
INTEGRATING AN AI-DRIVEN DISCUSSION PLATFORM: THE IMPACT OF PLATFORM ON ENGAGEMENT AND QUALITY

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

Abstract

Online educators rely on asynchronous discussions to satisfy the bulk of student-student engagement that is lost when transferring from a face-to-face to an online format. However, not all discussion platforms are created equal and some specialized tools may offer advantages over standard tools embedded in more comprehensive learning management systems (LMS). In this study, we compare two online discussion platforms – one is the native discussion tool embedded within the Canvas LMS and the other is a specialized discussion platform, Packback, which supports students and instructors with gamified elements and artificial intelligence. Specifically, we assess whether post quality differs across these platforms, as measured by average word count, cited sources and weekly participation rate. The initial results are mixed and differ by course, leading us to conclude that the impact of platform on engagement and quality likely depends on the course content, instructor and protocol.

Introduction: Artificial Intelligence in Asynchronous Discussions

The decision to adopt a new tool should be based on, at a minimum, two elements: the impact this tool will have on students and the ease of adoption by instructors. In this article, we discuss our preliminary findings after piloting a discussion platform (Packback) with artificial intelligence (AI) features designed to support instructors in monitoring and grading discussions while encouraging student engagement through quick feedback and gamified elements. Specifically, Packback prevents students from submitting posts that do not meet minimum requirements, such as word count and non-plagiarized content, and assigns each post a “Curiosity Score” using a proprietary algorithm that includes weighted measures for three different quality factors: depth, credibility and presentation, as determined by word count, sentence structure, citations and formatting. Furthermore, posts that appear to contain inflammatory material are automatically flagged and subject to human review to determine whether the content should be removed. Posts of sufficient quality are counted toward a minimum determined by the instructor, enabling them to

generate custom participation reports which can be input manually into their institutions' LMS.

While no formal consensus on what constitutes artificial intelligence has been reached (Pedro, Subosa, Rivas, & Valverde, 2019), there are some common themes in research on AI in higher education. Perhaps most prominently is the fact that the use of AI-driven tools, such as Packback, is commonplace at institutions of higher learning, but users often do not recognize these products as "AI" (Baker & Smith, 2019). Given that such resources are already in use by instructors, formal research into how these tools are related to successful enhancements of course goals are imperative. Furthermore, the use of AI fundamentally involves the pursuit of tangible rewards balanced against the risks of assigning, for lack of a better word, computers to complete tasks that were once accomplished by a human (ITU, 2018). In case of the present study, one sought reward for instructors is enhanced discussion quality and engagement by automating (i.e. trusting to an AI) tasks that were once the responsibility of the instructor, allowing for a greater portion of instructor time to be devoted to providing quality feedback.

Measuring Engagement in Asynchronous Discussions

Students engage in online discussions in different ways, some of which are seemingly "invisible", like reading and browsing. Researchers have spent two decades classifying these behaviours, with the earliest attempt being Hewitt's (2003) analysis of how users in an online discussion environment decide which posts to read. Hewitt (2003) concluded that users were biased toward creating elongated thread structures by responding to the most recent post in a thread. In other words, after the response to the initial post in a thread, subsequent posts were made to replies (and then later replies were replies to earlier replies) – or, in Hewitt's (2003; p.36) words, "recent notes are more likely to inspire responses than older notes". Using session logs, Hewitt (2003) determined that the first action most people took in the discussion forum was reading a post (97.6%), most of which were previously unread (82%). Then students would compose a response, 80% of which were to posts written within the last 48 hours.

On the surface, this does not appear to constrain the growth of the discussion; however, if the most recent response (the post on the "outer edge") does not inspire further conversation, then the thread is likely to die off, regardless of how stimulating older posts in the response chain were. Furthermore, there may be unintentional changes in topic (Hewitt, 2001; 2003). When only the most recent post is taken into account, ignoring the context provided by previous posts, subsequent posts may veer off topic, leaving behind topics that may be more relevant to the course. Fundamentally, Hewitt (2003; p.41) argues that discussants in online platforms are prevented from creating a shared direction because

the participants “do not recognize when topic changes occur, fail to notice when the discourse falters, and generally do not monitor how particular lines of inquiry are evolving.” He suggests including a student or instructor moderator to breathe life into faltering discussion threads and a careful examination of “how individuals interact with the user interface”. To investigate how user-interface influences participation patterns, we conducted an experiment to assess post quality across two discussion platforms. There are four ways in which interface changes across Canvas and Packback could influence engagement and quality: prompt (or “protocol”), gamified elements and discussion structure.

