

Improving Critical Reading with E-Texts: A Controlled Study in a Collegiate Philosophy Course

Mark N. Jensen

U. S. Air Force Academy

mark.jensen@usafa.edu

Lauren F. V. Scharff

U. S. Air Force Academy

Abstract: This study investigated the impact of incorporating e-reader texts and annotation tools in multiple sections of an upper level philosophy course ($N_{Control} = 98$; $N_{E-reader} = 76$). This study adds to the body of literature that assesses gains/losses in conventional measures of performance (e.g., scores on graded assignments) and changes in student attitudes as reported in questionnaires. However, this study was unique in that it focused on training students to use e-reader tools for critical reading practices and it included assessment of student annotations and their relationship with the performance measures. Using both quantitative and qualitative data, we tested the hypothesis that, with intentional training and a course-design that provided multiple opportunities for practice and feedback, students using e-readers for critical engagement with their reading assignments would demonstrate (a) deeper understanding of the content of the texts, (b) improvement in their use of critical reading practices, and (c) improvement in their attitudes toward the use of e-readers for academic work. While we did not observe significant gains in graded assignments compared with control groups using printed texts, we found no evidence of losses for students using e-readers. At the same time, we found evidence of improvement in students' critical reading practices, especially when paired with modeling and practice throughout the term. We also observed significant positive changes in student attitudes toward the use of e-readers for academic work, compared with controls. Our findings suggest that achieving the benefits of e-readers for the development of critical reading skills requires a course with design elements that are specifically tailored to this purpose.

Keywords: critical reading, e-text, e-book, e-reader, annotation, course design

Critical reading, or what some scholars have called “deep reading” (Wolf & Barzillai, 2009), is vital to contemporary democratic citizenship, insofar as a flourishing liberal democracy depends on informed, engaged, and well-read citizens who are the key decision-makers in the democratic political system (Dahl, 2000). The skills associated with critical reading should therefore be central elements of a liberal education that aims to prepare students for responsible democratic citizenship (Nussbaum, 2006).

Critical reading can be distinguished from ordinary reading by the degree to which readers engage critically with the text. Where ordinary readers aim to understand the central message, thesis or narrative of a text, critical readers go further. We offered an account in 2014:

[Critical readers] pay attention to the genre of the text and what might be known about the author's context. They aim to make sense of the author's support, defense, and development of the central message. They also challenge the text, raising questions and objections not only about the truth of the central message, but also about the author's argument in its defense. In the end, they see each text as part of a conversation in which their own

reflections become new contributions to advance our collective understanding of the issues in question. (Jensen & Scharff, 2014, p. 83)

Instructors and students face challenges when it comes to teaching and developing critical reading skills. For example, it takes more time and effort to critically read a text (Wolf, 2009); time and effort that students may not be prepared to invest, especially when studies suggest that over 70% of students have not completed the assigned reading for a given class day (Hobson, 2004). It also takes money and planning to critically read: students must have a version of the text with space to mark it up, together with the appropriate resources necessary for investigation, note-taking, and so on. The rising prices of academic books have increased the incentive for students to borrow, rent, or sell them back, which correspondingly reduces the chances that they will mark them up (Scharff & Dull, 2011). Finally, it is important to note that, especially for college educators, students may have already developed reading habits that are inconsistent with good critical reading (ACT, 2007). Changing students' reading habits may be difficult, especially if the initiative to improve students' critical reading skills is concentrated in just a single general education class, e.g., a first-year literature course.

Recently, some scholars have become concerned that the development of critical reading skills is being further undercut by the cultural shift from reading on paper to reading on a screen. A great deal of research has been undertaken to study these changes (see the literature review by Singer and Alexander, 2017). Neuroscientist Maryanne Wolf has discovered that our brains work differently when we read a paper text versus an e-text (Wolf, 2010). Her results indicated that when we read critically in the paper environment, we have conditioned ourselves to read deeply, bringing the whole of our cognitive faculties to the text. But in the electronic environment, we tend to "short-circuit" this process: we skip around the page and do less analytical work. Comparative reading comprehension studies appear to bear this out. Daniel and Woody (2013) found that students' reading comprehension in the electronic environment was similar to reading comprehension in a paper environment, but that it took students longer to read electronic texts. Further, Singer and Alexander (2016) found that while students predicted that they would do better on follow-up tests when working from an e-text, this was not actually the case.

More generally, many readers, especially academic or critical readers, seem to think that their critical reading faculties work best with a book that has a physical location with fixed physical pages (Baron, 2015). Electronic books, on the other hand, seem to them to be too detached or distant (Dirda, 2015), and screens are cluttered with other distracting elements (Klinkenborg, 2010; Daniel & Woody, 2013). Moreover, research seems to indicate that the physical chunking of a text into pages is important for reading comprehension (Tanner, 2014). To be sure: many of these commentators support the use of screen-reading for many uncritical tasks, such as newsgathering, shopping, and social media (Foasberg, 2014). But for critical and academic tasks, they say, paper is still best (Jabr, 2013).

