THE USE OF EXAM NOTES IN AN ONLINE MATHEMATICS COURSE

Mindy Capaldi, Valparaiso University

ABSTRACT

Many students want to use cheat sheets, or crib notes, on exams. Whether or not those aids actually help them has not been carefully studied. This paper measures 16 students' notes by scoring the writing density as well as the number of definitions, examples, and mistakes. To consider the effectiveness of the notes, they are matched against exam solutions. This closer look at the content and application of the online students' notes revealed that they were neither well- made nor well used.

Keywords: cheat sheet, crib notes, mathematics, exams, online

INTRODUCTION

Exams are ubiquitous in college, and students often plead for test aids such as note cards or sheets. This can be especially true in mathematics courses where there are numerous formulas and equations to memorize. There are many arguments for and against allowing cheat sheets. One argument for is that students can focus on applying the definitions or formulas instead of memorizing material. Another potential benefit is a reduction in students' stress during exams. On the other hand, cheat sheets may become a crutch that students rely on instead of thoroughly learning the material.

This paper investigates the use of cheat sheets, or crib notes, by online summer mathematics students. Although the name "cheat sheet" implies cheating, these notes were allowed and encouraged in the course. Hereafter they will usually be referred to as notes or note sheets. Despite the common use of notes (equation sheets, note cards, etc.) for exams, research on their use and effectiveness has been inconclusive. I propose a methodology for assessing notes and performed a statistical analysis of the notes. The goal of this paper is to share these results of allowing student notes for exams and to provide practical suggestions for making the notes beneficial.

After a literature review and description of methodology, I provide quantitative and qualitative results on students' use of note sheets. I measure features of the notes, such as the quantity of writing, definitions, examples, and mistakes. To further assess the effects of allowing notes, I compare the notes and exam solutions for individual students. The conclusion section summarizes the results, gives suggestions for increasing the benefits of exam notes, and proposes directions for future research.

LITERATURE REVIEW

The use of exam notes in the online class was motivated by a desire to discourage cheating. There is a perception that cheating is common in the digital environment (Kennedy, Nowak, Raghuraman, Thomas, & Davis, 2000; Wingo, Ivankova, & Moss, 2017), and research supports the idea that cheating is problematic in these classes. A study of business students found that they believed it is easier to cheat in an online course (King, Guyette, & Piotrowski, 2009). Another investigation of online economic students, where some exams were proctored and some were not, indicated that cheating occurred more in the unproctored environment (Harmon & Lambrinos, 2008). A recent article laid out 50 recommendations to mitigate academic dishonesty online (Nicolaides, 2018). While the focus of that study was business education, two pieces of advice seemed especially pertinent to this paper: allowing open book assessments and effective confirmation of a test-taker's identity. For the Finite Mathematics courses examined in this paper, note sheets were allowed and proctoring was required. There are avenues to discover plagiarism or other cheating in many writing-based online courses (Heberling, 2002), but mathematics courses do not have as much ready advice for avoiding or catching cheating online other than having exams be proctored. Therefore, it is prudent for the instructor to create assessment practices that minimize academic dishonesty.

While one reason to use exam notes is to decrease academic dishonesty, there is also a hope that the notes benefit student learning. Unfortunately, a literature survey found no consensus on whether the use of notes helps exam performance or student learning (Hamouda & Shaffer, 2016). I can see examples of the contradictions of previous results in the following studies.

A few researchers found evidence that crib notes do not affect exam scores (Hindman, 1980; Trigwell, 1987). Hindman (1980) even noted that the number of undergraduate students using notes declined as the semester progressed. Both studies agree that the merit of exam notes is in their creation, i.e., putting thought into what to include in a brief amount of notes necessitates learning the material. On the other hand, Wachsman (2002) reported that both preparing and using notes improved the test performance of economic students. Visco, Swaminathan, Zagumny, and Anthony (2007) examined how ten chemical engineering students prepared their exam notes and concluded that "good" notes do not indicate superior exam performance. More specifically, including a large amount of information on notes did not lead to better exam work. Recently, Hamouda & Shaffer (2016) found, in a data structures course, that if students' exam notes contained information related to higher-level questions, then they performed better on those questions.