The prompt is a significant difference between Packback and other discussion platforms – Packback has a well-defined, inflexible discussion prompt whereas the native discussion tool in Canvas does not lend itself to a specific style of questioning. Students in Packback are expected to pose a question to their peers about course-related material and other students respond to these questions with supporting points or counterpoints. Prompt can have a significant impact on interactions in online discussions (Zydney et al., 2012), though the specific prompt used in Packback has not been assessed. In order to minimize the impact of prompt, we attempted to replicate this style of discussion using the native discussion tool in Canvas.

A second key difference between Packback and Canvas discussions are the gamified elements embedded within the Packback platform. Packback uses the students’ total Curiosity Points to rank each student on a Learner Leaderboard. From the Learner Leaderboard, students can view their own ranking in the class and quickly navigate to Learner Profiles of other students that show a breakdown of this students’ participation and each of their contributions to the discussion throughout the semester. The Learner Leaderboard is supported by the AI features we discussed earlier – posts are automatically assigned a Curiosity Score, which facilitates this ranking.

Finally, the structure of the discussion platform, including presentation of posts, feed and feedback differ extensively across both environments. For example, Packback shows the discussion as a continuous feed throughout the semester while offering students multiple ways to filter and view the discussion occurring within their course, such as viewing new posts, posts with the most responses and high-quality posts. Instead, the native discussion tool in Canvas is fixed, with the most recent posts at the bottom of the feed. It is possible that alternative platform configurations could impact overall engagement and participation in online discussions, as discussed above by Hewitt (2003). Additionally, methods of establishing instructor presence differ across both platforms. In Canvas, direct feedback to students is always private and must be accessed outside of the discussion forum itself through the students’ grade view. In Packback, graders have the option of publicly

praising a student, featuring their posts or providing private coaching. This enables instructors to highlight the elements of posts students should mimic in a highly visible way, facilitating the feeling of teaching presence, a core component of the Community of Inquiry model (COI) (Anderson et al., 2001; Garrison et al., 2001; Zydney et al., 2012) associated with improved cognitive presence (Park et al., 2015) and students' perceived learning and satisfaction (Arbaugh, 2010). According to this model, Packback's practice of publicizing instructor praise in a highly visible way could enable instructor's to more easily act out the role of a "facilitator and co-creator of a social environment conducive to active and successful learning" (Anderson et al., 2001; p.2). The combination of prompt, AI-supported feedback and highly visible public praise provides "participants with the directions to facilitate themselves" ultimately "reducing the burden on the instructor" (Zydney et al., 2012). Interaction on the part of the instructor that is both highly visible but infrequent may allow students space to engage in behaviours typically associated with high teaching presence (Park et al., 2015) while coaching could lead to an increase in the frequency of higher-order thinking (Stein et al., 2013).

In our initial assessment of post quality, we compared platforms in terms of three indicators of engagement and post quality: posts per week, word count and citing sources. We predicted that students using the Packback platform would do the following:

- H1. When using Packback, students will average a higher word count per post, compared to Canvas.
- H2. When using Packback, a higher proportion of student posts will reference external sources, compared to Canvas.
- H3. When using Packback, students will average a higher number of posts per week, compared to Canvas.

Methods

The findings we will present in this presentation are part of a larger, ongoing study at the University of North Texas about the impact of using Packback on instructor workflow and student learning. The data that we will use to convey these preliminary findings are based on the discussion posts from students enrolled in two different courses. The two courses followed a similar discussion protocol, but the student population varied significantly. One course was an upper-level Political Science course, while the second was a graduate course in Learning Technologies. The Political Science course was face-to-face, while the Learning Technologies course had both a face-to-face and online section. All data is from the fall 2019 semester and data collection is ongoing for the spring 2020 semester.

