INFORMATION SYSTEMS EDUCATION JOURNAL

Volume 20, No. 4
September 2022
ISSN: 1545-679X

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is six times per year. The first year of publication was 2003.

ISEDJ is published online (https://isedj.org). Our sister publication, the Proceedings of EDSIGCON (https://proc.iscap.info) features all papers, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the EDSIGCON conference. At that point papers are divided into award papers (top 15%), other journal papers (top 25%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 40%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org. Special thanks to members of ISCAP/EDSIG who perform the editorial and review processes for ISEDJ.

INFORMATION SYSTEMS EDUCATION JOURNAL

Editors

Use of Artificial Intelligence to Grade Student Discussion Boards: An Exploratory Study

Stephen M. Rutner
stephen.rutner@gcsu.edu
Department of Management, Marketing and Logistics
Georgia College and State University
Milledgeville, Georgia 31061 USA

Rebecca A. Scott
scottra@uncw.edu
Congdon School of Supply Chain, Business Analytics & Information Systems
UNC Wilmington
Wilmington, North Carolina 28403 USA

Abstract

There appears to be an increasing acceptance of Artificial Intelligence (AI) across society. As people become more comfortable with AI's use in advertising, basic services and other areas of day-to-day life, the question arises will students also be willing to accept AI in learning situations. Furthermore, what are the impacts on both the student learning and acceptance as well as the effect on the instructor or professor. This paper presents the initial findings of the use of AI in grading students' discussion boards. It presents an initial model of student expectations, discusses potential benefits and drawbacks of AI and presents initial findings from a limited number of classes using AI grading.

Keywords: Artificial intelligence, Discussion boards, Pedagogy, Asynchronous learning, Online learning

1. INTRODUCTION

During the first year of COVID-19 pandemic, many traditional pedagogical tools and methods were stressed as classes were often shifted from face-to-face (F2F) to asynchronous, online (Kafka, 2020). During the early phases of the pandemic in 2020, many students went home for spring break only to not return to the physical classroom until fall semester 2021. This required rapid redesign of learning methods to continue courses and not disrupt students' paths toward graduation (Sanders, 2020).

These rapid changes often forced faculty to incorporate new learning methods to meet the asynchronous nature of these classes. For example, the traditional classroom discussion was not possible unless an online audio-visual conferencing platform was used (e.g., Zoom, MS Teams, WebEx, etc.). If an online platform was either unavailable or not used, the discussion portion of the class would suffer without an alternative. Most faculty are aware and have often used traditional learning management systems (LMS) (e.g., Blackboard, Desire2Learn, etc.) discussion boards as a means to an end in online classes. Furthermore, even by 2010, approximately 85% of universities were using some form of LMS (Chen et al, 2010). Therefore, it was a natural alternative to classroom discussion while adopting to the COVID environment. However, the likely stresses of moving multiple classes from F2F to asynchronous meant that faculty's time was pressed. Many faculty members were overwhelmed early in spring 2020 semester trying to convert content, include all learning activities, operate in a new environment, and maintain academic standards. These challenges highlighted the opportunities for companies to both reduce manual grading and increase student learning through various new or modified teaching tools.

The use of discussion boards represents one opportunity to improve from traditional uses to an enhanced version. In spite of large amounts of literature supporting the benefits of discussion boards, many faculty members are reluctant to use discussion boards for a variety of reasons. First, they are often concerned that the conversation will not be as "rich or inactive" as F2F or in-class conversations (Smidt et al, 2014). Another issue might be that discussion boards are often not voluntary (i.e., a required number of posts) which will impact the learning (Frey and Wojnar, 2004; Gill 2006). Finally, there is a concern on the difficulty of balancing the interaction between the faculty member and students to enhance learning without dominating the discussion (Dennen, 2005). Each of these valid concerns are in addition to the increased amount of faculty time to read all discussion posts and accurately assess them.

Given these challenges and facing the COVID environment, many faculty members were forced to adopt discussion boards into their classes without significant planning, testing or time beginning in the spring of 2020 to substitute for F2F discussions. Even though there were significant benefits to discussion boards, the challenges forced instructors to seek better processes for their benefit and outcomes to ensure improved student learning.

The purpose of this article is to highlight a specific pedagogical tool that appears to improve learning while simultaneously reducing faculty workload by using AI to help in evaluating student responses on a discussion board. Following this section, the literature review will highlight both use of discussion boards and the specific use of AI in grading students. This will help to develop a theoretical model and research propositions for further testing. Next, an early set of student responses will be presented. Finally, the conclusions and impact of this initial study will help develop the future examination of this subject.

2. LITERATURE REVIEW

The Literature Review is divided into two broad subsections. The first is to review the well-established research of the value of discussion boards in academia and highlight one of the key challenges of evaluating student responses. The second subsection is to identify the less developed, but growing, body of works on the application in AI in academia with a focus on the few recent articles involving discussion boards. The goal of these subsections is to identify the gaps in the literature that require further examination.

Discussion Boards and Evaluation

As mentioned briefly in the Introduction, there is a significant amount of literature about the benefits and disadvantages of using discussion boards in various academic settings. It would be beyond any paper to cover all of that research. Therefore, a brief synopsis of those is included. A detailed review of the more relevant literature revolves around the subject of discussion board evaluation and/or grading.

Since this article previously identified some of the challenges of discussion boards, it was reasonable to also present some of the benefits of using them in various educational situations. Hinton and Bradshaw (2004) did some initial examination of the perceptions of Autonomous Online Discussion (AOD). They found that it was difficult to evaluate the effectiveness. However, they did identify AOD as a "Core element" of online learning and course design. Furthermore, Hew et al (2008) further confirmed that AOD was becoming an "Increasingly common means to facilitate dialogue between instructors and students." They also provided an in-depth history of the overall online literature with a specific focus of challenges and studies applying each potential solution which is discussed later.

The benefits of AOD are numerous and have been thoroughly examined over the last twenty years. First, the unique nature of AOD allows students some flexibility on the timing of posts and time to reflect before replying (Murphy and Coleman, 2004). Another benefit identified by researchers is the actual act of writing, as opposed to verbal response, often helps students to increase learning (Newman et al, 1997; Vonderwell, 2003). Tracy et al (2020) also identified that when performed properly, AOD can increase student engagement and improve learning. Finally, one recent study compared the use of AOD with Zoom and found that students using AODs had increased performance in the class. This implies that properly applied AOD may actually work better than traditional discussion format in the classroom or in a online, real time learning environment (Ackerman and Gross, 2021). All of these studies highlighted some of the key benefits to using AOD as part of an online learning experience. Furthermore, the purpose of this subsection was not to state the shift to online courses due to COVID was a better overall learning experience, but rather, to identify the positive aspects of AOD. Each of the articles highlight a positive aspect that can be used regardless of F2F or asynchronous learning courses.

