Perusall: Harnessing AI Robo-Tools and Writing Analytics to Improve Student Learning and Increase Instructor Efficiency

Allison S. Walker, *High Point University*

Structured Abstract

- **Aim:** The data collected by robo-tutor systems can provide human instructors with valuable insights into individual learners' needs and give them more time and freedom to respond to student writing in ways that are informed by writing analytics and tailored to each "apprentice scholar" (Sommers, 2006). As an example of writing analytics and artificial intelligence (AI) in practice, the author will address significant findings from a research study on the use of Perusall in undergraduate general education writing and literature classes at High Point University. Perusall, a social reading and annotation platform available as a third-party plug-in through most learning management systems (LMS), allows students to annotate readings digitally and share their insights with other students. The platform uses writing analytics to automatically score student annotations, thereby enabling instructors to spend more time engaging students in meaningful conversations about their work. The platform also collects data about student reading and annotation patterns, data that instructors can use to further improve student performance in the classroom.

- **Problem Formation:** While Sommers (2006) suggests that an "apprentice scholar" approach to student writing is the most successful feedback strategy for instructors, the intensive labor and time constraints of this practice have remained largely unchanged for the past century (Gilbert, 1922; Merrill, 1992; Sommers, 2006). Despite various technological advances in computer

processing and digital mediation of the writing process, successful assessment and response at the level described by Sommers still requires approximately 20 minutes for every piece of student writing collected, though some recent strategies such as screencasting may reduce that response time to approximately 10-15 minutes per student (Walker, 2017). With increased class sizes and millennial student populations uniquely primed for the instant gratification of social networking, robo-tools could be allies in a quest to improve student writing and aid digital instruction. After several successful semesters of experience using Perusall, the author began to recognize marked student learning gains as a result of this writing analytics technology. How might AI-enhanced robo-tools like Perusall and the Spector et al. (2016) "stealth assessment" writing analytics collected through the back-end of such robo-tools be effectively leveraged to enhance student reading, writing, and critical thinking skills? How might Perusall improve the "apprentice scholar" (Sommers, 2006) relationship between instructors and students and capitalize on the benefits of collective agency (Kim, 2013; Kim & Baylor, 2006) among students who are digitally connected during various reading and writing tasks? Would such a tool also lessen the assessment burden of human instructors by accurately automating the assessment loop? In order to address these questions, the author developed a survey measuring student perceptions of the Perusall platform and collected data in the form of student grades on assignments associated with the Perusall readings as well as analytics from the Perusall platform itself.

- **Information Collection:** This research note reports findings from a study of 125 undergraduate students in two sophomore-level English classes during the 2017-2018 and 2018-2019 academic years at High Point University. The courses were conducted in both online and face-to-face formats over 16-week semesters in the fall and spring, and the courses were capped at either 15 or 23 students per course section. Data was collected through Perusall and through the grade center of the LMS used by High Point University. Additional findings were collected through a reflective narrative survey of students after completion of the Perusall reading assignments.

- **Conclusions:** The author provides an in-depth analysis of students' experiences with Perusall and the impact of this technology on their overall learning. Student grade trends and narrative reflections are examined, and examples of writing analytics gathered by the Perusall platform are also discussed. The author focuses on the impact of AI reading and writing technologies in relation to instructor labor practices and student learning. Some conclusions are also drawn regarding the user reliability and effectiveness of Perusall.

- **Directions for Further Research:** While this research note provides compelling evidence to support the use of Perusall in undergraduate classrooms, there are also many opportunities for continued study, particularly within the emerging field of writing analytics. To facilitate future research in this area and inspire the collective creative imaginations of her readers, the author has included a speculative scenario in which predictions are made about the future of robo-tools and robo-tutoring programs based on the findings of this research study and an analysis of our shared computer-mediated writing achievements as researchers and educators over the past century of reading and writing instruction.

Keywords: artificial intelligence (AI), educational data mining (EDM), intelligent tutoring systems (ITS), learning management system (LMS), massive online open course (MOOC), natural language processing (NLP), pedagogical agent as learning companions (PAL), Perusall, Zone of Proximal Development (ZPD), writing analytics

1.0 Aim

AI-enhanced robo-tutor programs can provide agile, adaptive feedback about student reading habits and surface-level writing concerns as well as more nuanced formative feedback akin to that given by a human tutor to aid student critical thinking, close reading, collaboration, self-efficacy, motivation, and mastery of higher-order academic writing skills. The data collected by robo-tutor systems can provide human instructors with valuable insights into individual learners' needs and give them more time and freedom to respond to student writing in ways that are informed by writing analytics and tailored to each ''apprentice scholar'' (Sommers, 2006). An AI robo-tutor program could therefore result in stronger student reading and writing skills and fewer faculty hours spent on assessment. As an example of writing analytics and artificial intelligence (AI) in practice, the author will address significant findings from a research study on the use of Perusall in undergraduate online general education writing and literature classes at High Point University. Perusall, a social reading platform available as a third-party plug-in through most learning management systems (LMS), allows students to annotate readings digitally and share their insights with other students. The platform uses writing analytics to automatically score student annotations, thereby offering instructors more time to engage students in meaningful conversations about their work. The platform also collects data about student reading and annotation patterns, data that instructors can use to further improve student performance in the classroom.

2.0 Problem Formation

In November of 2017, the website *Inside Higher Ed* asked whether faculty need automated help grading online discussions. Author Lindsay McKenzie offered a balanced profile of a new

Blackboard LMS tool that utilizes a computer algorithm to analyze student posts and generate a discussion forum recommended grade for instructor use in online courses where forums are often a tedious but necessary element of evaluating student participation. Though the article itself concluded with a quote from the new Blackboard algorithm's creator, John Whitmer, who said, "Our focus is to provide a tool to assist the human leading the class, not replace them. Of course, there is some risk with any automation, but we believe that the benefits of increased feedback outweigh these risks," a quick scroll to the comments section below the article indicated few commenters believed Whitmer's claim (McKenzie, 2017). Here are four responses from *Inside Higher Ed* readers who could, theoretically, find themselves in instructional situations where this new Blackboard tool would be available:

> What's next--a tool that does everybody's thinking for them?
>
> Once again, we have a tool that replaces the idea of intellectual development with a tool that MEASURES data collection. Education is NOT a noun count.
>
> Faculty and teaching have nothing to do with this. It's about $$$$$$$$$$$
>
> Can college administration also use the tool to grade instructors' participation in discussions? Beware of Unintended consequences. (McKenzie, 2017)

Though some commenters tried to steer the conversation toward more constructive discourse, the results were mixed. Whitmer himself entered the conversation to say:

> Wow, some strong perspectives and points of view here. I'll enter the waters with a little trepidation to add a little perspective. Before joining Blackboard, when I worked for the California State University and California Community College systems, and helped faculty learn online tools, I frequently heard that evaluating discussion forums was very difficult and onerous. In current conversations, we hear that faculty are spending their time counting posts and otherwise giving surface-level feedback.
>
> The purpose of creating this tool was to automate that part of feedback, so that faculty can focus on more substantive and meaningful reviews, such as many people here have suggested. Our designers have created the tool to require faculty to review (and possibly edit) each and every "grading recommendation" before it is published, and to do so at a per-student level.
>
> We'll work directly with faculty and some of our institutions to evaluate whether this tool is helpful and ensure that it enhances real effective feedback.
>
> HTH. My post received a recommended score of a "B-." (McKenzie, 2017)

One can't help but note the irony of HTH (which can mean either "Hope That Helps" or "How the Hell") as indicative of the polarized response these new tools can generate among some potential users, a response perhaps given before those potential users ever try the tool themselves.

Despite the trepidations of commenters like those above, this author wishes to consider the affordances that AI robo-tutor might achieve. Programs such as WriteLab, M-Write, Eli Review, and others are already doing this work to some degree. Here, for example, is a synopsis from the website of M-Write, a platform developed at the University of Michigan:

> Writing-to-learn pedagogies engage students by asking them to explain what they know, to interact with one another through peer review, and to learn through a revision process. M-Write will combine conceptual writing prompts, automated peer review with rubrics, natural language processing, and personalized feedback using ECoach to create an infrastructure for writing at scale. ("Origin Stories Showcases M-Write")

Following is a description of Eli Review, a similar digital initiative at Michigan State University:

> Eli Review (also known simply as "Eli") was invented to support evidence-based teaching practices and facilitate rich peer learning environments. Invented in writing classrooms at Michigan State University, Eli became necessary because no technologies existed to support the feedback and revision cycles that lead to better learning and more effective writers. ("About Eli Review")

WriteLab is yet another digital platform that promises "Immediate Feedback. Revision Mastery. Measurable Growth" on its website. This program of "instant AI-assisted feedback powered by Natural Language Processing that refines and strengthens student writing," is endorsed by *Forbes Magazine* and Stanford University, among many others ("WriteLab and Chegg: Improve your writing in less time").

Lastly, let us consider Perusall (www.perusall.com). With a mission of "every student prepared for every class," this platform makes big promises. The platform allows instructors to choose and assign readings from textbooks, web-based articles, and PDFs, and, "with novel data analytics, Perusall automatically generates optimal student groupings and social interactions, grades students' engagement to ensure they are prepared for class, and nudges those who need help to keep everyone on track." Developed by Eric Mazur, Gary King, Paul Lukoff, and Kelly Miller at Harvard University, Perusall hopes to turn solitary reading practices into interactive learning experiences that "proactively engage students with each other" while giving them "(automated) personalized guidance" that will "motivate your whole class" ("Perusall: Every student prepared for every class").