We take a more optimistic view on the possibilities for cultivating critical reading skills in an academic e-reading environment. In a previous paper (Jensen & Scharff, 2014), we wrote:

E-texts and e-readers offer tools that overcome the challenges posed by paper textbooks. In an e-text, students can insert (and delete or change) highlights and annotations and they won't ever run out of room to elaborate their thoughts. More importantly, with some thoughtful course design, teachers using e-texts in the classroom can publicly model the art of critical reading while students can work as true apprentices, receiving feedback as they work in class or through the easy electronic sharing of annotations. In our view, the fact that

e-texts can make the previously private act of critical reading into a public act represents one of the most positive transformational aspects of these new technologies. (pp. 83-84)

We are not alone. Alan Dennis and his team at Indiana University have conducted a number of studies of e-reader use at their institution (Dennis, 2011; Dennis, Morrone, Plaskoff & McNamara, 2015; Dennis, Abaci, Morrone, Plaskoff & McNamara, 2016). They begin with assumptions that we share about the importance of critical reading skills for learning:

Learning is not a passive process where students simply receive information, but an active process in which students co-construct knowledge. They build upon prior knowledge and experience as they make sense of the textbook, revising their own current understanding as they encounter new ideas and information and as they test their current schema. Annotation of texts can make an important contribution to both the cognitive and metacognitive aspects of learning. Underlining and highlighting may assist in recall. More complex annotation strategies, such as summarizing, paraphrasing, finding examples, and asking questions, contribute to metacognitive monitoring and enhance learners' self-regulation, recall, and comprehension (Dennis et al. 2015, 5254-5255).

However, to date, Dennis and his colleagues have focused on the learning benefits gained when instructors share their own annotations with students. They have not studied the learning benefits that might be achieved when students write and/or share their own annotations. For example, in their 2016 study, they describe the effect of embedding instructor annotations in a textbook. Students with access to instructor annotations performed better on subsequent tests of the material than students with access to the text alone. These results are suggestive, but they do not answer questions about the learning potentialities associated with students' own development and use of critical thinking practices (such as annotating a text) in a e-reader environment.

In this paper, we describe our own study that engages this question directly: "What impact does a course intentionally designed around cultivating critical reading skills in an e-reader environment have on student learning?" We share the view of Dennis and his team that e-readers have the potential to enhance students' critical reading skills and the view that they have this potential in light of the specific tools available in e-reader apps that allow students to bookmark, underline, outline, cross-reference, search, and annotate a text. However, we also believe that in order for students to make progress in the development of their critical reading skills in an electronic environment, they must do it themselves. Based on our earlier work (Jensen & Scharff, 2014), we also believe that they must be intentionally taught how to do this. In other words, we believe that instructors must teach students how to make use of electronic tools for critical reading in order for students make learning gains. Instructors must also provide opportunities and assignments that enable students to transform these practices into habits. Merely providing e-readers that have the relevant tools is not enough.

Our study was situated in an upper-level ethics course that is required of all students at the U.S. Air Force Academy. This course is designed to invite serious engagement with primary texts in a discussion setting. Most of the readings are therefore selected from key historical figures (e.g., Plato, Aristotle, Immanuel Kant, and John Stuart Mill) and the classes are restricted to roughly fifteen students. This setting was a natural fit for this kind of study, insofar as the learning outcomes for the course include the cultivation of critical reading skills. We conducted our study across a spring semester and a fall semester. This gap allowed us to refine the pedagogy of the test sections in response to instructor observations, student performance, and student feedback. Test group sections (hereafter, Kindle group sections) were required to install the Kindle App for PC and purchase

Kindle versions of the course texts. Control group sections worked from paper versions of the same texts.

The overall research question that guided our studies was as follows:

What effect would the use of e-texts, together with training and modeling, have on student learning, student attitudes, and student behaviors, when compared to students in control sections who worked from paper texts?

We began with three sets of hypotheses. First, by building electronic annotation requirements into the course and including explicit instructor modeling of how to make and benefit from quality annotations, we predicted that the Kindle group students would become adept at using annotation features in the Kindle reader and, over time, develop better annotation habits. Better annotation skills would be evident if the Kindle group self-reported more increases in their use of annotations than the Control group as well as if we observed an increase in the depth and quantity of annotations submitted by the Kindle group across the semester. Second, by increasing the quality and quantity of their annotations, we predicted that Kindle group students would more deeply process the readings, and thus perform better on the matched final exam questions than the Control group. Also related to performance, for the Kindle group we predicted that depth and quality of their submitted annotations would positively correlate with other performance measures in the course (pre-class reading assignment grades “preflights,” paper grades, individual exams, and the overall course grade). Third, because they developed greater familiarity and appreciation for the e-text annotation features across the semester, we predicted that the Kindle group would show a positive shift in their perceptions of e-texts compared to the Control group. Finally, as an ad hoc hypothesis, due to pedagogical improvements to the Kindle sections of the course between the spring semester and the fall semester, we predicted improvement in all three of the areas above.

Methods

Participants

Participants in this study included students enrolled in thirteen sections of an upper-level, core-required Ethics course, with ten to fifteen students per section. In the spring, we evaluated three Kindle sections and four Control sections. In the fall, we evaluated three Kindle sections and three Control sections. Overall there were 76 students in Kindle sections and 98 students in Control sections. Kindle sections were taught by Dr. Jensen (first author of this paper); Control sections by another member of the philosophy department. Students at the U.S. Air Force Academy are not free to select their instructors for required courses that are taught in multiple sections by multiple instructors. Instead, the registrar assigns students to sections that best accomplish balance in their schedules. This results in a quasi-randomized selection of students in each section who have roughly the same demographic and aptitude mix.

Design and Materials

This study incorporated a two-group comparison design (Kindle sections compared to Control sections), with some additional analyses comparing the spring versus the fall semester for both groups and pre-versus-post semester time periods within each semester. Dependent variables included multiple measures of academic performance as well as self-reports of attitudes and learning behaviors.

The main two-group comparison of academic performance used scores on three common final exam questions, scored using a shared rubric and blind grading of the combined group of exams. Within the Kindle group, we also analyzed the relationships between completion of and performance on pre-class writing assignments based on the reading, paper grade scores, exam scores, and the number and type of text annotations. Text annotations were evaluated for research purposes only, not for grades. We made separate counts of highlights and other types of annotations. We then categorized the annotations as one of four types: annotations that outline the text, annotations that connect the text to other texts and experiences, annotations that question the text, and annotations that raise objections to the text. At the same time, for each of these types, we scored the annotation as reflecting either a “high” level of critical engagement or a “low” or superficial level of critical engagement. For example, a superficial objection in the electronic margin might be as simple as “Wrong!” while a highly engaged objection might show up as several sentences of reasons why the author is wrong.