Some research has shown that students relied on their notes instead of truly absorbing the material, but that having access to notes or a book decreases test anxiety. A "dependency hypothesis" was first proposed by Dorsel and Cundiff (1979), who found in an experimental study that taking away crib sheets from people who expected to use them caused significantly lower performance in comparison to participants who never planned to use the aid. Dickson and Bauer (2008) observed that by unexpectedly giving psychology students a pretest without notes and posttest with notes, crib sheets enhanced performance but not necessarily learning. Although these studies indicate the ineffectiveness of constructing crib sheets, most students believe that they are helpful and relieve stress (Dickson & Bauer, 2008). Another study of psychology students investigated the differences between open-book, cheat sheet, or closed-book exams and had a follow-up retention quiz and preference survey (Gharib, Phillips, & Mathew, 2012). The results indicated that students did slightly better if they had the additional resources and that they preferred open book and cheat sheets to closed book. They had the least anxiety if they were allowed an open book, even compared to cheat sheets. Allowing an open-book exam for online classes could be problematic, though, since the course likely has an e-book, which leads to the use of electronic devices during exams.

De Raadt (2012) countered the Dickson & Bauer paper by pointing out that cheat sheets are intended to reduce the need for memorization. Therefore, removing the aid goes against that intended use. Instead of taking away their notes, de Raadt (2012) focused on whether using notes enhanced students' performance compared to their previous assessments. He also compared the work of students who used notes and others who did not. Unlike most previous studies, de Raadt analyzed the features present on student-created notes and whether those features related to their exam solutions.

Looking at the content of student note sheets, as opposed to just whether or not notes were allowed for an exam, is a relatively recent avenue of research (Ludorf & Clark, 2014). Visco et al. (2007) explored, through interviews, how ten engineering students made and used their note sheets. Ludorf & Clark (2014) ranked the overall quality, information, density, organization, and color use of notes on various scales. The two authors chose their rankings and then averaged them, but interrater reliability was not always high, such as .521 for organization (Ludorf & Clark, 2014, p. 62). Results from that research showed that higher quality and lower density notes were both associated with higher test performance, although almost all other factors were not related to performance. Hamouda & Shaffer (2016) measured student notes based on criteria such as organization, definitions, print versus written, examples, and color. They then looked for significant associations with students' final exam scores but did not find any when compared to the overall score. There were some significant results when they looked at specific test questions. Finally, Song, Guo, and Thuente (2016) considered variables similar to the other papers, such as density, organization, and number of sample answers, formulas, and graphs. They examined the note sheets of undergraduate and graduate engineering students and based on subjective rankings and adjustments used 1–3 or 0–1 scales for the variables. For example, students who had "too many" graphs received the poorest score, 0, for that variable, since the raters felt that it would not help them. The interrater reliability scores, based on Cohen's kappa, ranged from .69 to .91. That study concluded that note sheets are more useful for undergraduate students than graduate students and that the notes were perceived by students as helpful learning tools and exam aids.

The methods described above relate to those in this paper; I measure density and number of examples/definitions. However, I use new strategies with the goal of less subjectivity and more generalization to various disciplines. I also add variables of interest and extend the investigation to make connections to the content of exam solutions and not just exam scores.

METHOD

Context of the Study

This study uses the final exams from online Finite Mathematics classes, all taught by the author, at a comprehensive, primarily undergraduate midwestern university with approximately 4,500 students. This particular course covers systems of equations, the simplex method, probability, Markov chains, and game theory. It fulfils the quantitative analysis general education requirement and is mandatory for business majors. Students in the course range from freshman to seniors. For most students, it is their last required mathematics class. While this class is taught in person during the academic year, those students are not allowed to use note sheets for exams. Here I investigate the final exams and notes of 16 summer students in the online version of the course.

For the first and second tests of the summer, students could create and use one page of notes for each test. Two pages (front and back of one sheet) were allowed for the final exam. This note sheet policy was established to discourage actual cheating, which is a common concern in online courses (see Literature Review). A calculator was also allowed. To further discourage cheating, all exams were proctored either by the instructor or an approved third party.

Note Sheet Measurements

This introductory study proposes a new method of analysis of student note sheets that can be easily applied to many disciplines, is objective, and goes beyond looking at just exam scores. I also looked at exam solutions. The content and amount of notes were measured for summer students from the two most recent iterations of the online course. This more in-depth look at the students' notes was restricted to those recent terms because one aspect of the online course changed, which was a requirement to participate in a discussion board. While it does not seem likely that this change substantially influenced the content of the notes, there is a chance that it did. In the last two summers, there were 16 students who made notes and three who did not.

Few schemata for evaluating note sheets were found in a literature search, and of those found, most were for computer science or engineering courses (de Raadt, 2012; Hamouda & Shaffer, 2016; Ludorf & Clark, 2014; Nye, Crooks, Powley, & Tripp, 1984). Nye et al. (1984) studied student lecture notes, measuring the quantity of notes through a word count. On the other hand, de Raadt (2012) gauged the quantity by density, or how much vacant space was found on the sheet. Similarly, Hamouda & Shaffer (2016) and Ludorf & Clark (2014) gave percentages of nonblank space to measure quantity, but it was unclear how those percentages were determined. Song et al. (2016) used a 1–3 scale to rate density.