Unfortunately, there are a number of negative aspects to using AOD. Using Hew et al's (2008) synthesis of the overall literature, they identified three specific areas or dilemmas that faculty face using AOD: use of grades, number of posting guidelines, and instructor-facilitation. While all three of these areas are of interest to most modern educators, the first is key to this research (Hew et al, 2008). While there are a host of other issues, the key element of student evaluation remains a challenge even post COVID. For example, Dennen (2005) found that if there are not clear expectations given by the faculty member, students' interests and efforts will wane. In other words, the students are not willing to put forth efforts if it did not result in better individual grades. Furthermore, Dennen (2005) found the students benefited when post guidelines were specified (i.e., format, style, length, etc.). Also, faculty grading was a key component to student participation in AOD. The greater the weight of the grade, the more involvement by the student (Cifuentes et al, 1997). Finally, Murphy and Coleman (2004) also found that when students were required to post, the responses often devolved to "Me too" or "I agree" types of general comments. The net effect was that AOD grading created benefits and challenges to the overall learning.

However, the Murphy and Coleman (2004) articles raised a significant point that applies to the faculty member. The increased number of posts requires that every comment must be read, reviewed, contemplated and assigned a grade of some sort. This amount of time to incorporate a systematic process to fairly assigned grades to an AOD can be significant. Furthermore, it can feel somewhat arbitrary to the students. Therefore, one finding to many faculty members that have not used AOD prior to COVID may have been the significant increase in time to move from a F2F discussion evaluation of student comments to an AOD evaluation of much larger amounts of material. A fair amount of literature has been developed about the grading of discussion boards. Pecka et al (2014) states that "Rubrics are often used to facilitate and evaluate student's discussion board postings." In addition to the use of rubrics, they found that the use of AOD help to increase higher order learning in general for the students. Finally, one of their key findings was the inclusion of rubrics further increased the level of higher order learning with AOD. Phillippi et al (2015) also applied national and international competencies within their field to grade discussion. From those competencies, they developed a rubric to apply to each discussion post. The result was clearer guidelines for students and faculty to follow improving the use of AOD. Finally, Hew et al (2008) also stated that the use of rubrics for specific categories of contribution could help students' efforts. The overall result is that there are numerous studies and examples of how to standardize grading through the use of rubrics and the potential benefit for both faculty members and students.

Artificial Intelligence or Auto Grading

While the literature addresses the rubric process, the main benefit is to normalize the grades for the students, but it does not significantly reduce the workload on the faculty member. The challenges of grading an open-ended student work can be time consuming (Tsai, 2012). Furthermore, some faculty are likely to avoid giving open-ended assignments due to the time required to grade them (Tsai, 2012). A possible solution to this is the use of automatic or AI grading. But, some faculty were also reluctant to use any form of automation due to their belief that computers were not sophisticated enough to replace human judgement in grading (Bridgeman and Quinlan, 2009). Yeh et al (2007) also found that automated grading systems did not do an adequate job of dealing with higher level and/or critical thinking. This is an interesting finding and may be due to the level of computer sophistication or the lack of common use of AI in society in 2009. However, the initial literature search for AI or automated grading even in 2021 created an interesting result. The top 100 papers gathered by the library search engine, Galileo, had less than ten papers that involved academic applications of grading. Rather, the medical use of AI of grading various symptoms, diagnosis or treatments accounted for over 75% of the results. The implication that widespread use of AI may be much more advanced in the medical community versus academia. Furthermore, the majority of the academic literature trends toward specific computer tools, languages, engineering approaches and applications to improve the process rather than the impact on students and faculty members.

Regardless of the amount of AI usage in academic literature, automatic grading offers a number of potential benefits to both faculty members and students. Tsai et al (2012) did find that while not perfect, AI grading did offer the following potential benefits: consistency between students, rapid grading, never gets tired, and provides immediate feedback. To address some of the specific shortcomings of AI grading, Kyrilov and Noelle (2014) identified a theoretical framework to improve AI grading using the case-based reasoning (CBR) approach. Figure 1 – CBR Methodology presents the learning process for computer grading. The goal of their process was to develop the AI's ability to improve its grading. Finally, they stated that CBR was not widely adopted within the educational community, but CBR had the ability to assist instructors with grading of open-ended student works. It should also be noted that they foresaw the use of CBR in the medical community nearly ten years ago.

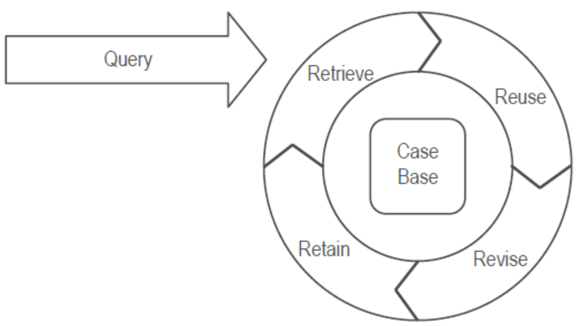

Figure 1 – CBR Learning Framework

Not surprisingly, in the nearly ten years since Kyrlov and Noelle's work, advances have been made in grading open-ended responses by AI. Liu et al (2021) identified the tedious nature of grading these types of answers and applied an automated grading method using multiway attention networks. Their experiments demonstrated superior results compared to six other grading methods. The overall results highlight the ever-increasing power and accuracy of the AI grading systems available to faculty members.

Delgado et al (2020) further identified the advantages of modern AI embedded within a LMS (Pearson MyEnglishLab) to provide specific and tailored feedback to students. In their paper, they demonstrated how the AI's comments were specific and designed to help students identify and improve weak areas of their answers.

As the use of AI grading progresses, current studies are exploring the use beyond simple responses in AOD. Rather, can a different form of input into the AOD be analyzed by the AI. Ghoneim and Elghotmy (2020) studied the use of AI boards for input into the grading system. While their study differed from traditional use of AOD, it did highlight the potential for creative uses for AI. Furthermore, they were one the few studies that specifically stated that the use of AI could be "Fun" for the student if creatively applied.

It is clear that the literature presents a solid overview of the challenges and benefits of the use of AOD. Additionally, there appears to be a growing use of AI in various aspects of the educational community. The increasing sophistication of AI grading has helped to alleviate some of the drudgery and inconsistency of AOD. However, most of the literature was focused on the pros/cons, methods, technical aspects, applications and outcomes of using AI. Very little focused on the reaction from students as well as their learning.

3. THEORETICAL MODEL AND RESEARCH PROPOSITIONS

Based on the previous research of the concept of AI grading, there are numerous potential impacts on student discussion quality, quantity, and learning. The traditional interaction between faculty members and students in an AOD are limited by the asynchronous nature of the process. Figure 2 represents a typical student and instructor interaction process.