Beyond measures of academic performance, we created a questionnaire to assess student attitudes and learning behaviors. For all sections during both semesters, we used the same pre-semester form. In the spring, the post-semester forms for the Kindle and the Controls sections were slightly different. In order to obtain more directly comparable results, we used identical post-semester forms for Kindle and Control sections in the fall. On the pre-semester and post-semester forms, both the Kindle and the Control sections were asked about their preference for print or electronic texts (closed-ended), their reasons for these preferences (open-ended), the extent to which they experienced distractions when using their electronic device (open-ended), and to describe their typical annotation behaviors (closed-ended; choose one of seven options: tend not to read; read only what is absolutely necessary; skim assigned readings; read assigned readings; read and highlight or underline; read, highlight and annotate; read, highlight, annotate and separate notes). On the post-semester questionnaires only, students were additionally asked to explain any perceived changes in their annotation habits (open-ended), and whether or not they recommended wider adoption of e-texts at our institution (open-ended; not asked of the Control group during spring semester).

Procedure

In the six Kindle sections (three each semester), students were required to install the Kindle App for PC and acquire all of their textbooks through the Kindle store. In the seven control sections (four in spring and three in the fall) students used paper books. The Kindle sections were all taught by Dr. Jensen; the control sections were all taught by another instructor. While the Kindle and Control sections did not work from an identical syllabus, all sections of Ethics taught at the U.S. Air Force Academy have identical learning objectives and conform to a “Course Contract” that requires a specific set of readings, reading-centered seminar-style discussions, a final paper, and a final exam. The main texts for Dr. Jensen’s sections were Plato’s *Republic*, Kant’s *Groundwork for the Metaphysics of Morals*, Mill’s *Utilitarianism*, Aristotle’s *Nicomachean Ethics*, and David Fisher’s *Morality of War*. Dr. Jensen also assigned three short papers that required a close analysis of a key passage from a text. Every student wrote a paper on Plato and Fisher; for the third, students could select from Kant, Mill, or Aristotle. In addition, Dr. Jensen and the other instructor developed a common set of ten short answer questions for the final exam based on shared texts between the Kindle and Control sections: Aristotle’s *Nicomachean Ethics*, Immanuel Kant’s *Groundwork for the Metaphysics of Morals*, and John Stuart Mill’s *Utilitarianism*. It should be noted that, despite the structural similarities between Kindle and control classrooms and the close collaboration between Dr. Jensen and the other instructor, differences in their teaching styles and the open-ended nature of seminar-style discussions are potential confounding variables. In the discussion below, we are careful to distinguish

comparisons that are made between sections taught by different instructors and comparisons that are made between sections taught by the same instructor. We also qualify our analysis in accord with these limitations.

Distinctive pedagogical features of Kindle sections included the following. First, on the second day of class of both the spring and fall semesters, Dr. Jensen taught a lesson on critical reading practices with the Kindle App for PC. Critical reading practices include highlighting the text, outlining the text using the annotation function, as well as annotating the text with questions, comments, and objections using the annotation function. Students were also taught how to aggregate and submit their annotations for research purposes. During the spring semester, student annotations were collected at the middle and end of the term for research purposes; they were not graded. During the fall semester they were collected at the end of each of the five units / philosophers. Second, Dr. Jensen assigned pre-class writing assignments based on the day's reading assignment, which required students to bring a typed question, comment, or objection to class for discussion. Students were assigned approximately 25 pre-class writing assignments over the course of the term. These assignments were graded on a four-degree scale: not-proficient (zero), proficient (75%), highly proficient (88%), and mastery (100%). Third, throughout the course of the term, Dr. Jensen led the class discussion from the e-text itself, and his highlights and annotations were projected onto the screen. For each main text in the course, at least one class session was devoted to an in-class group assignment focused on understanding and evaluating important passages.

After reviewing his experiences teaching with the Kindle App in the spring, Dr. Jensen made the following additions to the fall semester course. First, he collected student annotations five times rather than twice. Five sets of annotations made more sense, given that there were five discrete texts. Second, as part of submitting their annotations, students were also required to submit a short metacognitive reflection paper that required them to reflect on their experience with critical reading in the e-reader environment. This assignment asked students to briefly explain their sense of how well they understood the text, their perception of how reading in the electronic environment did or did not contribute to their understanding of the text, and their perception of how their use of the skills of critical reading (e.g., highlighting and annotating) contributed to their understanding of the text. This assignment forced students to reflect on their experience with e-reading with an eye on how their practices could be improved for the next book.

Control sections, like the Kindle sections, were small seminars based on discussion of the texts. Reading assignments in Control sections were of similar length to those in Kindle sections and from the same texts. With the exception of common final exam questions, Kindle sections and Control sections did not have other academic work exactly in common. However, as all sections had the same overall course objectives, Kindle sections and Control sections alike were focused on close readings of primary sources in order to identify and critically engage with the same set of themes.

To capture students' attitudes and perceptions of their behaviors, a neutral third person not associated with the class administered the questionnaires to all sections on the second day of class and again during the final week of class. Students were informed that participation was voluntary, that names would be removed from the data set once data were linked at the end of the semester, and that instructors would not have access to any of the questionnaire data until after grades were submitted at the end of the semester. Students were given 10-15 minutes at either the beginning or end of the lesson to complete the questionnaires during class time.