Before we can evaluate the effectiveness of AI-enhanced robo-tutoring platforms such as Perusall, it is important to establish some of the best practices in writing and reading instruction in a traditional human-to-human interface. In the field of writing instruction, Merrill et al. (1992), in their article, "Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems," determined that "cognitive apprenticeship" is the most effective approach, a means of scaffolding assistance then scaling it back as students gain facility with a particular concept (p. 283). This was an idea reiterated by composition scholar Nancy Sommers (2006) in her NCTE article, "Across the Drafts," in which she identifies the most salient

characteristic of feedback given by instructors as those comments which treat the work as that of an "apprentice scholar" (p. 250). She writes, "feedback plays a leading role in undergraduate writing development when, but only when, students and teachers create a partnership through feedback—a transaction in which teachers engage with their students by treating them as apprentice scholars, offering honest critique paired with instruction" (p. 250). Merrill et al. also delineated four distinct types of effective intelligent tutoring systems: directive/controlled, exploratory, coaching, and model tracing (p. 278). Each tutoring approach capitalizes on a strategy used by effective human tutors in Sommers' longitudinal study of Harvard student writers across four years of undergraduate writing. Both Merrill et al. and Sommers agree that successful human tutors engage in a variety of approaches in order to guide students through learning by doing. The best teacher-student relationships inculcate an open mindset for optimal learning while also providing student motivation and support. While the intelligent tutoring systems explored by Merrill et al. in 1992 were rudimentary compared to those like M-Write, Eli Review, and Perusall on the market today, these foundational qualities of effective tutoring remain evergreen. "Experienced human tutors maintain a delicate balance, allowing students to do as much of the work as possible and to maintain a feeling of control, while providing students with enough guidance to keep them from becoming frustrated or confused" (Merrill et al., 1992, p. 280). With this in mind, let us entertain the possibility that AI robo-tutors might accomplish this same "delicate balance" (p. 280).

For example, in an entirely technology-mediated tutoring interface, Kim and Baylor (2006) successfully implemented the "cognitive apprenticeship" and "apprentice scholar" approaches described above. To further bolster the viability of computerized writing tutors in educational situations where class size, geographic location, or other physical constraints limit the instructional time between student learners and their human instructors or tutors, "A Social-Cognitive Framework for Pedagogical Agents as Learning Companions" by Kim and Baylor (2006), utilizes key evidence from Vygotsky, Bandura, and Piaget to develop computerized "PALs" (pedagogical agents as learning companions). These avatar PALs have distinct personalities and appearances in order to provide four types of social interaction with the student user: as a tutor, an expert, a motivator, and a peer collaborator (p. 571). Each avatar is designed to fulfill a specific role in the student's learning, and it is thus tailored to meet multifaceted cognitive needs at various stages of learning, thereby creating a virtual environment in which learning becomes a social process (p. 576).

Kim and Baylor (2006) point to the work of Vygotsky regarding the "zone of proximal development" (ZPD), the most optimal cognitive space for learning, located just outside a student's intellectual comfort zone but still within cognitive reach. This ZPD doesn't overtax a student's working memory load by making the task too difficult, nor does it create learning experiences that are too easy or repetitive (p. 577). The ZPD is most accessible through interactive learning tasks in which learners must negotiate meanings with other learners (p. 576). "A PAL can utilize a variety of discourse functions, such as suggestion, argument, confirmation, and questioning to scaffold learners in the zone of proximal development" (Kim & Baylor, p.

583). Kim and Baylor (2006) argue that PALs can achieve ZPD learning by "extending the cognitive range of learners, allowing the learners to accomplish tasks not possible otherwise," and selectively aiding the learner only when needed (p. 577).

Kim and Baylor (2006) also point to the work of Paiget on the roles cognitive conflict and equal power relations can play in the cognitive development of learners. They argue, "cooperation and free discussion play an essential role in acquiring and constructing knowledge because they establish the most favorable conditions for counteracting an individual's egocentrism" (p. 576). While a computerized PAL is less ideal than a cooperative interaction among students who are physically present together, the PAL avatar can satisfy these cognitive needs in environments where the real thing isn't readily available, thus allowing human learners an opportunity to consider different, albeit hypothetical, positions outside of their own.

Lastly, Kim and Baylor (2006) look to the work of Bandura on the impact of personal, proxy, and collective agency among student PAL users. Personal agency "emphasizes learners' control over and self-regulation of their learning tasks" (Kim & Baylor, (2006, p. 579), so PALs must be designed to respond to student requests for assistance without directing that interaction overtly, thereby increasing a student's sense of self-efficacy and improving cognitive gains. Proxy agency allows a learner to utilize the expertise of others in order to accomplish a learning task; it is thus a social and collaborative function of learner agency in which the presence of another student, whether real or virtual, enhances student motivation (p. 579). To achieve collective agency, Kim and Baylor (2006) propose a "multiple-PAL environment" of virtual tutors, experts, coaches, and peers with "domain-specific skills" that "enhance the social richness of the environment" and collectively aid the learner in reaching learning goals (p. 580). These findings were instrumental in the development of the hypothetical robo-tutor scenario proposed in the Directions for Further Research section of this research note.

Building on these previous findings, Kim (2013) used PAL technology to increase short- and long-term reading comprehension. When the PAL avatars acted as virtual peers, modeling the cognitive process of active questioning, they enhanced students' curiosity, engagement, and willingness to tackle reading challenges (p. 59). Students thus benefited from individualized guidance provided by PALs because it allowed them to learn at their own pace and gradually scaffold active and intentional reading strategies into their existing reading habits (p. 60).

Another way AI-enhanced robo-tools can facilitate the "cognitive apprenticeship" of student writers is through the use of big data. Authors Spector et al. (2016) elaborate on the importance of formative feedback and assessment, including the novel approach of "stealth assessment" (p. 65), a strategy made possible through the advent of educational data mining (EDM) technology that allows for embedded and continuous, unobtrusive data collection about learners' habits and strategies. In "Technology Enhanced Formative Assessment for 21st Century Learning," Spector et al. identify three types of effective formative assessment (instructor to student, peer to peer, and self-reflective) and then propose five key things formative assessment should accomplish: providing a benchmark of current student knowledge, identifying the strengths and weaknesses of a learner in relation to the learning outcomes of the course, assessing the progress of the

learner, setting goals for future learning, and guiding learners through the learning process (p. 61). They note that the most successful intelligent tutoring system (ITS) programs will allow students to become independent learners by creating adaptive formative assessment programs built on AI machine learning techniques (p. 64) that promote students' digital and reasoning literacy (p. 66).

Studies such as those outlined above demonstrate that robo-tutoring is not only possible, it's already here and has been for some time. Machine learning, natural language processing, writing analytics, and educational data mining provide an abundance of resources to support effective reading and writing instruction.

One example of an AI robo-tool that provides successful integration into existing reading and writing instruction infrastructure is Perusall. While Perusall focuses on the critical reading component of student learning and is thus a less robust form of robo-tutoring than those provided by such programs as Eli Review, M-Write, and WriteLab, it is currently more widely available to instructors across various LMS platforms and therefore more accessible for use in research studies such as the one described in this research note. While Perusall may lack the complexity of other robo-tutoring platforms, it does incorporate many of the key elements of optimal digital instruction discussed above. For example, the platform collects "stealth assessment" (Spector et al., 2016) data from users that instructors can then interpret for improved instructor-to-student communication that is tailored to individual student learning needs and habits. Perusall also captures all three types of successful formative assessment confirmed by Spector et al., including: instructor to student (instructors can post their own annotations and respond to student questions directly, in either a public or private response), peer to peer (annotations are composed and viewed by all students, and conversation threads among subsets of students may develop organically as students engage one another directly), and self-reflective (students may also compose private annotations and notes on any reading that they can save and store privately). The Perusall platform also capitalizes on the personal, proxy, and collective agency of its users by connecting students with each other during the previously solitary task of close reading, thereby fostering a sense of community (Kim & Baylor, 2006). To advance student agency, the platform offers student readers immediate access to all of the annotations posted by their peers and even the instructor. When reading those posts, students can respond directly to them with text comments of their own, or they can utilize familiar social media tools such as emojis, weblinks, and GIFs. Students can post questions about the reading and get automatically alerted when another student (or the instructor) posts an answer to that question. Student annotations can fulfill various ZPD discourse functions by suggesting, arguing, confirming, and questioning the reading and each other (Kim & Baylor, 2006). Students can upvote annotations they found particularly helpful or thought-provoking, and they can follow the students who post the most helpful comments, thereby enacting Piaget's learning goals of cooperation and free discussion (Kim & Baylor, 2006). In this way, students fulfill the roles of tutor, expert, motivator, and peer collaborator (Kim & Baylor, 2006). They become role models and peer tutors for each other, engaging in critical conversations about the readings at a granular level and fostering a deeper

understanding of the material and an opportunity to recognize different nuanced perspectives on it. The platform utilizes NLP and automated scoring algorithms to assess student annotations, thereby alleviating some of the instructional burden for a human reader and freeing up valuable faculty time that can then be applied to other "apprentice scholar" (Sommers, 2006) instructional and assessment tasks. Lastly, Perusall seamlessly integrates into existing LMS programs such as Blackboard, Moodle, and Canvas, allowing users to interact with the robo-tool platform from within an existing digital learning infrastructure, eliminating the need for additional technology instruction and syncing with the many varied digital devices 21st century students use daily both in and out of the physical classroom environment.

3.0 Information Collection

The sample size of this study was 125, with students enrolled in separate courses instructed by the same professor and with the same readings and annotation assignments presented in the same order. These courses occurred over two academic years but were otherwise equivalent, with student enrollments balanced, the instructor matched, and readings and rubrics identical across all sections. The students had to meet a standard prerequisite for the course by passing or placing out (with an AP score of 4 or better) of the university's first-year composition and rhetoric course.