Post Quality and Engagement

For both courses, students used the native discussion platform in Canvas for the first half of the semester and, during the second half of the semester, students used Packback. In all, students in the Political Science course engaged in discussions in Canvas for seven weeks and six weeks in Packback. Students in the Learning Technologies course engaged in discussions for eight weeks in both Packback and Canvas. There were 41 unique discussion participants in the Political Science course and 16 unique participants in the Learning Technologies course. In order to be included in the analysis, participants must have posted at least once in both Packback and Canvas. Altogether, these students authored 1,027 posts in Political Science and 456 posts in Learning Technologies, but these posts are aggregated to the student level for all further analysis.

We compare the effect of both platform and course on post quality and student participation using a Two-Way Mixed ANOVA in IBM SPSS 25, as measured by: (a) the proportion of student posts with cited sources, (b) students' average word count per posts and 3) students' average number of posts per week.

Results

Post Quality and Engagement

The impact of platform and course varied depending on the outcome measure (word count, proportion of sources cited, or average posts per week). Specifically, results suggest that there is a main effect for both platform ($F(1,55) = 23.99, p < .001, \eta^2 = .30$) and course ($F(1,55) = 15.93, p < .001, \eta^2 = .23$) on word count. When using Canvas ($M = 180.16$), on average students wrote longer posts than when using Packback ($M = 148.11$), and students in the Learning Technologies class ($M = 193.06$) wrote longer posts than those in the Political Science course ($M = 135.21$). However, a significant interaction was present between platform and course ($F(1,55) = 13.64, p = .001, \eta^2 = .20$), suggesting straight interpretation of the main effects is not a full representation of test results. Initial inspections of group descriptive suggest that the main effect of platform was driven by the especially large difference across platforms for the Learning Technologies course (Canvas $M = 221.17$, Packback $M = 164.94$), while the differences for the Political Science course were much less pronounced (Canvas $M = 139.15$, Packback $M = 131.27$). However, in both courses, posts were longer when students used Packback, indicating that there is no support for H1: "When using Packback, students will average a higher word count per post, compared to Canvas". In fact, especially for the Learning Technologies course, posts were significantly longer in Canvas. Figure 1 shows the average word count per post for both courses in both conditions.

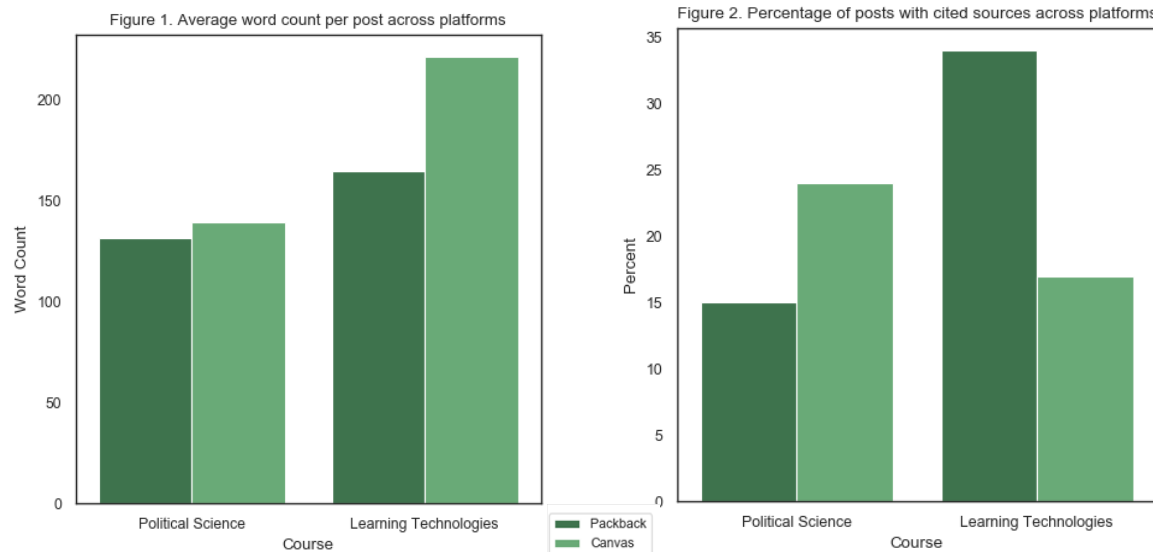


Figure 1. (right) This figure shows the average word count per post across platforms. Students in the Learning Technologies course wrote longer posts overall, but this difference was most striking for posts made in Canvas.

Figure 2. (left) shows the percentage of sources cited across both courses and platforms. The trend is opposite for both courses.