Results

Prior to analyses, pre-post questionnaire responses were linked and names were removed. Recall that the only data collected for the Control group were the pre and post questionnaire responses and the

common final exam question scores. Thus, Kindle vs. Control comparisons were only possible for those measures. Within the Kindle group, where we had additional performance measures, we completed some additional analyses comparing spring vs. fall semester and some correlations between annotations and performance. For the questionnaire data, for each question we first performed Chi Square comparisons between spring and fall semesters for each group to determine whether or not we could collapse the semester data, leaving two key groups for analysis: all Control and all Kindle.

H1: Annotation Data – Self-reported and Performance measures

We hypothesized that by building electronic annotation requirements into the course and including explicit instructor modeling of how to make and benefit from quality annotations, the Kindle group students would become adept at using annotation features in the Kindle reader and, over time, develop better annotation habits. Annotation data included two questionnaire items that asked students about their annotation practices at the end of the semester, and categorical scoring of the annotations submitted by the Kindle group. The first annotation question on the questionnaire asked, “Did you perceive a change in your annotation habits, i.e. use of outlining, questions, connections, objections, in this class during the course of the term?” Response options included “Got worse,” “No change,” “Somewhat different,” and “Significant change.” There were no fall-spring differences within either the Kindle or the Control groups, although there was a positive pattern for more change in the fall semester Kindle group compared to the spring (38% reported significant changes in the fall while only 20% did so in the spring), perhaps due to the fall semester course design changes that more explicitly incorporated annotations and metacognitive reflection. With semesters combined, the Control vs. Kindle comparison was highly significant, with 81% of Control group participants reporting no change in their annotation habits, while 74% of the Kindle group reported “Somewhat different” or “Significant change,” $\chi^2(3) = 66.65, p < .01$.

The second annotation question asked students to indicate the types of reading and annotation practices in which they engaged for both their core courses and major’s courses. There were seven response options: Tend not to read; Read only what is absolutely necessary; Skim assigned readings; Read assigned readings; Read and highlight or underline; Read, highlight and annotate; Read, highlight, annotate & separate notes. The number of students responding “tend not to read” was essentially zero, so this level was not included in the analyses. Chi Square analyses showed no significant differences between the groups at any time (pre-post or spring vs. fall) within either type of class (core or major). However, there were significant differences between core and majors courses for pre semester (all students combined), $\chi^2(5) = 14.14, p < .05$. Students in general were more likely to engage in more reading and annotation for the majors courses (almost all responses in the top four levels) than for the core courses (almost all responses in the bottom four levels). There was a similar trend at the end of the semester, but it was not significant.

For the Kindle group only, we also compared the actual numbers and types of annotations across the semester for both spring and fall. Remember that the text annotations were categorized into one of five types: highlighting, outlining, connecting, questioning and objecting. For analysis, annotations were grouped into three categories: highlighting, low-level annotations (superficial levels of each of the remaining four types of annotation) and high-level annotations (more in-depth examples of the remaining four types). The time factor was assessed by comparing the number of annotations for five different philosophers that were discussed in order across the semester: Plato, Kant, Mill Aristotle, and Fisher. Because the different texts were different lengths, the number of annotations in each category was tallied and then divided by page count in order to give a measure of annotations per page per philosopher. We performed a 2 (semester: spring, fall) x 3 (types of

annotation) x 5 (time / philosopher) mixed ANOVA using the annotations per page data. This analysis resulted in all main effects and interactions being significant except the interaction between semester and time/ philosopher. See Figure 1 for graph of the means for each condition. Overall, the students in the fall semester produced more annotations than those in the spring, $F(1,592) = 13.54$, $p < .01$, partial eta squared = 0.16. There was a significant decrease in the number of annotations across the semester, $F(4,592) = 44.82$, $p < .01$, partial eta squared = 0.38, and there was significantly more use of highlighting than of low-level annotations, and more low-level than of high-level annotations, $F(2, 592) = 104.97$, $p < .01$, partial eta squared = 0.59. The significant interaction between semester and type of annotation, $F(2, 592) = 11.13$, $p < .01$, partial eta squared = 0.13 showed that, while use of all three types of annotation were greater in the fall than in the spring, the greatest difference was in the use of highlighting. The significant 3-way interaction, $F(8, 592) = 3.61$, $p < .01$, partial eta squared = 0.05 further modified these effects by showing that the decrease across the semester was much more gradual in the fall semester than in the spring semester.