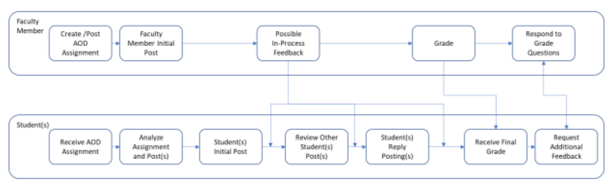

Figure 2 – Traditional AOD Interaction between Faculty and Student(s)
Note: a full size figure in the appendix

The figure highlights the typical pedagogical process on the part of the faculty member. Once the instructor chooses to incorporate an AOD, he or she creates some sort of assignment followed by an initial post containing instructions or questions to beginning the discussion. The instructor then would typically read some posts and may provide feedback at various times through the process. Finally, he or she would grade the students' individual posts and assign a grade. This would be followed by the likely questions from various students concerning grading. Most of the process is linear and involves limited interaction with the student. A key constraint is the faculty member's time to provide timely feedback to the students. Also, the students' post must be published to the board before the faculty member can provide feedback. These limitations force the student to either accept their initial posts without change or to create more posts that need to be evaluated yet again by the faculty member. This creates even more work and further limits faculty time to evaluate posts.

From the student side, the figure demonstrates the process from their view. It is also linear from receiving the assignment, to making initial post(s), reviewing other students' submissions, possibly receiving feedback, and then, making a final post(s). This is also followed by receiving their grade for the assignment which may trigger a question to faculty member about that grade. A key point is the limited interaction between student and instructor. There may be feedback, but it always lags from the initial post. Often, it may be days until the professor is able to catch up to the numerous posts in the discussion board. Therefore, a student is often left with little to no feedback during the traditional process.

Based upon Kyrilov and Noelle's framework (2014, Figure 3), rapidly received feedback could improve the students' posts, the level of discussion and overall quality of the AOD. Figure 3 presents an adapted version of their model to integrate into the traditional AOD interaction model (Figure 2).

The adapted process assumes that immediate feedback is available to the student through the use of AI grading. The student prepares an initial draft of his or her post. The AI grading would provide either instantaneous or immediate feedback during the draft process. The student then likely revises and improves the post a number of times until he or she is ready to submit it as their submitted post. The net result is likely a vastly improved overall product that has encouraged and motivated the student to think more deeply about the subject and increase overall learning. This occurs with all students' posts nearly simultaneously with little to no faculty interaction.

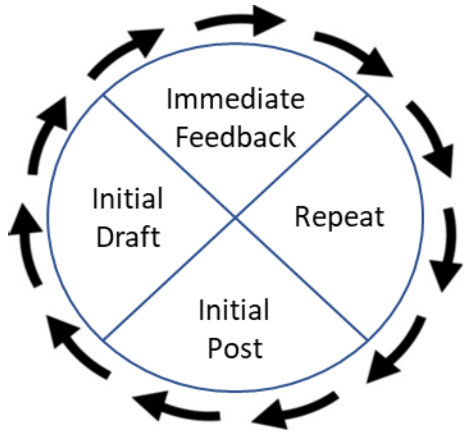

Figure 3 – Adapted Immediate Feedback Model

By integrating the adapted immediate feedback model into the traditional AOD interaction model, an improved AI grading model is displayed in Figure 4 – Incorporating AI Grading into AOD. This model presents the changes in the interaction between the instructor and students by including the AI feedback into the process. First, it demonstrates the timelier feedback from the AI grading. Furthermore, the adapted immediate feedback model interacts with both the faculty member and students' tracks. By providing immediate feedback, the AI acts as a surrogate for the faculty member. It also relieves some of the pressure on the faculty member to try and provide timely feedback. The AI grading becomes a linking feature between the students and the faculty member.

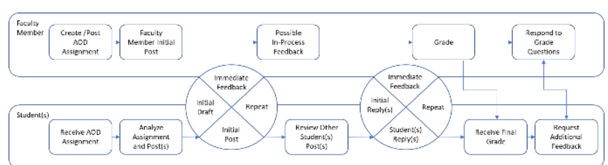

Figure 4 – AI Feedback Modified AOD Process Model
Note: a full size figure in the appendix

Based upon the adapted model using AI feedback or grading, there are a number of research questions that are designed to fill the gaps in the literature. Each of the propositions identifies key issues beyond the software mechanics of AI grading, but rather focuses more on the potential impacts and benefits for both the students and faculty members.

P1: Students benefit from immediate or real-time AI generated feedback.
P2: AI grading and feedback is adequate to replace faculty member inputs during the discussion board posting cycle.
P3: AI grading and feedback encourage students to think more deeply about the topic.
P4: Students will prefer the AI grade to the instructor's grading process.

These first four propositions focus on the potential pedagogical benefits of using AI grading and/or feedback. The assumption is that student learning benefits from any type of immediate feedback. The challenge is that in a real-world setting it is unlikely that faculty members are able to provide real-time or near instantaneous feedback. Furthermore, with the ever-increasing AI sophistication, the current state of AI feedback and grading is adequate to replace instructor comments at least during the discussion board process. However, this is not to imply that AI grading is fully able to provide final grades at this point. Finally, near simultaneous feedback encourages the students to review, revise and resubmit their initial and follow on posts which should encourage deeper thoughts on the topic and an increased learning level for the material.

P5: Artificial constraints in the AI system reduce the students' perceived benefits of using AI graded AOD (e.g. word limits, requirements to post a question vs statement to begin, lack of discussion board structure, etc.)
P6: Immediate feedback will reduce stress on the students throughout the posting process.
P7: An outside vendor (i.e., not university integrated LMS) will create issues for the students – cost, technical issues, ease of use.
P8: An outside vendor's desire to attract customers will create hidden benefits to the student.

The second group of propositions focus more on the mechanics of an AI grading/feedback system. The AI system is likely to have some limitations due to the programming. These may include, but are not limited to, word counts, required formatting or use of questions, various discussion board structure, etc. Furthermore, the large LMS that universities are using do not incorporate AI at this point. This necessitates additional steps, time, effort, and cost to the students and faculty member to employ the AI grading system. Therefore, both faculty member and students have to weigh the tradeoffs of using the system. Also, since a third party vendor is providing the AI solution, there is an implied belief that the company will constantly work to improve the product due to competition in the marketplace which may reduce disadvantages to the students and faculty members that exist at the time of this study. The net result of the second group of propositions is that the improved AI product should benefit the students and faculty member to include the pedagogical propositions (P1-P4).