Of the 125 students in this study, 71 in four course sections utilized Perusall for course readings, while 54 in two other course sections did not. (See Appendix A for a detailed list of the selected readings used by the instructor in this study.) These readings were course-specific academic excerpts, presented in the form of PDF documents, chosen and uploaded by the instructor to the Perusall platform and the LMS platform Blackboard. In two course sections, 28 students received extra credit for just using the Perusall platform to complete course readings; the extra credit was not contingent upon the automated score given by the Perusall AI assessment algorithm. In this condition, if a student received a Perusall algorithm score of zero on all the assigned Perusall readings, the student still received extra credit for using Perusall. In two other course sections, 43 students received extra credit contingent on the Perusall score, with a maximum of 5% course credit offered for a perfect score on all assigned Perusall readings. In this condition, if a student received a zero on all the readings, that student would not receive any extra credit, even though the Perusall platform was used to read the assigned texts. In two other course sections, 54 students acted as the quasi-control group by reading assigned course texts as printable PDF documents, without using Perusall at all.

Quantitative data was collected based on student Perusall scores on six reading assignments distributed throughout the semester, course grades on reading quizzes and writing assignments associated with those Perusall readings, a capstone course project that required students to synthesize the Perusall readings, and overall final grades in the course. Qualitative data was also gathered through a reflective narrative survey after completion of the Perusall readings. In addition to this, the author was also able to collect analytics from the Perusall platform itself, including "stealth assessment" (Spector et al., 2016) data such as confusion reports detailing the

most upvoted questions asked by students in their annotations, heat maps detailing the date and time of student annotations, active and passive reading times for each student, and a graph of collective student time spent on each page of an assigned reading. The data collection algorithms and data mining techniques utilized by the Perusall platform are considered patented, proprietary code, owned and protected by the privately-held Perusall company, and thus will not be evaluated in this study.

Demographically, the students in this study were undergraduates at High Point University, a small, liberal arts institution in North Carolina that offers two sophomore-level, 4-credit-hour English courses entitled "Narrative Medicine" (ENG 2200) and "Professional Writing in the Medical Humanities" (ENG 2130), capped at 23 and 15 students respectively and conducted over 16-week semesters in both the spring and fall. None of the students in these courses were repeating the course for credit, and all fulfilled either a general education or pre-pharmacy professional program requirement by taking the course. Less than 10 students enrolled in but did not complete the course, and those students are not represented in the data from this study because they withdrew from the course before completing the study. Students at High Point University are generally from an upper-middle to upper-class socioeconomic background, and the gender distribution in this study was 60% female and 40% male. High Point University is a predominately white institution, and the classes in this study included less than 10% minority representation. The project was approved by the High Point University Institutional Review Board (IRB) at the level of an "expedited" study with a three-year protocol approval, satisfying the ethical concerns for responsible research with human subjects.

For those 28 students enrolled in the course in which Perusall was offered as extra credit, selected course readings (see Appendix A) were posted as PDF documents on the Perusall platform and synced to the LMS so students could navigate to Perusall from within the LMS as part of the course website. The same readings were also made available in the form of a PDF document on the LMS, and students were given the option of reading them digitally and taking notes or printing them out and annotating them by hand. It is unknown how many students opted to print out the course texts or read them digitally outside of Perusall because tracking functions were unavailable in the LMS at the time of this study. Students were encouraged, but not required, to use Perusall as a social reading and digital annotation program throughout the semester. Those who used it were given a 1% extra credit bump on their final course grade, regardless of the actual score given to them by the Perusall assessment algorithm.

For the 43 students who received 5% course credit based on the Perusall assessment algorithm score, the Perusall platform was also integrated into the existing LMS, and Perusall scores were adaptive and visible to students within the LMS grade center throughout the semester. These students also had the option to print out the course readings if they chose, but their annotation scores were solely based on the Perusall algorithm.

The 54 students in the quasi-control group completed the same course readings, quizzes, and writing assignments, but they were not given access to Perusall, and any annotations of course readings, whether conducted digitally or by hand, were up to the students' discretion and not

assessed by the instructor or calculated into their overall course grades. While these students did not utilize Perusall for text annotations, they were given parallel instruction in close reading techniques and strategies through the identical course reading by Rita Charon called "Close Reading." They were also given an oral lecture by the instructor about the benefits of "annotating thoughtfully" in order to "help you master readings faster, understand the material better, and get more out of your classes." These instructions mirrored those given to the Perusall students, as described below.

Students in the Perusall experimental groups were given the following instructions, adapted directly from instructional resources provided on the Perusall website:

How Perusall Works

Perusall helps you master readings faster, understand the material better, and get more out of your classes. To achieve this goal, you will be collaboratively annotating texts with others in your class. The help you'll get and provide your classmates (even if you don't know anyone personally) will get you past confusions quickly and will make the process more fun. While you read, you'll receive rapid answers to your questions, help others resolve their questions (which also helps you learn), and advise the instructor how to make class time most productive. You can start a new annotation thread in Perusall by highlighting text, asking a question, or posting a comment; you can also add a reply or comment to an existing thread. Each thread is like a chat with one or more members of your class, and it happens in real time. Your goals in annotating each reading assignment are to stimulate discussion by posting good questions or comments and to help others by answering their questions.

Research shows that by annotating thoughtfully, you'll learn more and get better grades, so here's what "annotating thoughtfully" means: Effective annotations deeply engage points in the readings, stimulate discussion, offer informative questions or comments, and help others by addressing their questions or confusions. To help you connect with classmates, you can "mention" a classmate in a comment or question to have them notified by email (they'll also see a notification immediately if online), and you'll also be notified when your classmates respond to your questions.

For each assignment, Perusall will evaluate the annotations you submit on time. Based on the overall body of your annotations, you will receive a score for each assignment as follows:

3 demonstrates exceptionally thoughtful and thorough reading of the entire assignment

2 demonstrates thoughtful and thorough reading of the entire assignment

1 demonstrates superficial reading of the entire assignment OR thoughtful reading of only part of the assignment

0 demonstrates superficial reading of only part of the assignment

When Perusall looks at your annotations, they want them to reflect the effort you put in your study of the text. It is unlikely that that effort will be reflected by just a few thoughtful annotations per assignment. On the other extreme, 30 per assignment is probably too many, unless a number of them are superficial or short comments or questions (which is fine, because it is OK to engage in chat with your peers). Somewhere in between these two extremes is about right, and thoughtful questions or comments that stimulate discussion or thoughtful and helpful answers to other students' questions will earn you a higher score for the assignment. Note, also, that to lay the foundation for understanding other class activities and assignments, you must familiarize yourself with each reading in its entirety. Failing to annotate the entire reading will result in a lower score. (https://perusall.com/)

In addition to the above instructions, students were also given access to example annotations and scores as models prior to completion of their first Perusall reading assignment.

The Perusall AI assessment algorithm assigns a score of zero to three for each student after the student completes the annotation process on an assigned reading. The algorithm takes into consideration the depth of the annotation's content and the distribution of the student's annotations across the text. Superficial annotations clustered on one or two pages will thus result in a lower score, while complex and detailed annotations scattered throughout the reading will result in higher scores. Students may see their scores once the reading deadline has passed and can thus work to improve their scores on subsequent assignments based on the algorithm's feedback and careful observation of and interaction with fellow classmates' annotations. Figures 1, 2, and 3, screenshots of the Perusall "scoring" function, demonstrate the Perusall algorithm in action and the agility of the algorithm, for it is highly customizable based on the needs and expectations of the individual instructor.

Figure 1. Screenshot of Perusall scoring function.

Figure 2. Screenshot of Perusall scoring function (continued).
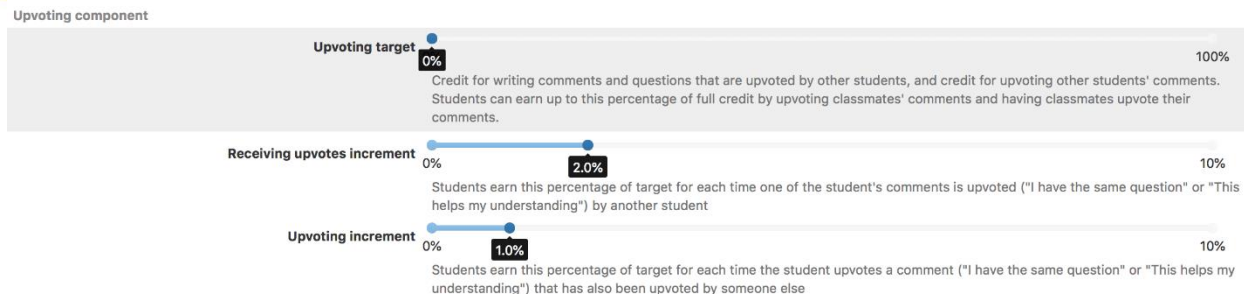
Figure 3. Screenshot of Perusall scoring function (completed).

The Perusall "scoring" function also enables "advanced" features such as one-on-one instructor-to-student direct messaging, emailed deadline reminders, student-to-student chat messaging, and the enabling of direct download of all digital readings to the student's own personal device. Figure 4, a screenshot of the "advanced" scoring window, describes these capabilities and also demonstrates the way they are adaptable, through simple check boxes, to suit each individual Perusall course.

Figure 4. Screenshot of Perusall advanced scoring functions.

In addition to the algorithm's agility, the Perusall "gradebook" function allows instructors additional flexibility, if, for example, an instructor ultimately disagrees with the Perusall algorithm's assessment of a student's annotations. Figures 5 and 6, screenshots of the gradebook in the instructor view, show the icons used to edit any score, as needed.