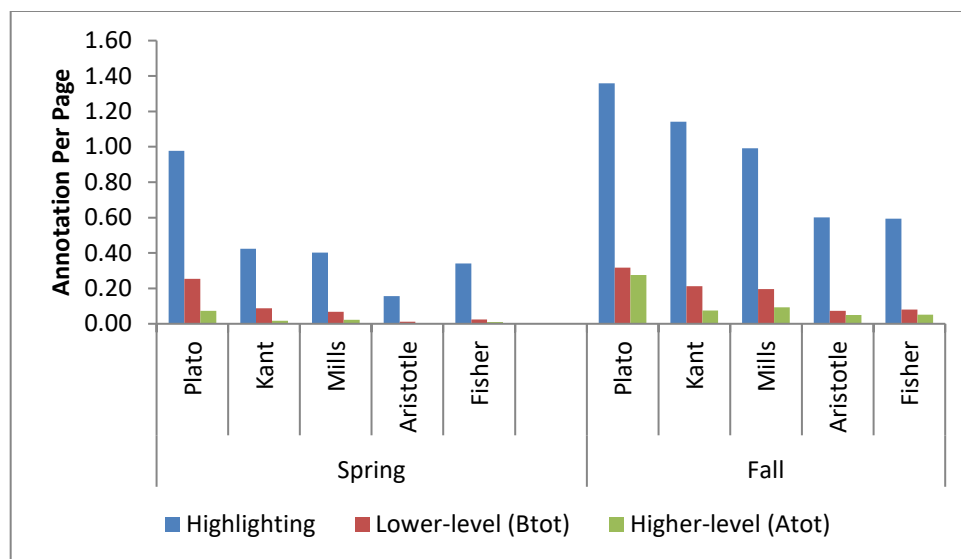


Figure 1: Kindle Annotation Results: Spring vs. Fall

H2: Performance Data – Group Comparisons

We predicted that by increasing the quality and quantity of their annotations, the Kindle group students would more deeply process the readings, and thus perform better on the matched final exam questions than the Control group. Also related to performance, for the Kindle group we predicted that depth and quality of their submitted annotations would positively correlate with other performance measures in the course (preflight grades, paper grades, individual exams, and the overall course grade). Prior to completing any Control versus Kindle group comparison analyses on the final exam scores, we created three average scores, one each for each philosopher (Kant, Mill, Aristotle) from the 10 common short-answer questions. We also checked to see whether or not we should include prior GPA as a covariate. GPA did significantly correlate with each of the three average scores (Kant, Mill, and Aristotle, $r = 0.25$ and $p < .05$, $r = 0.29$ and $p < .05$, and $r = 0.35$ and $p < .01$, respectively). However, there was not a significant difference between the two groups, $t(172) = 1.84$, $p = 0.07$. Thus, we did not include it as a covariate. Because students had some choice in which exam questions they answered (pick 8 out of 10 questions), the same individuals did not respond to each question and the number of students answering each question was different. Thus,

for each of the three common final exam average scores, we performed an independent groups t-test on the data. Although in each case the Kindle group had higher overall scores, there were no significant differences for the Kant average scores, $t(172) = .75$, $p = .45$ or for the Aristotle average scores, $t(172) = 1.00$, $p = .31$. There was, however, a significant difference in the scores on the Mill average scores, $t(172) = 8.66$, $p < .01$, with the Kindle group scoring higher than the Control group.¹

In addition to the two-group differences in performance on the three average scores, we were curious whether or not annotation performance for the Kindle group would predict grades on other performance measures. Thus, we ran correlations between each of the three categories of annotation (Highlighting, Low-level, and High-level) with each of the aligned performance measures (e.g. Kant annotations with the Kant paper grade, the Kant average score, the Kant pre-flight performance, and the overall course grade).

Depth and quantity of annotation engagement did seem to predict grades on papers, see Table 1. However, although all correlations were positive, the only significant correlations were between High-level annotations and paper grades.²

Table 1: Correlation r values between the three categories of annotation and performance on the aligned paper grades (* $p < .05$).

	Plato paper (N=76)	Kant paper (N=9)	Mill paper (N=32)	Aristotle paper (N=35)	Fisher paper (N=76)
Highlighting	0.11	0.30	0.30	0.10	0.03
Low-level	0.14	0.40	0.25	0.28	0.02
High-level	0.25*	0.19	0.37*	0.35*	0.10

The number and depth of annotations were also somewhat predictive of the other aligned performance measures and especially the overall course scores. For the final exam score, only the number of high-level annotations per page for the Aristotle reading significantly correlated with the average score on the Aristotle final exam questions. For the preflights, the number of high-level annotations was the best predictor of grades, but, annotation engagement in the other categories was also sometimes significantly predictive. Both the number of high-level and low-level annotations significantly predicted final course grades, although even highlighting showed a positive relationship that was nearly significant. Overall, engagement in high-level annotations was the most consistent predictor of performance on other course assignments and final course grade.

¹ This difference was almost entirely driven by differences in scores between Kindle and Control sections on one final exam question. While both Dr. Jensen and the other instructor assigned the text underlying the question and agreed that it was a good question for the test, it may be that the classroom discussion associated with the text was different.

² There were no significant correlations for the Kant paper, likely due to the very small number of students who chose to write that paper, nor the Fisher paper, possibly due to the much smaller number of annotations overall by the end of the semester.

Table 2. Correlation r values between the three categories of annotation and performance on the aligned final exam scores, preflight grades, and overall course average (* p<.05).

N = 76	Kant Final Exam	Mill Final Exam	Aristotl e Final Exam	Plato Prefligh t	Kant Prefligh t	Mill Prefligh t	Aristotl e Prefligh t	Fisher Prefligh t	Final Cours e Grade
Highlightin g	-0.09	0.17	0.04	0.20	0.32*	-0.01	0.03	0.23*	0.22
Low-level	0.11	0.15	0.11	0.11	0.32*	-0.06	0.14	0.22	0.27*
High-level	0.10	0.15	0.27*	0.33*	0.31*	0.06	0.15	0.23*	0.31*

H3: Attitude data – Kindle and Control group comparisons

We predicted that because Kindle group students developed greater familiarity and appreciation for the e-text annotation features across the semester, they would show a positive shift in their perceptions of e-texts compared to the Control group. Unlike the exam performance data, there were significant pre-post differences between the Kindle and the Control groups with respect to questionnaire data. For the question, “At this point in time, would you prefer to read an e-text or a printed text for academic work?” there were no differences between the two groups at the beginning of the semester, with most of the students in both groups reporting that they preferred printed texts (70% of Control and 69% of Kindle), $\chi^2(1) = 0.03$, $p > .05$. However, there was a significant difference at the end of the semester, with an increase in the number of Kindle group students reporting a preference for e-texts (65%), $\chi^2(1) = 20.43$, $p < .001$, and no change in the preferences of Control group students.