4. METHODOLOGY

To conduct an initial examination, an AI system was chosen and applied with a student sample. Georgia College and State University used Packback across three traditional asynchronous, online, graduate classes during the spring 2021 semester. The classes were all part of a single Master program. Two different faculty members were the instructors of record. Also, the three classes were three different courses across two differing cohorts of students. All three classes had been taught before using traditional discussion boards; so, the switch to an AI grading/feedback board was a minimal change to each of the existing courses. In other words, the test classes were not part of the reaction to COVID nor involved other significant pedagogical changes. Finally, all of the students were in their second or fifth semester of the five semester program and had used a traditional discussion board as a part of the integrated LMS in a previous class(es). The faculty members believed this group of students would provide a fairly wide cross section of views and experiences. Also, with the students' experiences with traditional discussion boards, they would be excellent judges of the benefits and disadvantages of use the Packback AI system throughout the semester. Finally, since this was an exploratory study, a simple 29 question survey was offered to the students for a small amount of extra credit at the very end of the semester. The majority of questions were five point Likert scale responses about the Packback system. The responses were anonymous, but the students' identification numbers were collected in a separate file to apply credit for completing the survey.

5. PACKBACK

Packback is an online discussion board platform. It was chosen based upon an initial recommendation from faculty that were using it with undergraduate students at another university. On the Packback website home page, they state that use of their AOD product will

"Inspire self-motivated, critical thinkers through inquiry-driven discussion." They even provide comments that their system will improve the learning and grading outcomes for students, create a more rigorous discussion and reduce the workload on faculty members (Packback, 2021).

A goal of the Packback system is to improve both students' discussion and easy faculty workloads. These are two of the critical issues identified in previous studies as advantages. However, the question arises of how does Packback work and how effective is it AI grading system.

One key difference between a traditional discussion board and Packback's system is the use of an AI grading process. The first of the two major parts of the AI grading in Packback was when the students are drafting their post. Packback provides a number of helpful items to encourage them to be more complete with their answers. Figure 5 – Student New Post Screen provides an example of what a student would see while drafting a post. The Instant Feedback column on the right side of the students' screen helps to guide the students' responses. A key item is the student is assigned a "Curiosity Score" during this process. While it is in draft mode, the score is displayed as a range. For example, the example post below is low with a 31-70 potential score. In two of the classes, an 80 was required to have the post count as a valid post. Also, the system helps the students to not focus only on the curiosity score, but encourages them to fix grammatical errors, add links to relevant material, include videos/pictures/charts and checks for plagiarism both inside the Packback program and outside. Finally, as soon as the students finishes the post, he or she will receive their curiosity score.

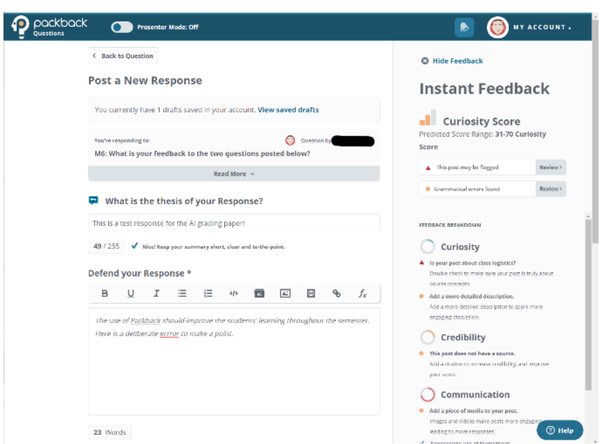

Figure 5 – Student Post Screen
Note: a full size figure in the appendix

This score is provided entirely by Packback and does not involve the faculty member at this point. The score is derived by a Packback algorithm based on a combination of the students' presentation, creditability and effort/depth of the individual post. Without having an entire discussion of the AI process, the score can be summarized as applying an algorithm that correlates high activity, highly curiosity of highly driven member posts. The scores are valuated against other students' posts not only within their class's discussion board, but compared to all other students using Packback. Finally, the algorithm checks for credibility of the post based on relevant and reliable sources that are used to defend the students' main points. This process helps to address some the common concerns about AI grading reliability.

To continue with the example, one of the faculty members in the test classes required a minimum curiosity score of 80 for the post to count as one of the three required postings for the weekly discussion. The score itself was not used as the sole grade for the students' discussion board results throughout these test classes. However, due to the nature of graduate students, the faculty members observed some "friendly competition" among the students to continuously improve their discussion posts' curiosity scores. Figure 6 – Student Post on Discussion Board shows what students viewed after posting their work.

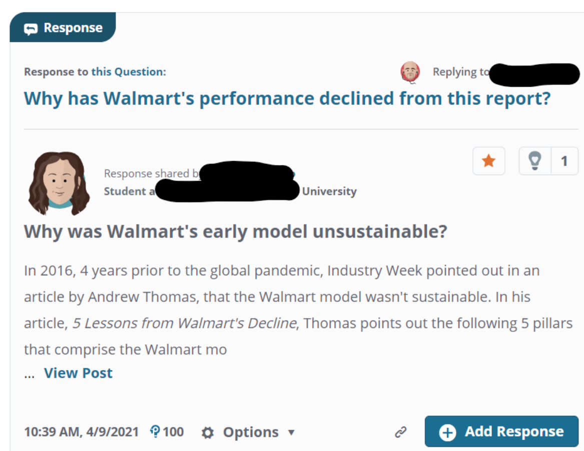

Figure 6 – Student Post on Discussion Board
Note: a full size figure in the appendix

This example was taken from an actual reply from one of the classes. It was chosen as an example for a number of reasons. First, since it was fairly long (four full paragraphs), Packback abbreviated it and had a "View Post" to see the complete post. This allowed the shorter version to be screenshot more easily and demonstrate a number of key points in one figure. First, you can see the student was replying to another student's post. Also, on the bottom row, the student's final curiosity score was a 100. A key point about the system was the ability for students to edit and re-edit their posts. Assume that the example draft post example ended up scoring a 70. The student could then go back, re-edit it and repost immediately. The new score may be an 85 or 90. If satisfied, the student could leave it as his or her post, or if unsatisfied, he/she could re-edit again in an attempt to increase the score. The resulting iterative cycle created many very highly scored posts. The real benefit was not the high curiosity scores, but rather, students reviewing and revising their work to create better posts which helped in the learning process.

Finally, looking at the top right corner there is a star and a lightbulb. A star was if the faculty member featured this as a significant post in the discussion, and the lightbulb, or "sparks," represent posts that the faculty member or other students "sparked" their curiosity. These little items added a different type of feedback and provided a useful tool to ensure especially good posts were read by the entire class.

There were some significant drawbacks to using Packback. First, there is an additional cost to the students to purchase use. The pricing model continues to change, but it was approximately $20 per class during the test semester. Also, Packback is not fully integrated into the various LMS. Therefore, faculty have to transfer grades between the systems. Packback did provide a very good tool to download scores into Excel spreadsheets with numerous options. A unique challenge with Packback was the inability to subdivide the course discussions into modules. The entire semester had to be performed on the same discussion board (there were a number of tricks to minimize this: post naming conventions, feature postings, etc.). Finally, Packback was another system that students and faculty members had to learn and operate beyond the university's LMS.