Click on a score to see details about how it was calculated; click the ✏ icon to manually override Perusall's score for that student. The average score column indicates the average score for the assignments whose final deadlines have passed (those without a ⟳ icon); blanks are counted as zero. The average score row indicates the average score for that assignment; blanks are counted as zero.

Search: _____

Figure 5. Screenshot of Perusall gradebook instructions.

Average score	Frank_Five Dramas of Illness	The Hippocratic Oath	Charon_Close Reading	Charon_Narrative Features of Medicine	Charon_Parallel Chart
2.70	3 ✏	3 ✏	3 ✏	3 ✏	0 ✏
2.70	3 ✏	3 ✏	3 ✏	3 ✏	3 ✏
2.40	3 ✏	3 ✏	3 ✏	3 ✏	3 ✏
1.90	3 ✏	3 ✏	3 ✏	0 ✏	3 ✏
2.50	3 ✏	3 ✏	2 ✏	3 ✏	3 ✏
0.50	0 ✏	0 ✏	3 ✏	2 ✏	✏

Figure 6. Screenshot of Perusall gradebook.

To demonstrate the student view of Perusall, screenshots detailing the documents, assignments, scores, and additional student features, such as notes, hashtags, and chats, have been collected in Appendix C.

Lastly, a reflective narrative survey was distributed through Survey Monkey to the 71 students in the Perusall experimental groups. Students were asked to reflect on Perusall as both a close reading tool and a way of interacting with peers during the reading process. They were asked to consider both the technology itself (form and functionality) and the learning goals associated with the digital annotation process (fostering community and collaboration, ensuring close reading and analysis, providing a digital resource for subsequent course assignments). Students were given one open-ended text box in which to respond.

4.0 Conclusions

With 71 students in the experimental conditions using Perusall and 54 in the control condition without Perusall ($n = 125$), several conclusions can be drawn based on student course scores collected by the LMS. Using a one-way analysis of variance (ANOVA), *p* values were statistically significant in all calculations ($p<0.05$), including those comparing performance differences across the three conditions (Perusall as a grade, Perusall as extra credit, and no use of Perusall) and those addressing correlations between Perusall scores and final course grades. *P* values did not, however, reach the level of highly statistically significant ($p<0.001$); replicating this study with a larger sample size would be one way of addressing this concern. A table of these scores, including minimum, maximum, and mean, can be found in Appendix B.

Median final grades in all the Perusall experimental groups (91.26%) exceeded those of the control group (88.93%), a 2.33% difference that resulted in a letter grade variation between the

groups. Furthermore, a correlation could be found between lower Perusall scores and lower final course grades, and among those students who received extra credit for Perusall participation, all earned a final course grade at or above a B+. Among those students who withdrew from the experimental group courses prior to completion of a full semester of instruction, all had either failed to complete Perusall annotation assignments entirely, thus earning scores of zero from the Perusall algorithm, or else they received extremely low Perusall algorithm scores of 1.5 or less on the reading assignments they completed. All those who withdrew from the course were currently failing the course at the time of withdrawal.

In addition to these findings, further conclusions can be drawn based on scores on student reading quizzes developed by the instructor to measure students' abilities to apply concepts from the Perusall readings to novel intellectual scenarios. In both courses, students were asked to complete comprehensive reading quizzes based on the Perusall readings immediately following the Perusall reading annotation assignments. In Professional Writing in the Medical Humanities (ENG 2130), Quiz 1 covered the first three Perusall readings in a combined cumulative quiz, while those same readings were divided between three separate quizzes (Quiz 1, Quiz 2, Quiz 3) in the Narrative Medicine (ENG 2200) course. In Professional Writing in the Medical Humanities (ENG 2130), Quiz 1 scores increased dramatically, with an average score of 91.24% in the Perusall experimental groups compared to an average of 79.24% in the no Perusall control group. In Narrative Medicine (ENG 2200), quiz scores improved as well, with Quiz 1 scores increasing from an average of 83.70% in the no Perusall control group to 87.34% in the Perusall experimental groups, Quiz 2 scores increasing from an average of 89.26% without Perusall to 92.42% with it, and Quiz 3 scores increasing from 91.16% in the control group to 93.70% in the experimental groups.

Looking more closely at student scores on writing assignments and multimedia projects built on the foundational knowledge gained from the Perusall readings, some additional insights can also be gained. In a written blog post based on a specific Perusall reading by Rita Charon, ENG 2130 students' median scores increased from 92.5% without Perusall in the control group to 97% with it. In Narrative Medicine (ENG 2200), final capstone project scores in the Perusall experimental groups increased from a median of 90% without Perusall to 92.5% with Perusall. In Professional Writing in the Medical Humanities (ENG 2130), the final capstone project median scores jumped from 85% without Perusall to 94% with it. Lastly, in final written reflections in ENG 2200 in which student writers were asked to utilize textual evidence from course readings to support their conclusions about their overall learning outcomes in the course, students in the control group received a median score of 94%, while those in the Perusall experimental groups received a median score of 97%.

While these quantitative measures make a compelling argument in support of Perusall as a viable robo-tool for improving student performance on active and close reading and other reading-based tasks, further qualitative evidence can be drawn from the reflective narrative survey conducted among those students in the experimental groups. These responses demonstrate positive student perceptions of the Perusall platform and improved collaboration among students

when completing reading tasks. While reading is traditionally considered a solitary endeavor, Perusall enables students to turn this isolating learning experience into a community conversation. Figure 7 shows a word cloud based on student responses to the reflective narrative survey.

Figure 7. Word cloud.

Following are some of the highlights from those student narrative responses:

> Perusall was a good platform to use, and I like that it created a community for us to gather ideas. It was also a nice way to collect all of our readings and keep them in one place.

> I learned that within this class I need to read the content more than once in order to be submerged within the text, but to also have a chance to step back and look at the situation as an "expert." One major benefit was the Perusall readings because I was able to critically think on the reading and also look at fellow classmates' thoughts on the reading which could give me a perspective I may have never thought of before. Through these benefits of the class, I feel I have grown in my

ability to read and think more critically, while relating the piece to something of high significance in society.

I enjoyed using Perusall. It caused me to reflect more deeply on what I was writing, especially when I read my peers' comments. It was very interesting to see what other students thought about particularly unique parts of our readings. Being able to read the text online and comment on it from anywhere was the most useful aspect.

I enjoyed using Perusall. I found the technology easy to use, though occasionally trying to click through comments was inconvenient. It did a great job of helping me get my peers' views as well as share my own in a way that didn't eat up time. It was also just a great tool for annotating. Perusall was useful because it worked like a google app in that everyone could contribute and see others' contributions, so it was great for working in a small class.

I did enjoy Perusall. The most useful part was being able to see what everybody else was thinking about the same article. The least useful part was probably how long the articles themselves were.

It was really helpful to see other students' thoughts about the readings (that you wouldn't get to see if you had simply been reading the text alone). It was also nice to make sure that you weren't thinking completely off track and have the opportunity to have questions answered, whether it be by the instructor or other students. Sometimes reading online can be a little tricky because it's easy to lose focus, but keeping the annotations helps with that aspect too.

I think one of my favorite parts about the course was Perusall. It was very interesting to see my classmate's point of view on the text. It also helped me to gain perspective when I did not know what the text meant. Hearing other people's take on the meaning showed a lot about my classmate's personal background.

While data was collected by the author in the form of student responses and student grades, the Perusall platform itself also allows for additional "stealth assessment" (Spector et al., 2016) thanks to the integrated analytics available from the platform's back end. For example, instructors are able to view and download various data points from each Perusall annotation assignment, including heat maps indicating the date and time of student annotation submissions, page view and timing charts for each reading, active and passive reading time measurements for each student, and detailed lists of the most upvoted and unanswered questions posed by the students on a given reading assignment (known as a Perusall "confusion report"). These charts, graphs, and CSV reports are in addition to the actual annotation scores given by the Perusall algorithm (collected into the Perusall "grade center" and synced within LMS grade centers) and the downloadable text of the student annotations themselves. It's important to note that instructors have the capacity to manually adjust the student annotation scores if they disagree

with the Perusall algorithm's assessment. While the author extensively reviewed the Perusall algorithm scores given to her 71 experimental group students, she found no evidence of any inaccurate assessment by the Perusall algorithm and thus never manually altered a single annotation score throughout this research study. Some useful conclusions were drawn from some of the back-end analytics provided by Perusall. For example, the author was able to compare heat maps between ENG 2200 and ENG 2130, two courses running simultaneously in the experimental groups but comprised of somewhat different student populations. While both courses are comprised of High Point University undergraduates and qualify as general education course credit, those students in ENG 2130 are more likely to be studying science and healthcare-related majors and many are taking the course specifically as the second writing course requirement for the pre-pharmacy major. A closer look at the differences between the heat maps of these two courses reveals some interesting time management trends. In a comparison of the day and time of student annotations on the same reading assignment, a theoretical chapter by Rita Charon called "Close Reading," the ENG 2200 students followed the familiar pattern of procrastination, with the hottest point of annotation occurring during the last two hours before the assignment deadline (Figure 8). However, the annotation patterns of the students in ENG 2130 were more evenly distributed over several days prior to the deadline, and the hottest points occurred during the traditional 9-5 workday hours on Friday, a full 48 hours prior to the Sunday, 11:59PM deadline (Figure 9).
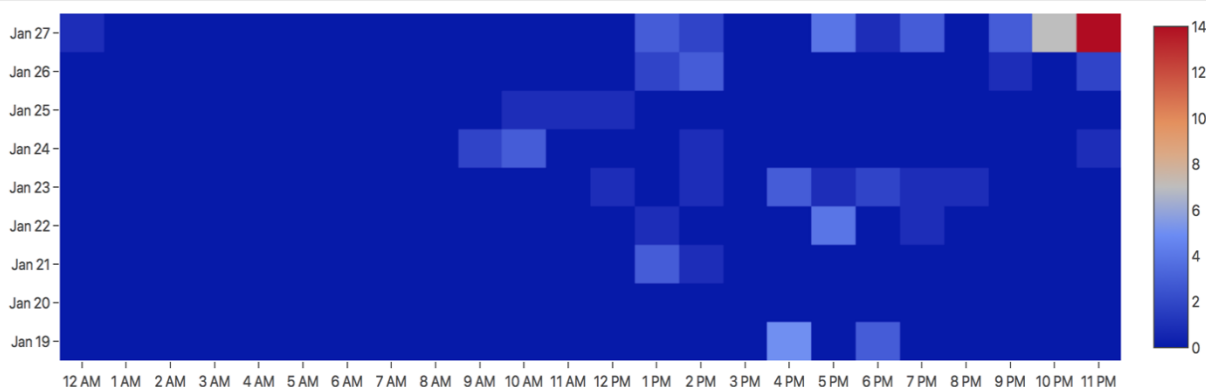