We did a qualitative categorical analysis of the reasons students gave for their preferences between print and e-text. First, the open-ended responses were reviewed and five global categories were formed: Ease of use / speed of reading / ability to annotate; Eye strain; Habit / tangibility of print copy; Cost / convenience / weight; Problem with Distractions. Responses were then categorized and tallied for 4 subgroups pre and post (Control group students who reported preference for the print version, Control group students who reported a preference for the e-text, and the same two subgroups for the Kindle group), and Chi Square analyses were performed. The only group showing a significant pre-post difference in reasons for their preference was the Kindle subgroup that reported preferring the e-text. They reported significant increases in ease of use and convenience. Pre and post comparisons between those who reported preferring print (regardless of group, Kindle or Control) compared to those who preferred e-texts (regardless of group, Kindle or Control) show that those who report preferring print are significantly more likely to choose eye strain, habit, or distractions as reasons for their preference, $\chi^2(4) = 20.67$, $p < .01$ and $\chi^2(4) = 20.93$, $p < .01$, pre and post, respectively.

When reporting level of problems with distractions, there were no differences between Kindle and Control at pre-semester (50% of both groups claiming no problem, ~30% claiming a moderate problem, and ~20% claiming a big problem). However, there were significant differences post semester, $\chi^2(2) = 14.94$, $p < .01$. The majority of the Control group (54%) reported smaller levels of distractions, with 32% reporting “about the same” and 14% claiming bigger problems. In contrast, 28% of the Kindle group reported a smaller level of distraction, 68% claimed “about the same” and 28% claimed bigger problems.

Finally, students were asked “Would you recommend wider adoption of electronic books in courses at USAFA?” At the end of the fall semester (no spring Control group data) there were significant differences between the Control and the Kindle groups, $\chi^2(2) = 8.29$, $p < .05$, with 55% of the Kindle group saying “yes” that they should be more widely adopted, 27% saying “it depends” and 17% saying “no.” In contrast, only 40% of the Control group recommended further adoption, with 8% saying “it depends” and 52% saying “no.”

Discussion

A review of our results offers support for many but not all of our current hypotheses, and provides support for our earlier preliminary conclusions (Jensen & Scharff, 2014) that critical reading is challenging to develop, that a course design that pervasively incorporates development of annotation skills using an e-text is effective, and that practice using an e-text and its annotation features can lead to significant positive shifts in attitudes about e-texts.

With respect to the first set of hypotheses regarding the development of critical reading annotation skills, our results offer encouragement for the explicit course design centered on the incorporation of the e-text annotation practices; however, the results also reinforce the challenge of getting students to engage in critical reading. By the end of the semester, the Kindle group was significantly more likely to report increases in their engagement in annotation behaviors than the Control group. Further, although it was not a significant difference, almost twice as many Kindle group students in the fall reported a “significant change” in their annotation habits when compared to the spring. These results suggest that a combination of frequent requirements to engage in annotations combined with explicit in-class modeling, multiple checkpoints (added fall semester) and periodic reflection (added fall semester) are effective in shaping student critical reading behaviors.

Unlike predicted, however, the overall number of annotations decreased across the semester in the Kindle group, rather than increased, although this decrease was significantly less severe in the fall semester when annotation requirements were more pervasive throughout the semester. The overall decrease in the number of annotations is not too surprising as many students start a semester with good intentions, and then decrease their levels of reading and other learning behaviors as the semester goes on. Reading critically requires time and effort (Wolf, 2009). Thus, some of the decrease could be due to constrained schedules as overall student workload increases across the semester, but some might also be due to students learning what is really required for “success” in a course. If students realize that they can do well in a course with less effort, most of them will decrease their effort. The fact that there was significantly less drop-off in the number of annotations in the fall compared to the spring, and the trend toward the significant increases in spring-fall self-reported annotation engagement, underscore the influential role that an instructor’s expectations and course design have on student behaviors: there was better engagement with more explicit and pervasive incorporation of critical reading. An additional factor to consider with respect to levels of critical reading annotation behaviors is the type of course. Both the Kindle and the Control group students were significantly more likely to report engaging in deeper reading and higher-level annotation for their majors courses than for their core courses. Thus, early core courses might be particularly key targets of opportunity for explicit critical reading development.

Our second set of hypotheses focused on the impact of e-text annotation engagement on performance in the course. Overall, if there were differences between the Kindle and Control group, or significant relationships between engagement in annotations and performance for the Kindle group, they supported the benefit of the e-text annotation behaviors. These performance results complement those by Dennis et al. (2016), who showed that access to instructor annotations can increase student performance on exams. Going into our study, we had expected greater differences

between the Kindle and the Control groups. However, given the overall low rates of engagement in high-level annotations, and the fact that high-level annotations are better predictors of performance than highlighting or low-level annotations, the lack of a large effect is not surprising. It does highlight the need for even more course redesign to better support engagement in those high-level annotations.

Also of interest with respect to student performance, and unlike what has been suggested by some prior research (e.g. Baron, 2015; Jabr, 2013; Wolf, 2010), there seemed to be no detrimental effects of using e-texts rather than printed texts. Further, the fact that high-level annotations best predicted performance across a variety of course assessments suggests a distinct benefit of e-texts compared to printed texts. Students are often hesitant to write in printed texts because more marks lead to lower re-sale value (Scharff & Dull, 2011). They are also constrained by the small margins of many college textbooks. Both of these deterrents would be especially true for high-level annotations, but importantly, these two deterrents do not exist for e-texts.