For the next measure of post quality, proportion of sources cited, results suggest that neither a main effect for platform ($F(1,55) = 6.39$, $p = .014$, $\eta^2 = .10$) or course ($F(1,55) = 8.63$, $p = .005$, $\eta^2 = .14$) is present. More specifically, there did not appear to be a difference in how often students using Canvas ($M = .20$) or Packback ($M = .25$) cited sources. Furthermore, overall results suggest students in the Learning Technologies class ($M = .26$) cited sources at similar rates to the Political Science course ($M = .19$).

However, there was significant interaction between platform and course ($F(1,55) = 12.04$, $p = .001$, $\eta^2 = .18$), suggesting straight interpretation of the main effects is not a full representation of test results. Initial inspections of group descriptives suggest that the impact of platform depending on the course students were enrolled in. As shown in Figure 2, Learning Technologies students (Canvas $M = .17$, Packback $M = .34$) cited more sources in Packback, while students in the Political Science course (Canvas $M = .24$, Packback $M = .15$) cited more sources in Canvas. These findings indicate that there is partial support for H2: “When using Packback, a higher proportion of student posts will reference external sources, compared to Canvas”.

Finally, we measured engagement by the students’ average number of posts per week. Results suggest that there was a main effect for both platform ($F(1,55) = 6.39$, $p = .014$, $\eta^2 = .10$) and course ($F(1,55) = 8.63$, $p = .005$, $\eta^2 = .14$). More specifically, when students used Canvas ($M = 1.86$), they appeared to post more often than when using Packback ($M = 1.66$), and students in the Learning Technologies class ($M = 1.60$) posted fewer times

per week than students in the Political Science course ($M = 1.91$). However, a significant interaction was present between platform and course ($F(1,55) = 24.15, p < .001, \eta^2 = .31$), suggesting straight interpretation of the main effects is not a full representation of test results. Initial inspections of group descriptives suggest that inverse effects of platform exist depending on what course the students were in. Learning Technologies students (Canvas $M = 1.91$, Packback $M = 1.30$) posted less often in Packback, while students in the Political Science course (Canvas $M = 1.82$, Packback $M = 2.01$) posted less often when using Canvas. These findings indicate that there is partial support for H3: “When using Packback, students will average a higher number of posts per week, compared to Canvas”.

It is clear that, when comparing these two discussion platforms, the question of engagement and quality is complex. In the following section, we will discuss some possible explanations for these findings and next steps for this research.

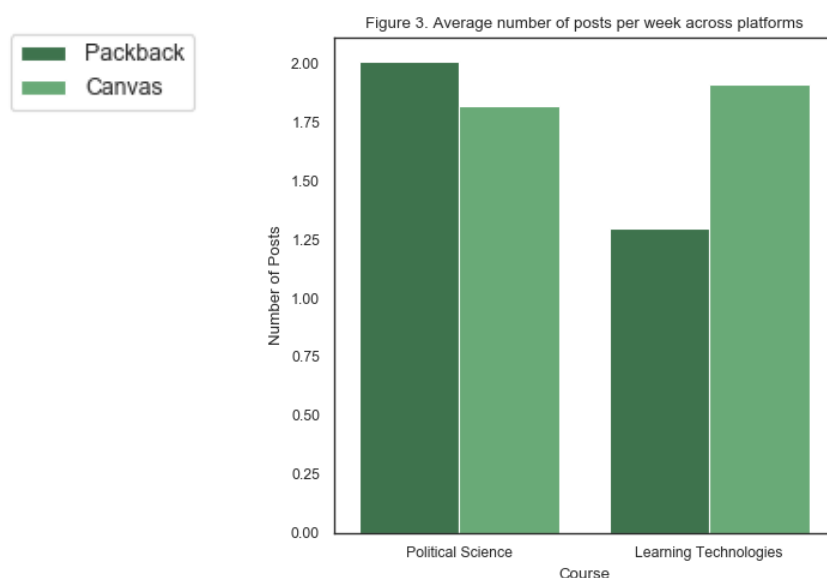


Figure 3. This figure shows the average number of posts per week for both courses across platforms.