Figure 8. Submission time heat map 1.
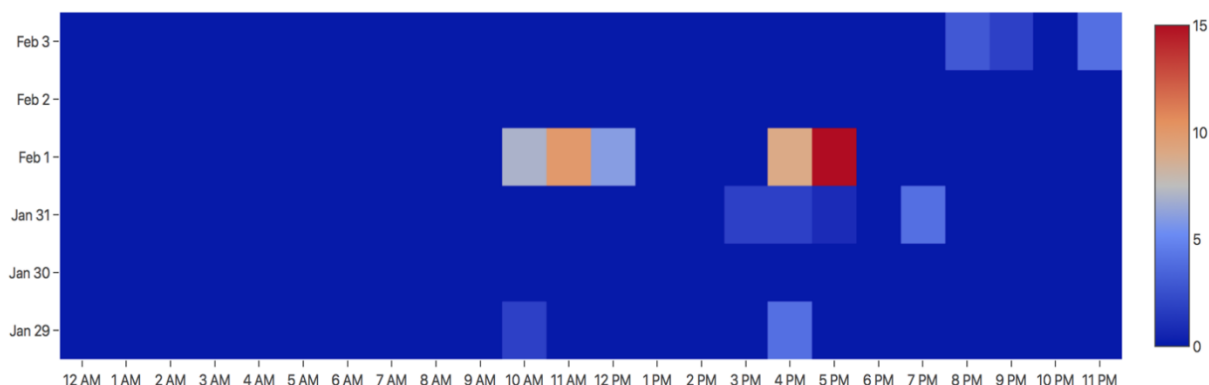
Analytics for Charon_Close Reading

Figure 9. Submission time heat map 2.

Another interesting point of formative "stealth assessment" (Spector et al., 2016) between these two classes can be found in the Perusall page view and reading time charts, available for viewing and download from the back-end platform analytics. These charts combine the number of views per page of a given reading assignment and the amount of time students spend reading each page in order to map student reading progress and highlight pages of the reading where students seemed to spend the most time, perhaps due to reading comprehension difficulties. After viewing these charts, an instructor may be able to delve into those specific pages of the reading that proved most time consuming, cross reference those pages with the annotations and/or questions students posted on them, and thus deduce important reading and comprehension trends to be addressed in future classes for further discussion, clarification, and analysis. As Figure 10 notes, students in ENG 2200 spent increased time on pages 3, 8, and 21 of the Charon chapter, while page 20 had the most page views overall. Time averages for these pages reached nearly 50 minutes each, while the averages for the last two pages of the reading, those in which Charon cites her sources, were less than five minutes.
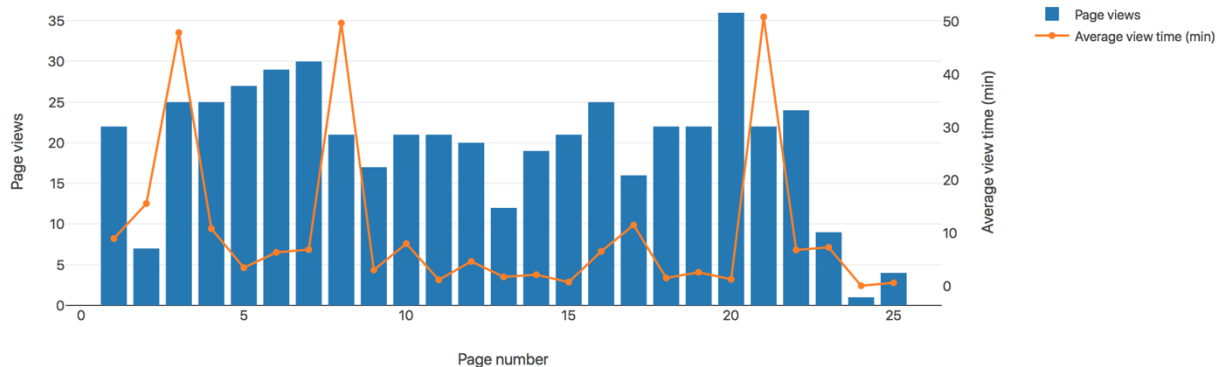
Analytics for Charon_Close Reading

Figure 10. Graph of student page views & view times 1.

In contrast, Figure 11 details the reading patterns of ENG 2130. These students' reading time peaked on page five, while page views peaked on page three. The rest of the reading was much more evenly distributed than their ENG 2200 counterparts, though the patterns mirror each other on the final two pages of citations. The ENG 2130 students spent far less time, on average, reading the chapter, with the peak being 30 minutes instead of 50, though they had many more page views (several pages in excess of 100 as opposed to a maximum of 35 in ENG 2200). This could indicate that ENG 2130 students chose to read the text a few pages at a time, or it could indicate a lack of focus. It could also indicate technical difficulties accessing the reading. These are the kinds of questions instructors could ask in response to the Perusall analytics, thus deepening their understanding of their students' reading patterns and elucidating any deeper questions about the reading or troubleshooting technical difficulties.
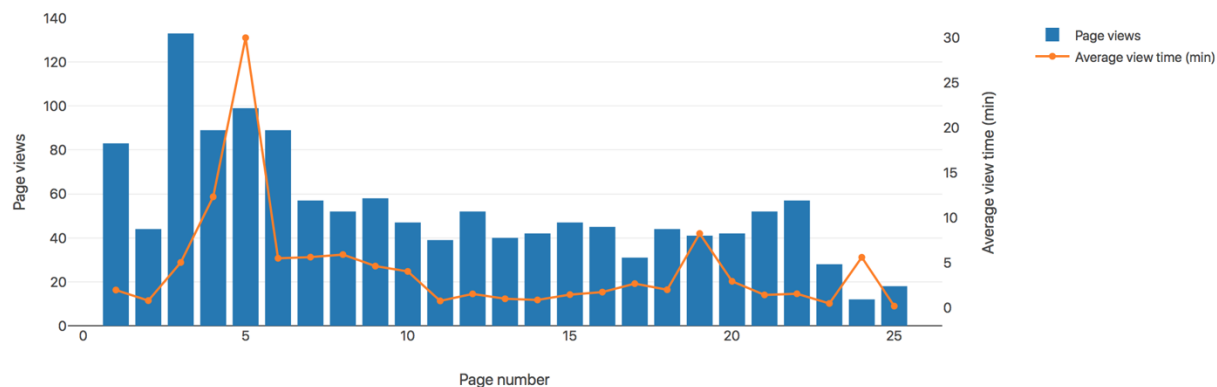

Analytics for Charon_Close Reading

Figure 11. Graph of student page views & view times 2.

Another useful analytical tool available in Perusall is the active and passive reading time measurement for each individual student. This allows an instructor to potentially assess student learning difficulties and intervene appropriately. Figure 12 provides a sample screenshot of this data set for the same Charon chapter described in the earlier figures, indicating an extremely wide range of reading times among the eight ENG 2200 students represented. While two of these students spent in excess of three hours on this reading, one failed to read it at all. (Circumstantially, the student who failed to complete this reading assignment was also the student with the lowest final grade in the class. This implies that further research could examine the potential uses of Perusall to aid student retention and instructor intervention.)

Analytics for Charon_Close Reading

	Last name	First name	Viewing time (min)	Active reading time (min)	Activity
DA			3 hours, 18 minutes	1 hour, 17 minutes (39%)	4 annotations
TB			0 seconds	0 seconds	
TB			25 minutes, 8 seconds	22 minutes, 59 seconds (91%)	5 annotations
MC			2 hours, 32 minutes	47 minutes, 45 seconds (31%)	5 annotations
ZC			1 hour, 51 minutes	1 hour, 7 minutes (60%)	4 annotations
SC			0 seconds	0 seconds	
SC			13 minutes, 5 seconds	12 minutes (92%)	4 annotations
RD			3 hours, 13 minutes	1 hour, 47 minutes (55%)	4 annotations

Figure 12. Screenshot of student active reading time.

In Figure 13, we see a sample excerpt from the Perusall platform's unique analytics known as a "confusion report." The report collects the top questions posed by students in their annotations, based on the annotations most highly upvoted by the student users themselves and unanswered by other students in their responses. It should be noted that the platform allows for immediate email and text notification to individual users when their posted questions get answered by another student or the instructor. This means the confusion report is a reliable means of gauging gaps in student comprehension, and it offers rich fodder for future class discussions about the reading. It should also be noted that the platform itself includes adaptive responses based on the instructor's perception of the "usefulness" of the confusion report. Should an instructor click a frowny face icon, the platform will adjust subsequent confusion report calculations based on that feedback.

Figure 13. Screenshot of Perusall's "confusion report."