Our third set of hypotheses focused on student attitudes about e-texts and how regular interaction with an e-text and its features might impact those attitudes; we predicted that Kindle students would develop more favorable attitudes toward e-texts as they developed familiarity with them. The aspect of familiarity is key, because otherwise attitudes comparing printed texts and e-texts are likely to be biased toward printed texts, which are much more familiar and which don't require training on how to interact with them. Our pre-semester results clearly supported prior research suggesting that students have a preference for printed texts (Baron, Calixte, & Havewala, 2017); more than two-thirds of both of our groups reported such a preference to start. These students reported that eye strain associated with electronic screens, prior reading habits, or distractions in the electronic environment were reasons for their preference. However, by the end of the semester, two-thirds of the Kindle group students reported a preference for e-texts, while there was no change in preference for the Control group students. The most commonly reported factors for those preferring the e-text were ease of use and convenience. These shifts in preference likely also underlie the significant difference between the Kindle and Control groups with respect to their agreement that e-texts should / should not be more widely adopted. By the end of the semester, 52% of the Control group students but only 17% of Kindle group students said they should not be more broadly adopted.

These attitude results reveal a major weakness of other studies that examine attitudinal data drawn only from single-point-in-time questionnaires (e.g., Baron, 2017). Like most new technologies, adoption rates and preferences change through use and habituation, provided that the technology in question can eventually bring about positive change. This fact is especially important in the academic setting, where the skills, knowledge, and habits that we aim to cultivate are complex. It often takes focused practice, intentionally directed toward the formation of new habits, in order to see progress toward educational goals. Even achieving a novice proficiency at some learning goals can take months; mastery may take years. While the 10,000 hour rule popularized by Malcolm Gladwell may not apply here, the underlying principle that success requires concentrated effort and regular practice, together with the fact that preferences for a new technology follow on the achievement of success with that technology, suggest that simple, one-time assessments of students' preferences are not all that useful. Our data explicitly make this point, insofar as they demonstrate significant change in attitudes over time when coupled with consistent use of the technology.

A final attitudinal result that must be acknowledged is the significant difference in the level of problems with distractions reported between the Kindle users and Control group students by the end of the semester. Although the majority of Kindle users reported a decrease or no change in distractions, almost 30% claimed an increase in distractions. In contrast, only 14% of the Control group reported an increase in distractions when using electronic texts (presumably for other

courses). Unless students use an e-reader only device, such as a Kindle Paperwhite, rather than a computer, the issue of distractions is common and powerful. Instructors across the disciplines struggle with students' multi-tasking in their classrooms, to the detriment of their learning (McCoy, 2013). There are numerous suggestions ranging from discussions of negative learning impact, demonstrations of deleterious effects, banning of electronics, to a laissez faire acceptance and belief that it's the student's choice to be distracted or not (Levy, 2014). We suspect that many of these challenges are best understood as the ordinary growing pains associated with human adaptation to new practices and technologies. As we discover virtues and vices associated with these technologies, we will modify our technologies and develop habits that will be conducive to curtailing vice and encouraging virtue.

Conclusions

Based on our data and analysis, it seems to us that both promoters and detractors of e-readers in the critical reading environment overstate the case. Promoters are mistaken in thinking that simply distributing the technology will be sufficient to confer a benefit for student learning. Our research suggests that critical e-reading, especially for academic purposes, comprises a set of skills and habits that must be taught, modeled, and reinforced in the classroom. It suggests further that progress may be slow and that instructors must be willing to make regular adjustments to their pedagogies in order to achieve learning benefits. Nevertheless, our research also suggests that learning benefits are available for critical reading using e-texts and e-tools—benefits that may surpass those that can be achieved in the paper environment.

Detractors, often relying on snapshot attitudinal data (Woody, Daniel, & Baker, 2010; Baron, Calixte, & Havewala, 2017), are too quick to dismiss the potential for critical reading practice in the electronic environment. Our longitudinal data undercuts the detractors' claims that e-readers are neither preferred by students nor confer a learning benefit to students. Further, we are optimistic that as engineers refine e-readers and their analytic tools, critical e-reading will be more effective and attractive. Technical improvements, such as improvements in readability, page tracking, and annotation tools, will make adoption more attractive. Wider adoption, together with intentionally cultivated practices, will change the critical reading culture.

Concerning the structure and methods of our study, Singer and Alexander have recently argued that, on the basis of a comprehensive literature review, studies comparing e-reading and paper reading need to better “address critical dimensions such as learner differences, text characteristics, and task demands.” (Singer & Alexander 2017, p. 1034) Our study begins the process of answering these calls. First, by situating our study in a core course rather than a majors course, we were able to assess impact using a wide range of learners. Second, by including data from five different texts, we were able to compare performance across a variety of textual characteristics. To be sure: all of these texts were associated with the field of philosophy. It will be important for other researchers in other disciplines to compare e-reading and paper reading in their disciplines in order to capture a wider range of text characteristics. Finally, by assessing the quality of student annotations, we were able to assess much more directly the quality of student engagement with the critical reading exercise. In constructing their literature review, Singer and Alexander excluded studies that focused exclusively on self-reports. We agree that self-reports are important but also limited in their usefulness: our study also examined performance data and annotation data. In particular, we believe that our novel methodology of counting and scoring student annotations represents a form of assessment that should be replicated and expanded. It seems to provide a more direct evaluation of the nature and quality of student engagement with the text. Combined with

performance data (e.g., measures of comprehension and retention), we are able to develop a much better sense of learner progress.

Finally, one result of our study, found in the attitudinal data, strikes us as especially provocative. One student's metacognitive reflection illustrates the way in which changes in attitude work hand in hand with critical reading practices and academic progress:

At first, I disliked highlighting and annotating in the electronic text. I missed having a pen in my hand. However, I surprised myself; I began to like reading from the e-text. It was easy to find passages to be on the same page as our classmates. I also found it easy to read over and track my annotations. Being able to type my annotations rather than write them gave me the leniency to write more than I would in a book. Typing all my thoughts helped me understand the text better because I could ask and answer my questions in a single note whereas in a real book, I would have to squeeze my comments into the tiny margins and would not be as thorough...Often, when I read a paper book, I have lots of questions, but lack of space to write them down let alone attempt to answer them. The slightly more difficult task of writing opposed to typing prevents me from writing down my thought process, thus hindering my understanding. Overall, I have really enjoyed using the e-text to read Plato and think my critical reading skills will improve as we continue with our other texts.