The overall result was that Packback is not a revolutionary new system. However, it is clearly an evolutionary step in applying AI to the grading and feedback portions of discussion boards. The faculty were encouraged enough by the anecdotal successes during the semester to use it again in the fall 2021 semester with the same program's students.

6. FINDINGS

The initial survey resulted in 72 useable responses. Table 1 presents the demographic results for gender, class, etc. It should be noted that 100% of the students were in the Master of Logistics and Supply Chain Management program in this study. The demographics are fairly representative of a group of graduate students in the field. It leans a little towards the male side of respondents. Two of the three classes were more represented, but that also aligns some to the class sizes. Since the students are graduate business majors, it is also reasonable that the PCs were much more common that Macs in the sample. Next, the grade distribution is reasonable given both the graduate level and split between first- and second-year students and the likelihood of higher achieving students being a little more likely to provide feedback. Finally, the response by 72 students out of a total population of 95 resulted in a 75.8% response rate. It should be noted that a small number of students could have been in two courses simultaneously but were limited to responding in only one class.

Variable	Responses		
Gender	Male = 59.51%	Female = 40.85%	
Year Group	1st Year = 46.48%	2nd Year = 49.30%	Other = 4.23%
Course	Inventory = 37.50%	Strategy = 47.22%	Purchasing = 15.28%
Computer	PC = 85.92%	Mac = 14.08%	
GPA	4.00 = 49.30%	3.50-3.99 = 22.54%	3.00-3.49 = 19.72%
GPA (cont.)	Below 3.00 = 2.82%	Not Provided = 7.04%	

Table 1 – Summary Demographics

The use of Likert scale survey questions attempted to evaluate the effectiveness of the Packback AI towards the students. One of the key differences between this work and previous studies was to collect student feedback about the use of AI grading and/or feedback. A series of specific questions asked questions based upon the research propositions. For example, questions regarding the value of the immediate scoring and other areas addressing the pedological impact were included in the survey. These questions were aimed at the first four propositions. Also, there were numerous questions about the specific process to include strengths and weakness of the system to examine the second group of research propositions. Finally, there were some duplicate questions to check student response consistency. A summary of the key results is included here; however, providing all of the questions here would be redundant and too lengthy.

To begin with the pedagogical impacts of AI grading, the first key question was "Did the students like the ability to receive immediate feedback?" The response was an overwhelming yes. Over 97% of the respondents answered that they either strongly agreed or somewhat agreed (70 of 72). Only two students were neutral or opposed. When asked specifically about AI grading portion, the students were still very positive. Figure 7 – I Liked the AI Grading presents the students' responses to this question (5-Strongly Agree to 1-Strongly Disagree). Although the result was not as strong as the immediate feedback question, 83.3% of the respondents had a positive view and only 9.7% were opposed. The combination of the immediate feedback and AI grading were supported by the vast majority of the students across all classes, both genders and regardless of GPA.

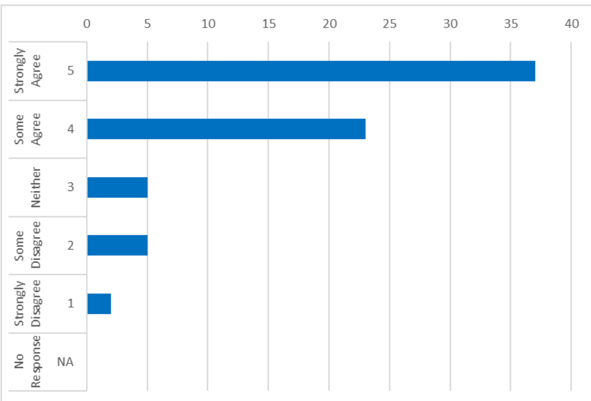

Figure 7 – I Liked the AI Grading

To continue to examine the pedagogical impacts, the students were asked to evaluate their view of AI grading compared to the faculty members' grading system. Here, there was a cross section of answers. The students did not have a strong opinion on which, if either, was better. Figure 8 highlights this finding. It is interesting that the students were not willing to completely trust the AI system. However, clearly some students preferred the AI compared to the faculty members' grading processes. There are many possible causes for this finding and they could be a subject to an entire paper in itself. However, some of the main comments included a lack of clear understanding of how the AI system worked and how the professors would incorporate the grading into their classes. It should also be noted that the faculty members used the results from the AI grading differently in their individual classes. Finally, some of the variation is clearly due to the belief that a minimum score on the AI or curiosity score would earn the student full credit for the assignment which was not the case.

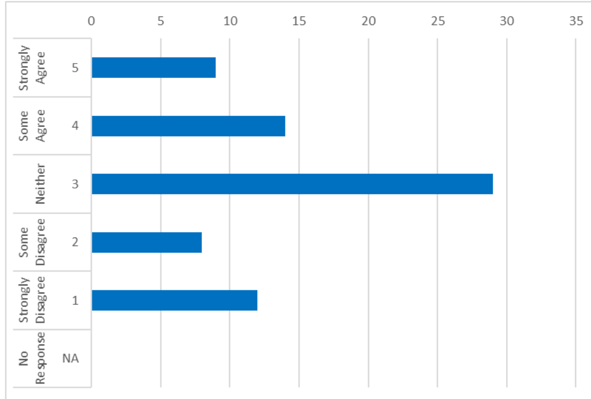

Figure 8 – I Prefer Instructor verse AI Grading

One of the key goals of the use of AI was to encourage deeper thought and learning in the AOD. Figure 9 demonstrates that 55.6% of the students reported a positive impact. It should be noted that the wording of the question did not include that the use of the AI could have had a negative impact. Therefore, the fact that a majority of the students responded that it increased their learning experience by using the AI system. This is a tremendous benefit to the overall class. Furthermore, there are a number of second order effects that may have not been obvious to the students. First, if over half were improving their posts and learning, then the remaining students were reading more well-developed submissions and by default would have an increased learning experience. Also, even if a student did not feel his or her learning was better, the level of competition within the course likely encouraged them to improve their work and hence their individual learning. Finally, the impact on learning was likely the most significant on the students in the middle of the grade distribution. The very high and low achieving students may not have gained as much due to their already being on the extreme ends of the spectrum. These are additional areas for future research. However, the initial finding was strong enough for the faculty to continue to use AI.