To put the conclusions of this study in the broader context of writing analytics and the use of artificial intelligence in the writing classroom, the author wishes to reaffirm the tremendous benefits to both student learning and instructor labor gleaned from the adoption of the Perusall social reading and digital annotation platform in her classes. The platform's autonomous assessment of student annotations proved extremely accurate, and it seamlessly integrated into the existing LMS and grade center without any major technical difficulties. The instructor found she could rely on the Perusall platform to ensure her students completed course readings at a consistently high level of close reading and active engagement with the text and each other. The confusion report provided a useful starting point for class discussions and the other "stealth assessment" (Spector et al., 2016) analytics gave her insight into the reading, writing, and critical thinking habits of individual students so she could better tailor her feedback during face-to-face meetings, individual office hours, and tutorial sessions. Overall, student grades on assignments associated with the Perusall readings trended upward in the course sections in which Perusall was utilized, and the instructor found student writing more often included textual evidence from the Perusall readings than in previous courses in which those same readings were assigned as PDF files to be read and annotated in solitary fashion. The students themselves articulated positive perceptions of Perusall as a user-friendly digital tool, as a platform for the curation of their own learning, and, most importantly, as a means of connecting with other student readers and their diverse perspectives, questions, and interpretations. While Perusall is merely one among many digital robo-tools available to writing instructors, the author believes it to be a prescient example of what lies ahead in the digital classrooms of this century. Before students are able to write effectively and engage with their instructors as "apprentice scholars," (Sommers, 2006) they must first master the skills of close reading and analysis. The Perusall platform builds that necessary foundation, and it does so in ways that support active reading habits in Vygotsky's "zone of proximal development" (Kim & Baylor, 2006) by utilizing the positive influences of collaborative learning. The Perusall platform capitalizes on what Bandura calls proxy agency,

thus allowing students to utilize the expertise of others in order to accomplish a reading task while also harnessing the social and collaborative functions of learner agency in the presence of other students (Kim & Baylor, 2006, p. 579). The result is a form of collective agency over the reading process that motivates students and thereby enhances the "social richness" of the course learning environment while collectively aiding them in reaching their learning goals (Kim & Baylor, 2006, p. 580).

5.0 Directions for Further Research

Some considerations for further research include replication of this study across different student populations and contexts. For example, it is unclear whether or not Perusall would be as effective among less privileged students than those at High Point University or among students with varied levels of digital literacy. It is also unclear if Perusall would be effective among nontraditional students such as adult education and returning student populations who may find the digital annotation of texts counterintuitive and, instead, prefer a more traditional method of annotating print texts by hand. Consideration should also be given to the possibility of the Novelty Effect within this study, as this was the first exposure to Perusall, and it is unclear whether or not the students' enthusiasm for this technology might wear off with repeated use.

Based on the myriad findings from Perusall, ITS, EDM, writing analytics, NLP, composition pedagogy, and effective human tutoring pedagogy from both face-to-face and online educational settings discussed in this research note, the following speculative AI robo-tutor scenario has been developed as a prediction of our near future. In this scenario, Emma is a struggling first-year writer, and she has been given an argumentative writing assignment in which she should utilize both popular and scholarly sources to make a claim about the use of social media as an educational tool in the classroom. She will write three drafts of the essay, one of which will be formative and include feedback from her robo-tutors, and one which will be submitted to her instructor for face-to-face formative feedback before a summative revision and submission of a final draft for a grade. Emma may use the robo-tutors at any stage of the process, and she may use them as often as she likes for as long as she likes. The program may be used in all her classes throughout her four years of study, and because the program operates using machine learning and artificial intelligence algorithms that allow the robo-tutor program to execute agile programming functions adapted to her unique writing style, Emma should be able to transfer her reading and writing skills from one discipline and classroom-specific writing context to the next.

The program is designed to scaffold that learning experience and reward Emma with praise as she gains self-efficacy in her reading and writing while also creating a perceived social learning environment in order to optimize the intellectual effects of engaging in thoughtful dialogue with others through collaborative social reading and annotation such as that demonstrated by Perusall, and, ultimately, collaborating "across the drafts" (Sommers, 2006) to achieve sophisticated academic writing. Ideally, each student using the program would complete a brief profile before the first tutorial session so that the program could choose AI avatars appropriately matched to the user's profile.

Emma's robo-tutors and human collaborators are:

- Matt, the MLA and grammar expert (AI)
- Amy, the peer collaborator (human being)
- Brad, the tutor who provides motivation and scaffolded writing instruction (AI)
- Kathy, the librarian and research assistant (AI)
- Scott, the troublemaker who pushes Emma toward useful cognitive conflict by offering different perspectives and "risky" revision suggestions (AI)
- The Instructor (human being)

Emma may click on any individual avatar at any point to ask for help. Depending on what stage in the writing process Emma is in, the program may also offer unsolicited advice/active questioning when certain keystroke patterns or delays in activity prompt a feedback loop. Emma may turn off this function if she chooses. The unsolicited advice and active questioning will decrease as Emma masters certain writing skills, such as citation or grammar, but the higher-order writing functions, such as those prompted by avatars Amy and Scott, will continue to challenge Emma to remain in Vygotsky's "Zone of Proximal Development" (ZPD) throughout the composition process. When keystrokes are active, the program will not interrupt Emma's composing process, unless she explicitly requests aid by clicking for it.

When Emma engages a robo-tutor, any dialogue is automatically saved to the dialogue history associated with the document in progress. Each time the robo-tutor asks Emma a question, she has options to choose "yes," "no," "maybe," or "I don't know," as well as a dialog box for narrative response or additional attachments. Depending on her choices, the robo-tutor program will branch off its feedback and tailor the response to Emma's writing style and the specific needs of the task at hand.

In addition to clicking on any avatar for guidance, Emma may also choose from a select number of explicit commands in the toolbar, such as emailing the instructor. Each toolbar icon elicits guidance from an avatar connected with that task and may also connect Emma to other resources, such as the university library's database or the digital annotations of a course reading posted by her peers.

> *Emma logs onto the robo-tutor platform. It contains a toolbar, icons for each member of her robo-tutor team, internet connectivity, and a word processing window queued up and ready to go. Emma may choose to compose her first draft without interruption. Instead, she immediately asks for help. Robo-tutor Brad pops up.*
>
> Brad: Hi Emma! I'm Brad, your personal robo-tutor. I look forward to working with you! Before you start composing, would you mind sharing an example of your previous work with me? Some piece of writing you're proud of.
>
> (*Emma shares a research paper she wrote in high school that received an A.*)

Brad: Thanks! I can tell you're good at organizing your ideas into effective paragraphs. There are also some things we can work on together. Let's get started! Will you share the assignment prompt with me?

(*Emma shares the assignment prompt.*)

Brad: Great! Social media in the classroom? Interesting topic. Have you done any prewriting or brainstorming about this topic?

Emma: Yes.

(*Emma shares an introductory paragraph and bulleted list of evidence that shows some degree of logical development in favor of social media as a potential educational tool in high school settings.*)

Brad: I see you're working on an introduction for an argumentative paper in favor of social media use for educational purposes. Who is your intended audience for this essay?

(*At this point Brad continues to ask leading questions that prompt Emma to think critically about who her audience will be and what she intends to accomplish in the essay. After Emma has written an introduction and begins working on her first body paragraph, robo-tutor Matt interrupts Emma's drafting when she has paused for a significant amount of time at a place where an in-text citation is required.*)

Matt: Hi! I'm Matt, a citation and grammar expert. It looks like you could use some help citing your source here. What style would you like to use?

(*Emma clicks MLA 8 from the dropdown menu.*)

Matt: This looks like a block quote. Do you know what that is? Click on any word or phrase in our dialogue and you can see its definition.

(*Emma clicks on "block quote."*)

Matt: Would you like me to show you how to format a block quote following MLA 8 guidelines?

(*Emma clicks "yes" and watches the screen as the cursor executes the correct formatting. Matt disappears. Emma continues composing. The next time Emma cites a source, she does it correctly without using Matt's assistance. Matt pops up briefly to provide positive feedback in the form of a smiley face emoji and thumbs up. Emma reaches the conclusion of her essay and immediately clicks the command to "revise."*)

Kathy: Hi Emma! I'm robo-tutor Kathy, a librarian and research expert. I see you've finished your first draft. Ready to make it even better?

(*Emma clicks "yes."*)

Kathy: You told us your intended audience would likely be digital natives who are comfortable reading news sites on their smartphones. Have you considered ways to make your argument more visually appealing for that digital audience?

(*Emma doesn't respond.*)

Kathy: For example, here's a recent article on a similar topic.

(*The article floats onto the screen as Emma's draft recedes behind it.*)

Kathy: Notice how it starts with an image and includes an embedded video in the middle? That makes the reading more interesting for digital natives. Would you like me to show you how to find an image you can use in your essay for free?

(*Emma clicks "yes" and then watches Kathy execute a screencast demonstrating how to use the Google tool that filters for usage rights labeled for noncommercial reuse. At any point, Emma can stop the tutorial video and return to her own document or mirror Kathy's movements by conducting her own search.*)

Kathy: I'm going to save a list of other resources for you under the "Research" feature for future use. Click on me any time and I'll walk you through the process until you get the hang of it yourself.

(*As Kathy disappears, Scott pops up with a devious grin. He's the troublemaker robo-tutor designed to challenge Emma's perspective and entertain "risky" revisions.*)

Scott: Are you interested in some more revisions?

(*Emma clicks "yes."*)

Scott: I noticed the assignment prompt asked you to evaluate or critique a social media platform in order to engage with the other side of your argument. According to my calculations, fewer than 5% of verbs and phrases in this draft are associated with evaluation and critique, and 75% of your verbs and phrases are associated with summary and exposition. Would you like me to highlight some examples?