To be sure: this is just one's student's reflection. But we believe that it serves as an example of the transformative possibilities that are present here. We conclude with an analogy we believe is especially apt: a little over one hundred years ago, there were many who preferred the smell, touch, and relationship that one could build with one's means for transportation—the horse. And for a time, cars and trains struggled to compete with the horse, even for speed and ease of use. But for many today, the horse is less effective, impractical, or unavailable. It is our view that the same pattern will be repeated as electronic texts become more pervasive in academic environments.

Acknowledgments

Thanks to the USAFA Program for the Scholarship of Teaching and Learning for supporting this research project. Thanks also to audiences at COLLT, and the Teaching Professor Technology Conference for feedback on presentations of our findings in this study. Thanks especially to Captain (retired) Lindsey Kirchoff, who taught all of the Control sections of this course and contributed to the design and grading of shared final exam questions. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the U.S. Air Force, the U.S. Department of Defense, or the U.S. government.

References

ACT Inc. (2007). *Aligning Postsecondary Expectations and High School Practice: The Gap Defined Policy Implications of the ACT National Curriculum Survey® Results*. Washington, DC: ACT, Inc. Available online at <http://www.act.org/path/policy/pdf/NCSPolicyBrief.pdf>

Baron, N. (2015). *Words Onscreen: The Fate of Reading in a Digital World*. New York: Oxford University Press.

- Baron, N., Calixte, R., Havewala, M. (2017). "The persistence of print among university students: an exploratory study. *Telematics and Informatics*. 34 (5), 590-604.
- Dahl, R. (2000). *On Democracy*. New Haven Connecticut: Yale Nota Bene.
- Daniel, D. & Woody, W. (2013). "E-Textbooks at what cost? Performance and use of electronic v. print texts. *Computers & Education*. 62, 18-23.
- Dennis, A. (2011). *E-textbooks at Indiana University: a summary of two years of research*. <http://crg.iupui.edu/Portals/133/PropertyAgent/15158/Files/367/e-TextbooksatIU.pdf> (accessed June 16, 2017).
- Dennis, A., Morrone, A., Plaskoff, J., & McNamara, K. (2015). "Improving Learning with E-Textbooks." *System Sciences: 48th Hawaii International Conference on System Sciences*, IEEE Computer Society, 5253-5259.
- Dennis, A., Abaci, S., Morrone, A., Plaskoff, J., & McNamara, K. (2016). "Effects of e-textbook instructor annotations on learner performance." *Journal of Computing in Higher Education*. 28(2), 221-235.
- Dirda, M. (2015). *Browsings*. New York: Pegasus Books.
- Foasberg, N. (2014). "Student reading practices in print and electronic media." *College & Research Libraries*, 75(5), 705-723.
- Hobson, E., (2004). "Getting Students to Read: 14 Tips." Idea Paper 40. https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/Idea_Paper_40.pdf (accessed September 26, 2017).
- Jabr, F. (2013). "The reading brain in the digital age: the science of paper versus screens. *Scientific American*. April 11. <https://www.scientificamerican.com/article/reading-paper-screens/> (accessed June 16, 2017).
- Jensen, M. & Scharff, L. (2014). "Using e-book annotations to develop deep reading." *Journal of Teaching and Learning with Technology*. 3(2), 83-86.
- Klinkenborg, V. (2010). "Some Thoughts About E-Reading." *New York Times*, April 15th, p. A26. <http://www.nytimes.com/2010/04/15/opinion/15thu4.html> (accessed June 16, 2017).
- Levy, L. (2014). "7 Ways to Deal with Digital Distractions in Class." *Edudemic: Connecting Education and Technology*. <http://www.edudemic.com/7-ways-deal-digital-distractions/> (accessed September 12, 2017).
- McCoy, B. (2013). "Digital Distractions in the Classroom: Student Classroom Use of Digital Devices for Non-Class Related Purposes." *Journal of Media Education*. 4(4), 5-14.
- Nussbaum, M. (2006). "Education and Democratic Citizenship," *Journal of Human Development*. 7(3), 385-395.

- Scharff, L. & Dull, C. (2011). *Multiple Case Study Investigation of E-reader Devices in Academia*; unpublished.
- Singer, L., Alexander, P., (2016). "Reading Across Mediums: Effects of Reading Digital and Print Texts on Comprehension and Calibration." *Journal of Experimental Education*. 85(1), 155-172.
- Singer, L., Alexander, P., (2017). "Reading on Paper and Digitally: What the Past Decades of Empirical Research Reveal." *Review of Educational Research*. 87(6), 1007-1041.
- Tanner, M. (2014). Digital vs. print: Reading comprehension and the future of the book. *SLIS Student Research Journal*, 4(2). Retrieved from <http://scholarworks.sjsu.edu/slissrj/vol4/iss2/6>
- Wolf, M, (2010). "Our 'Deep Reading' Brain: its digital evolution poses questions," *Nieman Reports*. <http://niemanreports.org/articles/our-deep-reading-brain-its-digital-evolution-poses-questions/> (accessed June 7, 2017).
- Wolf, M. & Barzillai, M. (2009). "The Importance of Deep Reading." *Educational Leadership*. 66(6), 32-37.
- Woody, W., Daniel, D., & Baker, C. (2010). "E-books or textbooks: Students prefer textbooks." *Computers & Education*. 55(3), 945-948.