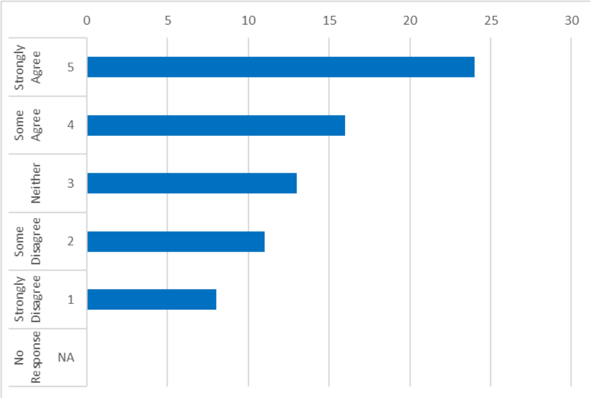

Figure 9 – The AI Encouraged Me to have Deeper Thoughts

In terms of the next group of research propositions, there were a number of questions about the mechanics. A summary table condenses these findings due to space limitations. Table 2 – Process Results for Using Packback highlights most of the key findings. It should be noted that since the specific Packback program was used, the findings may or may not apply to other AI AOD programs. The table is organized with a shortened version of each question followed by the students' responses. The responses are organized by positive, then neutral, and finally negative based on the questions. For example, the first question was "Is the use of Packback more difficult since it was not incorporated into the LMS?" The majority of the students did not think it was more difficult; so, the "No" finding is a positive for this question.

Question	Responses (not all add 100% due to non responses)		
More difficult since not in LMS?	No – 56.66%	Neutral – 18.06%	Yes – 15.28%
I gamed the system for grade?	No – 52.50%	Neutral – 18.06%	Yes – 19.44%
I had technical problems?	No – 94.45%	Neutral – 1..39%	Yes – 2.78%
Use of AI's features was fun?	Yes – 62.50%	Neutral – 29.17%	No – 8.3%
The cost was worth the benefits?	Yes – 58.33%	Neutral – 23.61%	No – 18.05%
It was easier to use that the LMS?	Yes – 61.11%	Neutral – 18.06%	No – 20.83%
Overall satisfied with AI?	Yes – 83.31%	Neutral – 5.56%	No – 11.11%

Table 2 – Process Results for Using Packback

For all of the specific, mechanical types of questions, the majority of the students had a positive response. There were virtually no technical problems with the AOD. The two students that did have issues both were using VPNs to block their identities which when turned off, the Packback website worked fine. A concern with any AI grading system is the students will game the process for a better score. For example, Packback allows you to put a link into your response which will help your curiosity score. However, the link could be very much off topic and the student still gets the points. Therefore, it was interesting to see that almost 20% of the students did game their posts at some time during the semester. This both highlights that AI is not perfect at this point and faculty member involvement is still needed. Next, the students enjoyed using the AI system. While not an extremely important point, a positive experience using the system will likely encourage additional use when compared to a negative experience. One of the faculty members' key concerns was cost. Students already pay for a LMS and have premium pricing in the program. The majority of students did think Packback was worth the additional cost. However, the written comments did state that since the costs was outside the university, some students' employers would not reimburse it which led to their dislike. Finally, the summary question of overall satisfaction was very high at 83% of the students. These findings coupled with the pedagogical results highlight that the AI grading and feedback had a successful proof of principle test in the spring semester.

7. CONCLUSIONS, RECOMMENDATIONS, AND FUTURE OPPORTUNITIES

As previously stated, this was an initial test of the Packback. AI grading and feedback systems were considered a qualified success based on both the faulty members' and students' feedback. The majority of the propositions were supported with summary data. The students reported they learned more, applied more effort and were satisfied with the AI system. Faculty members were also pleased in general with the clear improvement with student work but were not as positive due some of the technical items due to the stand-alone nature of the AOD, content organization abilities, and cost. However, as stated before, the overall positive aspects were more than enough to adopt the Packback AOD again in the upcoming fall semester.

Based on the use of Packback for the first time, there were a few clear learning points that will be applied before the next iteration. These items are shared as recommendations for any faculty member planning use either Packback or another AI grading system. First, both the syllabus and faculty member should clearly articulate exactly how the AI scores will be incorporated into the overall grading scheme. The key is not whether the grade is all AI based, a hybrid or all faculty member derived, but rather which will be used. That will help to clarify the students' expectations. The authors recommend a hybrid that is given to the students at the beginning of the class (i.e., 50-50% faculty-AI scores based upon …) Next, the benefits and challenges of using the AI should be stated at the beginning of the semester. For example, one of the challenges of Packback is the lack of modules or submodules to separate different discussions. This will likely be addressed in future updates of Packback. In the spring classes we developed a numbering system that aligned LMS module numbers and specific posts were to include the number in the title. Again, it was a simple item, but helped to provide clarity. Finally, faculty member expectations should be restrained. There was a minor reduction in workload; however, each student post still needed to be read and evaluated. The primary benefits were improved student posts and more timing flexibility of when to review the postings. The asynchronous portion of the board and that the AI will fill in for the instructor should be communicated at the beginning. The instructor should clearly indicate that he or she will be reading all the posts to ensure that students are not trying to game the system with unnecessary photos, videos, links, etc.

Also, due to the introductory nature of this paper, there are a number of findings that should be more rigorously tested. For example, the sample size was not large enough to do specific demographic tests beyond a cursory evaluation. Another area for future examination is the impact on undergraduate students and use in a F2F class. It is likely that the findings would remain the same; however, the differing nature of these groups and class settings might have significant impacts on the results. Another key point is that the AI software continues to evolve. A future examination of the ability of students to game the system when less faculty review is conducted would help instructors to moderate their time and effort in grading.

The last point about improving nature of the software is a key closing point. As more companies enter the field and traditional LMS recognize the benefits of AI grading, it is likely the quality and options for AI grading both inside AOD and in other areas will improve dramatically over the next ten years. Faculty members should begin to realize the potential pedagogical and workload benefits. Just as PowerPoint changed classrooms 20+ years ago, and real-time media is reshaping them today, AI will change the learning experience over the next few years. The same question arises of how should a faculty member apply this new technology to maximize it benefits while minimizing its weakness for both students and instructors.

8. REFERENCES

Ackerman, David S. and Barbara L. Gross (2021). "Synchronous Online Discussion Board as a Primary Mode of Delivering Marketing Education: Responding to the Covid-19 Pandemic and Beyond." Marketing Education Review, March, 1–10.

Bridgeman, Brent & Thomas Quinlan (2009). "Differences between Humans and Computers for Evaluating and Teaching writing." In Proceedings of 2009 International Conference on ELT Technological Enterprises, pp, 3-24.

Chen, Pu-Shih Daniel. Amber D. Lambert and Kevin R. Guidry (2010) "The Impact of Web-Based Learning Technology on College Student Engagement." Computers & Education, 54, 1222-1232.

Cifuentes, Lauren, Karen L. Murphy, Rhoda Segur, and Sailaja Kodali, (1997). "Design Considerations for Computer Conferences." Journal of Research on Computing in Education, v. 30(2), pp. 177–201.

Day, J., H. Lou,. and C. Van Slyke. (2004). "Instructors' Experiences with Using Groupware to Support Collaborative Project-Based Learning." International Journal of Distance Education Technologies. 2(3). 11-25.

Delgado, Heloisa Orsi Koch, Aline de Azevedo Fay, Maria Jose Sebastiany and Asafe Davi Cortina Silva (2020). "Artificail Intelligence Adaptive Learning Tools: The Teaching of English in Focus" Brazilian English Language Teaching Journal, 11(2), 1-19.