(*Emma sighs and reluctantly clicks "yes."*)

Scott: Here is a sentence that seems to be working toward evaluation. Click on the words or phrases you think are associated with evaluation.

(*Emma clicks on "however" and "despite."*)

Scott: Good! Now, is this an idea you could possibly expand?

(*Emma clicks "yes."*)

Scott: If so, let's move this sentence to a blank page and give you some time to write about it.

(*Emma watches as Scott opens a blank page and posts her sentence at the top. All other commands and tool bars disappear. At the end of five minutes, Emma has written a rough paragraph about potential risks of Twitter in the classroom, including the way it limits users to narrow character counts and seems to promote bullying behaviors. Emma knows, however, that these ideas could undermine her argument in favor of Twitter use in the classroom. Scott pops up when her composing keystrokes have stalled.*)

Scott: Are you finished, or do you want to keep going? *Emma clicks "finished."* Was this freewriting helpful?

(*Emma clicks "I don't know."*)

 Scott: Do you think some of those ideas might be useful in your draft?

(*Emma clicks "maybe."*)

 Scott: I see your new paragraph contains more evaluative words and phrases now. Well done! Would you like to continue looking at other places in your draft that could use more exploration?

(*Emma clicks "no."*)

Scott: Maybe it would be a good time to schedule a face-to-face meeting with your instructor or a writing tutor to explore these ideas further.

(*Scott offers a link to the tutoring services appointment system, but Emma does not click it. Emma scrolls to a different part of her draft, adds a sentence and then deletes it. She then adds a quote from a credible popular source, but the quote doesn't connect to the rest of the paragraph. Emma re-reads the paragraph then deletes the source completely, rather than working to connect it to her claim. Robo-tutor Brad pops up.*)

Brad: It looks like you're feeling stuck or frustrated. Maybe I can help.

(*Emma types into the narrative box:* I think I need more evidence.)

Brad: It looks like you've used numeric data and personal experience to support your claim already. What about an expert opinion to back it up?

 (*Emma clicks "yes."*)

Brad: Would you like to use a popular or scholarly source?

(*Emma clicks on the definitions for these terms, and then selects "scholarly" from the menu.*)

Brad: Okay. Do you have one in mind?

(*Emma clicks "yes" and shares the popular source she had previously quoted and then deleted earlier.*)

Brad: It looks like this is a popular source, not a scholarly one. This source may still provide an expert opinion to support your claim, but I want to be sure you understand the difference between popular and scholarly sources first. Okay?

Emma: Okay.

Brad: Re-read the definitions for popular and scholarly sources and then look at these examples.

(*Brad floats two example texts onto the screen. As Emma scrolls over each one, dialog boxes pop up on the documents to highlight key elements that distinguish them as popular or scholarly, such as a Works Cited page for the scholarly source and a current event photograph as the lead-in for the popular source.*)

Brad: Now look at the source you chose. Can you tell me what clues you see in that source that indicate it is popular instead of scholarly?

Emma: It has pictures and a video.

Brad: Good! I can see how you might have been confused by the psychologist the author quotes as an expert within the article. She is a scholar, but that doesn't make the source itself a scholarly one. One thing you could do is trace that psychologist back to her own scholarly publications on this subject using the library research databases. You can search on her name first, and then narrow your search with the key term "cell phone addiction" until you find something that suits your needs. Would you like to try that now?

Emma: Yes.

(*Kathy pops up and assists Emma in her search. Brad returns after Emma finds a scholarly source from the psychologist quoted in the* New York Times *article she used earlier.*)

Brad: You'll need time to read that article, so this might be a good place to pause in your revisions and return to this paragraph tomorrow. Before we do that, though, I have a question I'd like you to consider about this paragraph. Are you willing to continue?

(*Emma clicks "yes."*)

Brad: As a reader, I'm feeling confused here.

(*Brad points to several sentences describing how the Twitter platform works, who created it, and how many Tweets are posted daily. The sources for these facts are all correctly cited.*)

Brad: You told me your target audience already knows about social media and uses at least one social media site daily. So why are you giving readers so much background about Twitter when they already know how it works and might even be regular users? How does this information connect to your argument?

(*Emma re-reads the paragraph and deletes more than half of it, all of which included unnecessary background information. This leaves ample space in her essay to explore her new scholarly source and the ideas she began to uncover in her freewriting. Before Emma logs out, student peer Amy pops up. Depending on programmatic constraints, she could be an actual student in Emma's class or a robo-tutor avatar designed to model Emma's cognitive development and offer Emma an opportunity for collaboration and increased self-efficacy.*)

Amy: Hey! Are you working on our paper for English 1103?

Emma: Yes.

Amy: Me too. I'm struggling with my thesis statement. Would you mind reading it and telling me what you think?

Emma: Sure.

(*Amy shares a thesis statement with Emma that argues against the use of social media for educational purposes. The thesis statement is riddled with surface errors, but otherwise it is sufficient for the purposes of the assignment. Emma edits it and sends it back.*)

Amy: Thanks for the grammar help. I suck at that stuff. What stance are you taking for this paper?

(*Emma and Amy chat about their different perspectives, and Amy asks Emma questions, offering counterevidence that she's uncovered from the other side. The dialogue prompts Emma to think more deeply about her position but leaves Emma dissatisfied. She wanted to finish her draft in one day.*)

Amy: Hey, what's your favorite social media platform anyway? I'm totally addicted to SnapChat!

Emma: Twitter.

Amy: (Cool emoji) There's lots of political stuff on there. I wonder if our teacher has a Twitter? You should follow her…or troll her if you get a bad grade (winky smiley face emoji)

Emma: LMAO

Amy: TTYL

Since the robo-tutor platform allows for automatic saving and back up functions through cloud storage, every keystroke, every draft, every addition and deletion, every robo-tutor conversation and the time spent in each application will provide real human tutors and instructors insight into the student user by providing them access to a "stealth assessment" of Emma's ongoing composition metadata. An instructor could interpret and apply that data in face-to-face or online conversations with Emma, and in so doing, focus her attention on higher-order ideas, nuanced perspectives, and abstractions that will improve Emma's essay and give her meaningful practice in academic writing, ultimately providing writing experiences that will easily transfer to other academic disciplines and professional contexts. Following are a few examples of ways an instructor could use the data generated by the robo-tutor platform to foster a positive instructor-student working relationship as described by Sommers in "Across the Drafts." For example, an instructor might point to sentences that were headed in a more conceptually complicated direction and ask Emma about them.

> Instructor: Emma, you were headed in an interesting direction here. Why did you delete it? What if you returned to that idea now that you've read that additional scholarly source from the psychologist? Is there more there to explore in light of her findings?

The instructor might look at the time Emma spent on certain parts of her essay and compare it to others.

> Instructor: It looks like you spent twice as much time on this paragraph as you did on this one, yet that second paragraph has ideas in it that interest me and leave me wanting to know more. When I read these two paragraphs together, the first one seems to be more fully developed, but also more expected, and the second reads as somewhat rushed and incomplete, which makes sense if you spent less time on it, but it's also got some untapped potential in terms of getting your audience to see social media differently. In the next draft, maybe you could give this paragraph more time and see where it takes you. What do you think?

The instructor might look at revision suggestions Emma didn't take from the robo-tutors and challenge her to take more intellectual risks in the next draft.

> Instructor: I see Scott did some risky editing for you and deleted this whole source from your paper. I know the prompt requires five sources, and without that one you only had four, but I think I see what Scott was getting at when he removed it. Even in the draft without it, your argument still made sense. That means you weren't using that source to do any intellectual work for you. It wasn't, or isn't *yet*, integrated into the logical structure of your argument. Instead of just adding a source to meet the requirements, I'd rather see a draft that is a little bit short on sources but long on synthesis of the sources you've already got. I like that you

found the actual psychologist from that *New York Times* piece instead of just quoting it secondhand. In the next draft, work on connecting those sources more deeply to your argument. Imagine you're having a conversation with those authors about social media in the classroom. Debate with them in your head and see if you are able to uncover any surprises, or complications, in your thinking. I think at least one of those authors might disagree with your stance on live Twitter streams during class meetings, but there's certainly one of them who would likely agree. Can you figure out which one?

The instructor might see that Emma has clicked repeatedly on the rubric associated with this writing assignment, particularly on the criteria related to source use.

Instructor: You're right. Technically, you might lose a point or two on the rubric if you have four instead of five sources. You might gain a point or two if those four are integrated effectively, but if you really want to use that fifth source, and you're willing to do the work to integrate it with the others, you might bring it fully into the conversation you're having here with that psychologist. It could be an interesting counterpoint to her argument about digital natives and technology addiction. And don't be afraid to step into that conversation yourself! You are, after all, an "apprentice scholar," a digital native, and a Twitter user yourself. What do you think?

Note

The author has obtained written permission of the students to use their work in this study.

Author Biography

A graduate of the University of Alaska Anchorage, **Allison S. Walker** received her M.F. A. in Creative Writing in 2004. Her poetry has appeared in numerous literary journals and her recent scholarly work in *EvoS: The Journal of the Evolutionary Studies Consortium*; *Visual Imagery, Metadata, and Multimodal Literacies Across the Curriculum*; the *Community Literacy Journal*; and *The Journal of Writing Analytics*. Her research interests include narrative medicine and empathy studies. She also directs *HPU LifeLines*, a service learning initiative that harnesses the healing power of poetry by connecting students with residents of local assisted living facilities and after-school programs. In her spare time, Allison likes to volunteer for the Feral Cat Assistance Program of Guilford County.

References

About Eli Review. (n.d.). Retrieved from https://elireview.com/about/

Gilbert, A. H. (1922). What shall we do with freshman themes? *The English Journal*, *11*(7), 392–403.

