generated but socially mediated (Felix, 2005; Whitelock, 2010). With Packback, 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 educator (Garrison & Anderson, 2003).

Though Packback provides just-in-time feedback and rapid assessment, the benefits of the tool are largely purported to be for educators, rather than students despite the fact that Packback is a learner-facing tool (Baker & Smith, 2019). This trend of developing AI tools to support educators, rather than to improve student learning, is a recent shift according to Barshay (2020) who reflects on the aforementioned 2020 McKinsey & Co. report about the impact of AI on K-12 teachers. Barshay (2020; p.1) claims that the evidence that AI tools (like virtual tutors) improve student learning is “not strong” and that “technology use

at school sometimes hurts student learning”, citing the recent reports on the final impacts of Columbia University Teach to One program (Consortium for Policy Research in Education, 2019). While this evidence is from the K-12 sector, these studies are a glimpse into what is possible in higher education and set the tone for beliefs about the impact of AI on education generally.

There are some potential concerns with automatic grading, as well, which tends to be simplistic and runs the risk of students learning to “trick” the autograder. As Popenici and Kerr (2017; p.2) claim, while the possibilities of AIED are vast, “we have reasons to stay aware of the real limits of AI algorithmic solutions in complex endeavours of learning in higher education.”

In this paper, we will discuss how educators used Packback in their courses and their impression of both human and AI-initiated feedback. The research questions that this study aimed to address are:

- RQ1: What are the graders’ perception of an AI-driven discussion platform?
- RQ2: How does an AI-driven discussion platform influence grading and feedback?
- RQ3: How does an AI-driven discussion platform impact graders’ workflow?

Methods

This paper forms part of a broader research project at the University of North Texas (UNT) that uses mixed-methods and is investigating the impact of an AI-driven discussion platform on students’ learning and educators’ workflow. The focus in this presentation is on the qualitative portion of the study, which focused on researching the impact of the platform on the graders’ workflow. The research team is still collecting data from more participants. All data is from the Fall 2019 semester and data collection is ongoing for the Spring 2020 semester.

Recruitment and artifacts collected

We conducted focus groups with three teaching assistants (TA) facilitating a large online first-year undergraduate course in Biology and Environmental Sciences. The TAs taught the online course using Packback in Summer 2019 for eight weeks and in Fall 2019 for 16 weeks. The TAs served as the primary individuals responsible for grading and providing feedback to students. In Fall 2019, the course was divided into five online sections of 50 students each and one TA managed each section. Three out of five TAs consented to participate in this study. All participants had experience using both Packback and Canvas for online discussions.

Data was collected in the form of focus groups during the Fall 2019 teaching semester. The focus groups were administered three times during the teaching semester: one at the beginning of the semester, during the first week of class, to examine participants' prior experience with and perceptions of online discussions; one half-way through teaching using Packback; and one at the end of the semester. The focus groups were audio-recorded and transcribed using the NVivo Transcription service.

Data analysis techniques

The research team analysed the data from the focus groups using a combination of thematic and saliency analysis. Thematic analysis helps identify themes or patterns in data by using a coding scheme. It helps to derive the patterns most relevant to the research questions and the most recurrent in the participant sample (Braun & Clarke, 2006). Saliency analysis focuses on the saliency of the themes: some themes may not be as recurrent but of high importance and relevance in answering a research question or furthering understanding of the overall research goals (Buetow, 2010).

Our primary goal was to assess TAs' perceptions of online discussions generally, their perception of Packback as a tool for online discussion, and any changes to their perception that may have occurred because of using this tool. While there are many facets of these focus groups that we could discuss, for this presentation we will focus on two dimensions of an overarching theme related to artificial intelligence: perceived impact of feedback and ambiguity. The research team, through an inductive process, collaboratively determined the artificial intelligence theme and others, after a close reading of all focus group transcripts, using the three research questions as the lens through which data was interpreted.