Dennen, Vanessa Paz (2005). "From message posting to learning dialogues: Factors affecting learner participation in asynchronous discussion." Distance Education, 26, 127–148.

Douglas, Tracy, Allison James, Louise Earwaker, Carey Mather and Sandra Murry (2020). "Online Discussion Boards: Improving Practice and Student Engagement by Harnessing Facilitator Perceptions." Journal of University Teaching & Learning Practice, 17(3), 1–14.

Frey, Barbara A., and Linda C. Wojnar. (2004). "Successful Synchronous and Asynchronous Discussions: Plan, Implement and Evaluate". Educause Information Resources Library. Accessed at http://www.educause.edu/ir/library/pdf/MAC0426.pdf on 9 Jun 2021.

Ghoneim, Nahed Mahammed Hahmoud and Heba Elsayed Abdelsalam Elghotmy (2020). "Using an Artificial Intelligence Based Program to Enhance Primary stage Pupils' EFL Listening Skills," Sohag Universit Journal of Education, 83(1), 1-32.

Gill, Grandon (2006). "Asynchronous Discussion Groups: A Use-based Taxonomy with Examples." Journal of Information Systems Education, v. 17, n. 4, p. 373–383.

Hew, Khe Foon, Wing Sum Cheung, and Connie Siew Ling Ng (2010). "Student Contribution in Asynchronous Online Discussion: A Review of the Research and Empirical Exploration." Instructional Science, v. 38, n. 6, p. 571–606.

Hinton, Leone and Julie Bradshaw (2004). "Benefits of an Online Discussion List in a Traditional Distance Education Course," Turkish Online Journal of Distance Education, vol. 5, no.3, pp. 8-16.

Liu, Tiaoqiao, Wenbiao Ding, Zhiwei Wang, Jiliang Tang, Gale Yan Huang and Zitao Liu (2019). "Automatic Short Answer Grading via Multiway Attention Networks" Working Paper, https://arxiv.org/abs/1909.10166 accessed June 2021.

Kafka, Alexander C. (2020). "Shock, Fear, and Fatalism: As Coronavirus prompts College to Close, Students grapple with Uncertainty." The Chronical of Higher Education, Mar. 12, https://www.chronicle.com/article/shock-fear-and-fatalism-as-coronavirus-prompts-colleges-to-close-students-grapple-with-uncertainty/ accessed Jun 2021.

Murphy, Elizabeth and Elizabeth Coleman (2004). "Graduate Students' Experiences of Challenges in Online Asynchronous Discussions." Canadian Journal of Learning and Technology, *30*(2), spring.

Newman, D. R., Chris Johnson, Brian Webb and Clive Cochrane, (1997). "Evaluating the Quality of Learning in Computer Supported Co-operative Learning." Journal of the American Society for Information Science, 48. 484-495.

Packback. (2021). https://www.packback.co/ accessed January 2021.

Pecka, Shannon, Kendra Schmid and Bunny Pozehl (2014). "Psychometric Testing of the Pecka Grading Rubric for Evaluating Higher-Order Thinking in Distance Learning." AANA Journal, 82(6), 449–456.

Phillippi, Julia C., Mavis N. Schorn, and Tonia Moore-Davis, (2015). "The APGAR rubric for scoring online discussion boards." Nurse Education in Practice, v. 15, n. 3, p. 239–242.

Ratcliff, Jennifer J., Korrine I. Minister and Chelsea Monheim (2014). "Engaging students in an online format during the COVID-19 pandemic: A jury voir dire activity." Scholarship of Teaching and Learning in Psychology, February.

Sandars, John, Roaqueel Correia, Mary Dankbaar, Peter de Jong, Poh Sun Goh, Inga Hege, Ken Masteers, So-Young Oh, Rakesh Patel, Kalyani Premkumar, Alexandra Webb and Martin Pusic (2020). "Twelve tips for Rapidly Migrating to Online Learning during the COVID-19 Pandemic," MedEdPublish, April 29, https://doi.org/10.15694/mep.2020.000082.1 accessed June 2021.

Smidt E, McDyre B, Bunk J, Li R, Gatenby T. (2014). "Faculty Attitudes about Distance Education." IAFOR Journal of Education, 1(2) 181-209.

Sull, Errol C., (2014). "Discussion Boards, Grades, and Student Involvement." Distance Learning. 11(3), 55-57.

Tsai, Min-hsiu (2012) "The Consistency Between Human Raters and an Automated Essay Scoring System in Grading High School Students' English Writing," Action in Teacher Education, (34)4, 328-335, DOI: 10.1080/01626620.2012.717033

Vonderwell, Selma (2003) "An Examination of Asynchronous Communciatin Experiences and Perspectives of Students in an Online Course: A Case Study," The Internet and Higher Education, 6(n), (first quarter), 77-90.

Walker Kristen, Mary T. Curren, Tina Kiesler, H. Bruce Lammers and Jamie Goldenson, (2013). "Scholarly Networking Among

Business Students: Structured Discussion Board Activity and Academic Outcomes." Journal of Education for Business, 88(5), 249–252.

Yeh, Y., Liou, H. C., & Yu, Y. T. (2007). "The Influence of Automatic Essay Evaluation and Bilingual Concordance on EFL Students' Writing." English Teaching and Learning, *31*(1), 117-160.

Appendix:

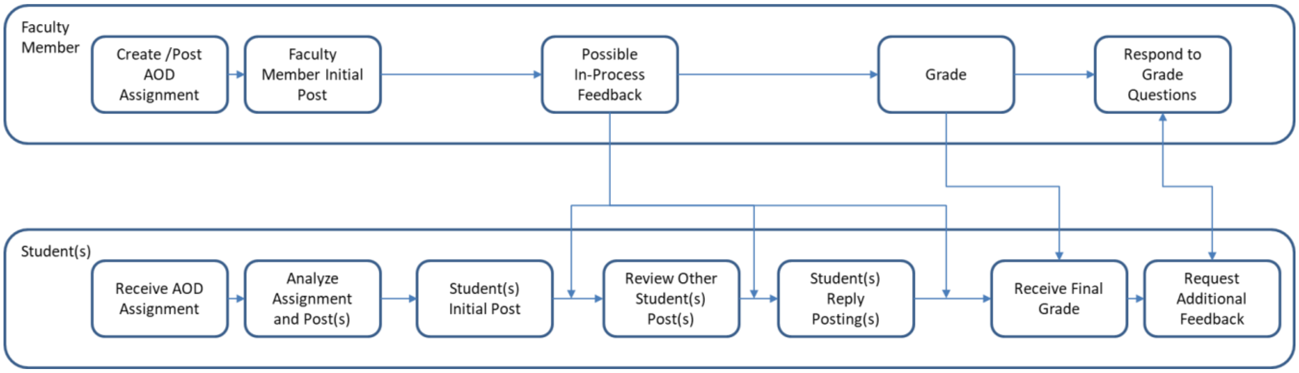

Figure 2 – Traditional AOD Interaction between Faculty and Student(s)