Kim, Y. (2013). Digital peers to help children's text comprehension and perceptions. *Journal of Educational Technology & Society*, *16*(4), 59–70.

Kim, Y., & Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, *54*(6), 569–596.

McKenzie, L. (2017, November 14). Do professors need automated help grading online comments? *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/

Merrill, D. C. et al. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences*, *2*(3), 277–305.

Origin Stories showcases M-Write. (n.d.). Retrieved from https://ai.umich.edu/blog/origin-stories-showcases-m-write/

Perusall: Every student prepared for every class. (n.d.). Retrieved from https://perusall.com

Sommers, N. (2006). Across the drafts. *College Composition and Communication*, *58*(2), 248–257.

Spector, J. M. et al. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology & Society*, *19*(3), 58–71.

Walker, A. (2017). I hear what you're saying: The power of screencasts in peer-to-peer review. *Journal of Writing Analytics, 1*(1), 356-391.

WriteLab and Chegg: Improve your writing in less time. (n.d.). Retrieved from http://www.easybib.com/guides/writelab-chegg/

Appendix A: Citations for the Selected Readings Used in This Study ("Perusall Readings")

Charon, R. (2006). Close reading. In *Narrative medicine: Honoring the stories of illness (pp. 107–129)*. New York: Oxford University Press.

Charon, R. (2006). Narrative features of medicine. In *Narrative medicine: Honoring the stories of illness* (pp. 39–62). New York: Oxford University Press.

Charon, R. (2006). The parallel chart. In *Narrative medicine: Honoring the stories of illness* (pp. 155–175). New York: Oxford University Press.

Frank, A. (2007). Five dramas of illness. *Perspectives in Biology and Medicine*, *50*(3), 379–394.

Fusselman, A. (2001). Afterword. In *The Pharmacist's Mate* (pp. 87–102). New York: Penguin Books.

Tyson, P. (2001). The Hippocratic oath today. *NOVA Online*. Retrieved from https://www.pbs.org/wgbh/nova/article/hippocratic-oath-today/

Appendix B: Data Tables

Final Grades	Minimum	Maximum	Mean
Conditions 1 & 2: Perusall as a grade; Perusall as extra credit	67.63834	97.46901	91.26

Final Grades	Minimum	Maximum	Mean
Condition 3: No Perusall	63.93361	90.97995	88.93

Quiz 1 Grades	Minimum	Maximum	Mean
Conditions 1 & 2: Perusall as a grade; Perusall as extra credit	79.5796	96.98219	91.24
Condition 3: No Perusall	48.88025	91.24719	79.24

	Quiz 1 Mean	Quiz 2 Mean	Quiz 3 Mean	Capstone Project Mean	Final Reflection Mean
ENG 2200: Perusall as a grade; Perusall as extra credit	87.34	92.42	93.70	92.50	94.00
ENG 2200: No Perusall	83.70	89.26	91.16	90.00	97.00

	Capstone Project Mean	Blog Post Mean
ENG 2130: Perusall as a grade; Perusall as extra credit	94.00	97.00
ENG 2130: No Perusall	85.00	92.50

Appendix C: Screenshots of Perusall Student View
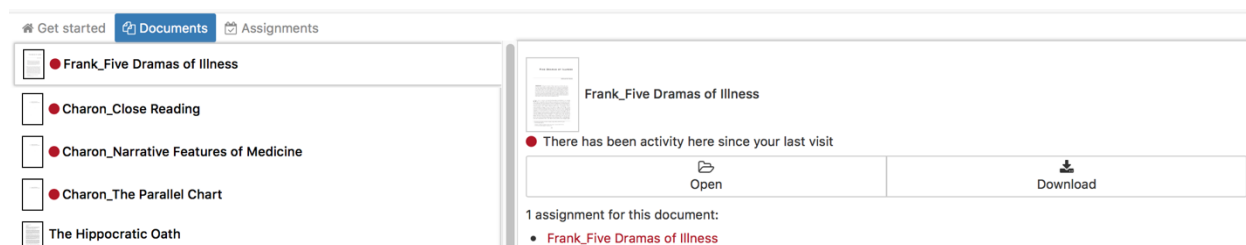
Figure C1. Perusall student view of assigned documents available for annotation.

Figure C2. Perusall student view of assignments and deadlines.

Figure C3. Perusall student view of annotation scores.

How to add Perusall deadlines to the calendar app on your computer or mobile device

Perusall provides an assignment calendar that you can subscribe to from the calendar on your computer or phone; depending on your calendar app's settings and capabilities, the calendar can also update automatically if new assignments are added or if due dates change.

To subscribe to your personal assignment calendar, enter the URL below into your computer or phone's calendar app:

https://app.perusall.com/courses/2019fa0-l-narrative-medicine-eng-2200-01/SqBbMMyHfoAyWnYBB/due_dates.ics

For more information on how to set up your calendar app:

- macOS calendar app
- Microsoft Outlook
- iPhone or iPad

Figure C4. Perusall student view of personal assignment calendar feature.
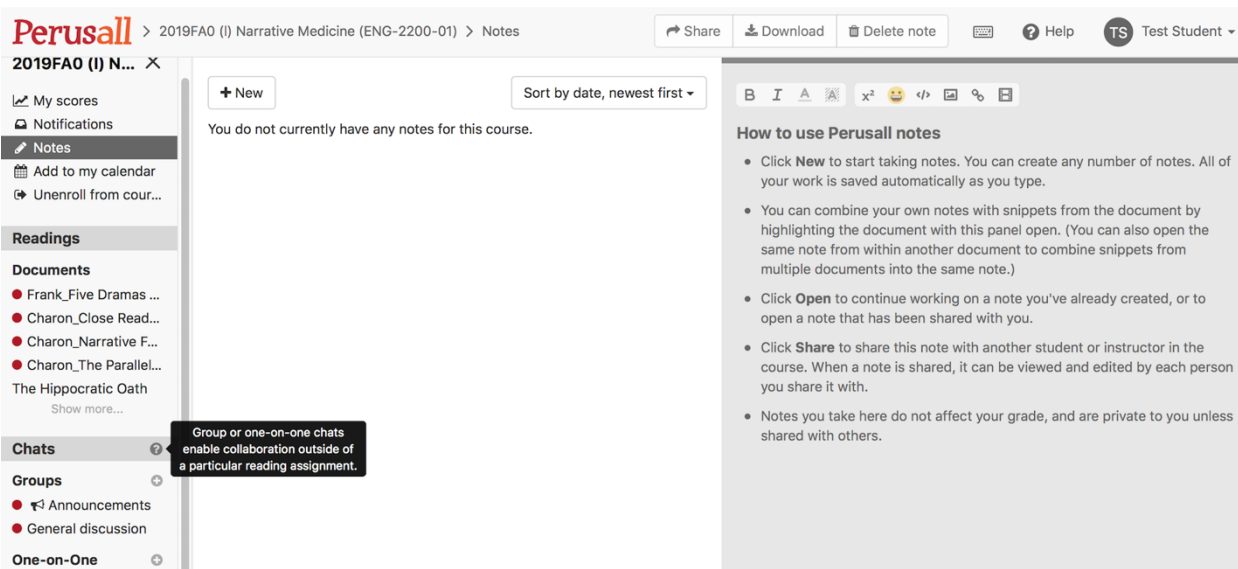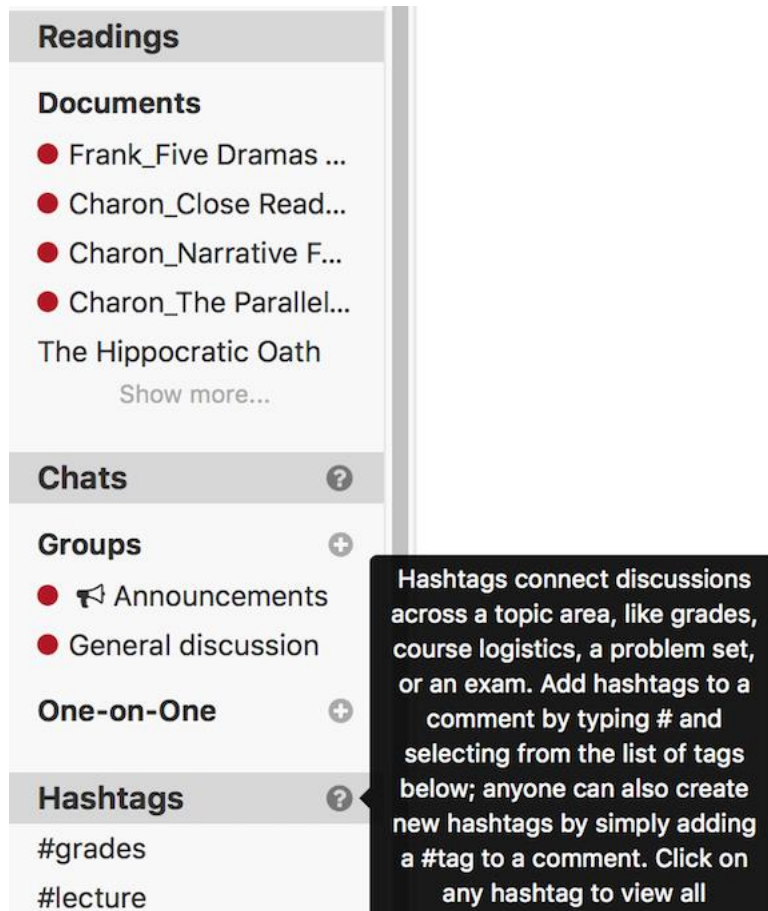

Figure C5. Perusall student view of notes and chat features.

Figure C6. Perusall student view of hashtag feature.