After agreeing on the primary themes and dimensions, two members of the research team coded the transcripts using NVivo 12, allowing for the possibility of additional recurring and salient themes. After completing the coding process, we reviewed all new recurring and salient themes and any inconsistent themes and dimensions highlighted by NVivo's Coding Comparison feature, until we achieved a Cohen's Kappa for all dimensions of above 0.80 through negotiated agreement (Lavrakas, 2008).

Results

As mentioned toward the end of Methods, in this section we are focusing only on two dimensions from the overall artificial intelligence theme: TAs' perceived impact of feedback to students, and ambiguity. We include quotations from focus group interviews to illustrate the themes emerged. The three participants are referred to as PA, PB and PC.

Perceived Impact of Feedback: human-initiated vs. AI-powered

A recurring and salient pattern emerging from the data on all three TAs was the importance they placed on feedback in online discussions and its impact on student learning, both human TA-initiated feedback delivered via the platform and AI-powered feedback. In TA-initiated feedback, TAs used platform features such as private coaching and public praising, and “featuring” or “pinning” posts to recognize good work, which they perceive as aiding both student engagement and writing quality:

PA: “I think [giving feedback] helps [students] to umm make an effort and do better job. And when you tell them for example you did a great job, they try to participate more in the discussion.”

PA: “I prefer to guide them to umm, for example, elaborate their idea or umm I tell them you mention a great idea, but you need to explain more. Or sometimes, for example if they umm they didn’t write umm details or something or examples as I told...but if they write a good, I mean discussion I just tell them it’s a great idea, you did a great job or something like that.”

PC: “...the student’s name was being put out there, they’re like “ooh, I’m being publicly recognized in the class now.” So that is a positive good thing that the student would want.”

PA: “...sometimes I tell them, you did a great job and your post is a very good post and something like that, or sometimes I tell them, you have to elaborate your response. It’s better to write a better question or something like that. So both of these feedbacks can be helpful for our students in my opinion.”

The TAs also discussed the impact of feedback calculated by the AI itself, in the form of grading and curiosity points. TAs appear to be intrigued by the curiosity scores but are not always in agreement with how the AI algorithm grades or assigns curiosity points, leaving the TAs wondering what criteria are used for assessment, if these criteria are available to students or not, and whether grades are informed by curiosity points or not:

PB: “I want to see the criteria, like, I want to be able to click on the curiosity points and be able to see why they were assigned or why they weren’t assigned (...) If [students] could click on the curiosity points and see why they got what they got, I think that would be phenomenal.”

PB: “...some posts get really good curiosity points and some post get just a few. But those students still get full credit for those posts, which obviously isn’t fair to the other students that put in the work to get the full amount.”

As they try to make sense of how the scores are calculated by the AI, TAs are also concerned that their own feedback, delivered using their praising or coaching feature, may not be taken into account into the scores:

PC: "...so, is the praises and coaches that we as TAs give, is that also taken into consideration while it is being graded? (...) Does that give them extra points? Or is it just that...because, if it has to be graded properly, I would say that if a particular student has a praise from us that should give them some extra points. Because that means that there is no repetition of ideas and this is something that we actually liked."

The TAs' comments on their perceived uncertainty as to how grades and curiosity scores are calculated by the algorithm is one element of ambiguity, a theme discussed in more detail in the next subsection.

Ambiguity

TAs know that the AI grades and assigns scores based on length of post, sentence structure, and source citation, to name a few criteria. However, their experience interacting with the platform, reading students' posts and observing how these are scored by the AI, left them wondering to what extent each criterion weighs into the score, or if the AI places more emphasis on simple participation or engagement with the tool itself:

PC: "I'm not really sure, so in assessment, who exactly is grading it? I think right now it is just like the presence and absence thing and if the words are correct, but is anybody actually reading it?"

The TAs noted another ambiguous element. This stems from the structure of the platform, which differs from discussion boards in traditional LMSs. With discussions presented in one feed, TAs feel it may not make the posting instructions stand up:

PA: "Packback is awesome, but it is a little confusing for students because it has lots of...umm. The students can see the previous questions and they don't know if they should answer that question or the new one."