Discussion

It is difficult to say conclusively whether the impact of using Packback was positive or negative – in fact, it is difficult to say whether it was positive or negative even for a specific course. In the Political Science course, students wrote posts of approximately the same length, but cited sources about 1.6 times as often in Canvas. In other words, their posts were of comparable length, but, when posting in Canvas, students more often provided evidence for their claims. Even so, the participation rate was higher in Packback than in Canvas for these students – indicating that they posted more often. In fact, given that the instructor required a minimum number of two posts per week, it appears that students met this minimum more often in Packback than in Canvas. This could be due to the fact that participation guidelines become more clear by the end of the semester, but it also could be

related to Packback's ease of use or the gamified elements incentivizing participation. We are currently interviewing students about their usage of these platforms in order to ascertain which is more likely. We will also plot participation rates over time to verify whether participation rates in Canvas are lower at the beginning of the semester, suggesting a period of adjustment that artificially lowers the overall participation rate. Interestingly, the fact that participation in Packback (at the end of the semester) is both higher and in line with course requirements is notable in of itself, as previous research suggests that participation in asynchronous discussions declines during times of increased workload (such as final exams) (Ding et al., 2018; Xie & Durrington, 2011; Xie & Fengfeng, 2011).

The trends in the Learning Technologies course suggest an entirely different story, however. Overall, posts were shorter, but they provided supporting evidence at nearly twice the rate in Packback compared to Canvas, and their participation rate was also lower in Packback – there were fewer, shorter posts, with more supporting evidence. There are two potential confounding factors worth mentioning in this case. First, the instructor of the Political Science course was a veteran user of Packback; her prompt in Canvas mirrored that in Packback as closely as possible and she assisted her students in the transition from one platform to the other more easily, given her experience using the tool. Meanwhile, the instructor of the “Learning Technologies” course was not familiar with the tool, leading to a more difficult transition and a prompt that was not aligned. The Canvas prompt was more instructor-driven, requesting that students respond to instructor-derived questions, while the Packback prompt requested that students pose questions to one another and respond to the student-derived questions. Responding to instructor-derived questions is likely to result in longer, reflective posts, rather than sustained dialogue or debate between students, which would be characterized by shorter posts with more corroborating evidence (cited sources). Generally, we would expect higher participation with this style of dialogue, as well, but the undoubtedly difficult transition between Canvas and Packback for this course could artificially lower participation rates. Because of this, we will track rates over time to assess when during the semester participation was lowest.

There are many unanswered questions here – is it the tool or the prompt that is driving differences in student behaviour? Are instructors utilizing the AI features of this tool in order to bolster instructor presence? Do students even notice the gamified elements designed to extrinsically motivate them to participate? Does transitioning from one platform to another cause participation rates to drop? In order to address these questions, we are conducting focus groups with both instructors and students. We are also carefully reviewing participation over time and have designed an experiment to assess the impact of transitioning on student participation rates. Ultimately, data from different disciplines, use

cases from instructors and student perspectives will be invaluable in determining whether this gamified, AI-driven tool positively impacts the asynchronous discussion experience.

References

- Anderson, T., Liam, R., Garrison, D. R., & Archer, W. (2001). *Assessing Teaching Presence in a Computer Conferencing Context*. Retrieved from <https://auspace.athabasca.ca/handle/2149/725>
- 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*. Nesta Foundation.
- 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>
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7–23. <https://doi.org/10.1080/08923640109527071>
- Hewitt, J. (2001). Beyond Threaded Discourse. *International Journal of Educational Telecommunications*, 7(3), 207–221.
- Hewitt, J. (2003). *How Habitual Online Practices Affect the Development of Asynchronous Discussion Threads*.
- ITU. (2018). *Artificial Intelligence (AI) for Development Series Module on AI Ethics and Society*. Switzerland.
- 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>
- Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: challenges and opportunities for sustainable development* (7th ed., Vol. 7, pp. 1–46). France: UNESCO.
- 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>

Xie, K., & Durrington, V. (2011). Relationship between Students' Motivation and their Participation in Asynchronous Online Discussions. *Journal of Online Teaching and Learning*, 7(1), 17–29.

Xie, K., & Fengfeng. (2011). The role of students' motivation in peer-moderated asynchronous online discussions. *British Journal of Educational Technology*, 42(6), 916–930.

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>

Acknowledgements

We would like to thank the instructors who allowed this research project to take place in their courses, Dr. Jacqueline DeMeritt (Political Science) and Dr. Lin Lin (Learning Technologies), both at the University of North Texas.