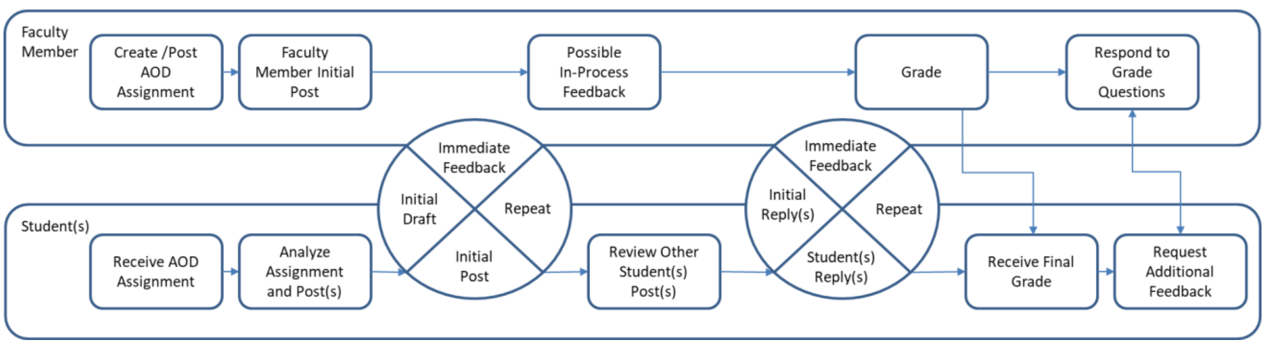

Figure 4 – AI Feedback Modified AOD Process Model

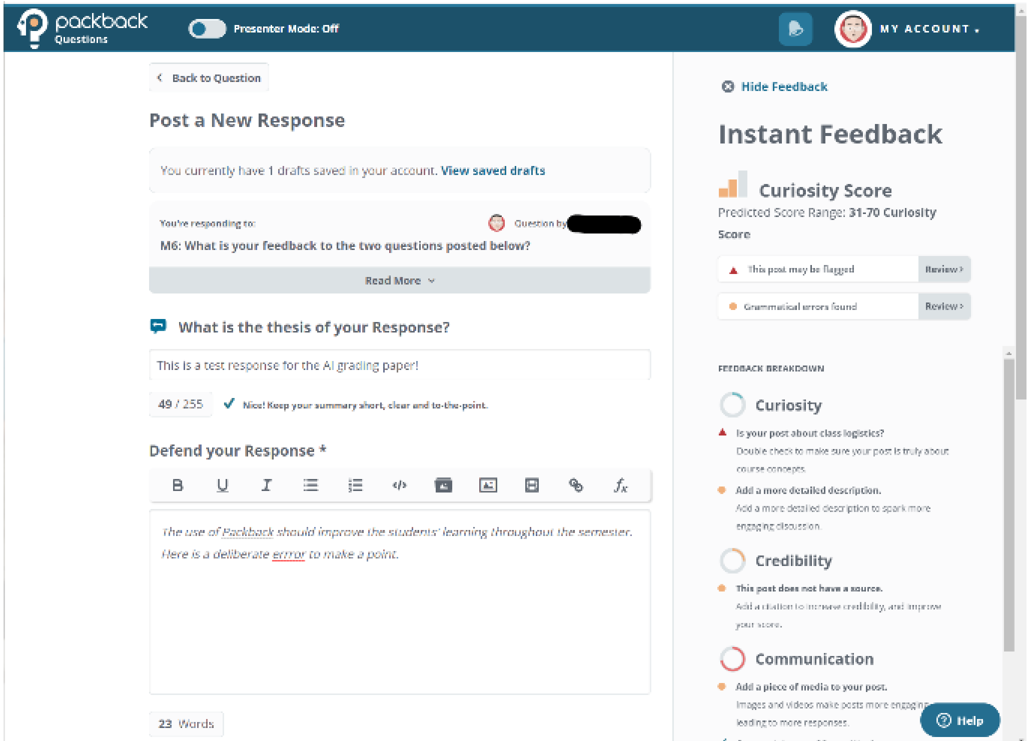

Figure 5 – Student Post Screen

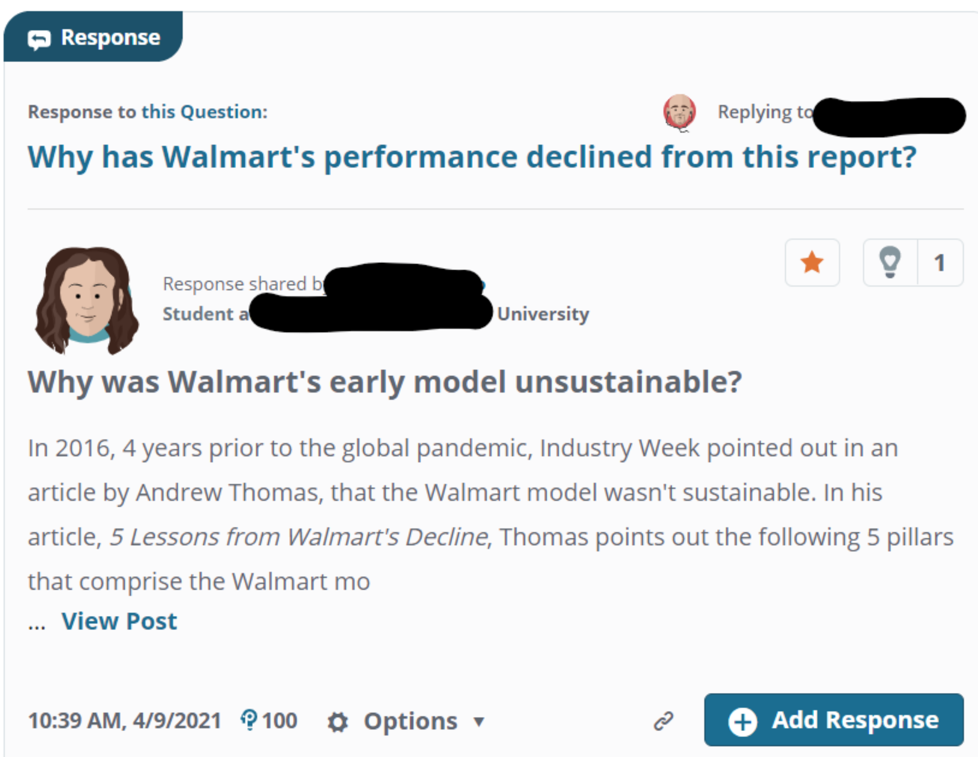

Figure 6 – Student Post on Discussion Board