PC: "It's all going to be on one feed. So, the post – the student would argue with us that they did post – and they must have posted (...) but maybe they posted in a section that was already done and graded. So, they need to know where exactly they have to be posted."

Ambiguity can also be a consequence of students and TAs familiarizing with online discussions via a tool that is novel to them, because of the gamified and social media-like elements:

PC: “Even I don’t remember any of my students having missed a discussion when it was on Canvas. (...) but in Packback, I think maybe first, navigation issues, understanding what exactly they’re supposed to do.”

PC: “Canvas is just doing a better job, it is very easy for us also take a look. It would hardly take one hour for us to completely finish the entire discussion because we know that everyone is going to take part in the discussion only at the deadline. (...) In Packback, you’re not getting that. So it’s too much of a back and forth happening over there”

PB: “I think I can’t see the full potential of Packback and all of its functions, and if I did, maybe I would have a better idea of what really did or didn’t work.”

Preliminary Reflections and Next Steps

The feelings of uncertainty but also curiosity that TAs have been experiencing using Packback could be a result of participating in a Community of Inquiry (Garrison & Anderson, 2003) where a novel element, previously unknown to the TAs, is introduced. The AI is designed to partly replace some 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). In this novel COI dynamic, TAs were faced with the challenge of learning a new technology, understanding how the AI worked, and figuring out how to best complement it with human intervention, with minimal disruption to student learning.

These preliminary findings come from a small sample of graders, three TAs facilitating one large, first-year undergraduate online course in Biology. Further research is ongoing and the research team is interviewing more graders and educators, including full-time professors and educators teaching other subjects, such as political science and learning technologies, to increase the validity of these trends or identify other emerging patterns across subject disciplines and academic ranks. Focus groups are also being conducted with students, to investigate the impact of using AI in online discussions from the learners’ perspective. We are also collaborating with other academic partners to identify cross-institutional trends in the next iteration of the analysis.

References

- 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>
- Baker, T., & Smith, L. (2019). *Educ-AI-tion Rebooted? Exploring the future of artificial intelligence in schools and colleges* (pp. 1–56). Nesta Foundation.
- Barshay, J. (2020). Reframing ed tech to save teachers time and reduce workloads. *Digital Education | The Hechinger Report*. Retrieved from: <https://hechingerreport.org/ai-in-education-reframing-ed-tech-to-save-teachers-time-and-reduce-workloads/>
- Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Buetow, S. (2010). Thematic Analysis and its Reconceptualization as Saliency Analysis. *Journal of Health Services Research and Policy*, 15(2), 123-5. The Royal Society of Medicine Press Ltd.
- Consortium for Policy Research in Education. (2019). *Final Impact results from the i3 implementation of Teach to one: Math*. Retrieved from <https://hechingerreport.org/wp-content/uploads/2019/02/Final-Impact-Results-i3-TtO.pdf>
- Felix, U. (2005). E-Learning pedagogy in the third millennium: the need for combining social and cognitive. *ReCALL*, 17(1), 85–100.
- Garrison, R., & Anderson, T. (2003). *E-learning in the 21st Century: A Framework for Research and Practice*. London: Taylor & Francis Group.
- Lavrakas, P. J. (2008). *Encyclopedia of survey research methods*. Sage Publications.
- McSweeney, E. (2020). How AI eases teachers' heavy workloads. *The Financial Times*, 10 March 2020 issue.
- 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>
- Popenici, S. A., & 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), 22.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: a modern approach*. Upper Saddle River, NJ: Prentice Hall.

Whitelock, D. (2010). Activating Assessment for Learning: are we on the way with Web 2.0? In M.J.W. Lee, & C. McLoughlin (eds.), *Web 2.0-Based-E-Learning: Applying Social Informatics for Tertiary Teaching* (pp. 319–42.). IGI Global.

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.

Acknowledgements

The researchers are grateful to Dr. Ruthanne “Rudi” Thompson at the University of North Texas for allowing this research project to take place in her course.