Using a word count is tricky for mathematics notes. Instead of English words, the notes often contain equations and expressions. For example, would numbers and variables count as "words"? Would each entry of a matrix be a word, or should the entire matrix count as one word? Instead of measuring the quantity of the notes through counting words or estimating a percentage of blank space, this study uses an image analyzing software to calculate the percentage of handwriting on the page. The MIPAR software (http://www.mipar.us/) employs a "recipe," which was created specifically for this research to address the question of finding the percentage of handwriting on a page. While this method does not disassociate handwriting size from the quantity of notes, size is also a problem when considering word count or blank space. Measuring the notes' density is only used as a broad overview of quantity and it works well for that purpose. See Figure 1 for a comparison of low and high densities. The author's personal judgement also deems these two note examples as the lowest and highest quantities out of the sixteen examined.

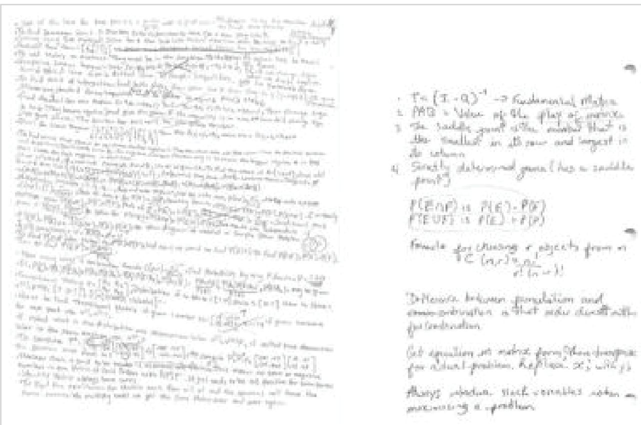

Figure 1. Density.

The notes on the left had the highest density at 14.5. The notes on the right had the lowest density at 2.9. Only the first page of each student's notes is shown.

Two other features de Raadt (2012) looked at were code examples (it was a programming course) and answers written in students' notes. Hamouda & Shaffer (2016) also measured definitions and code examples as a fraction of the overall note content. Along these lines, the current study counted the number of mathematical examples and definitions. Of interest was which type of information students believed would be more useful: specific examples or generalities, such as definitions and formulas. A final category of interest arose when it became apparent that multiple students wrote incorrect definitions or formulas; the number of mistakes was consequently recorded. In summary, the following data were included in the results:

- Note "Density"
- Number of Definitions/Formulas
- Number of Examples (both partial and those with solutions)
- Number of Mistakes

Although previous studies (De Raadt, 2012; Hamouda & Shaffer, 2016, Ludorf & Clark, 2014, Song et al., 2016) also considered color use and/or organization of student notes, this study does not. Organization rankings are subjective, and I want to avoid such measurements. Since some of the notes for the online class were scanned for submission, color was not always apparent and cannot not be used as a variable. Other measurements by the former studies were not applicable as they pertained to the specific courses involved or could not be applied to this mathematics course, such as whether notes were typed (typing mathematics can be hard for students). The variables examined here apply to many courses and disciplines beyond mathematics.

Comparing Student Notes and Exam Solutions

After analyzing their notes separate from the exams, I then compared each students' note sheets to their exam solutions. Making this direct connection between what was written on a note sheet to what was written on the exam is a new strategy not found in the previous literature. I looked at whether mistakes in the notes appeared on exam work. This direction did not yield fruitful results, as many of the topics connected to the mistakes were not asked about on the test. Instead, I turned to the following questions and looked for patterns or remarkable occurrences. The first question is similar to a variable considered by Hamouda & Shaffer (2016), who marked whether each exam topic was covered on note sheets. The second question is an extension of that idea to see whether having notes on a topic is actually beneficial.

- What portion of the exam topics did students include in their notes?
- If they did write notes on an exam topic, were those notes used in their solutions?
- Are definitions or examples more useful to have on the notes?

Each student's notes were examined and records were kept of whether they had written anything (definition or example) for each of the 11 main exam topics. Next, their solutions were checked for each of those 11 questions and points deductions were noted, as well as what sort of mistakes were made. Blank answers and whether or not the student had corresponding notes were also recorded. Finally, I turned to a specific problem about that is notoriously difficult for students. For this question, I tracked whether students wrote the related definition or an example, or both, in their notes and compared the correctness of the solution.

RESULTS

Quantity

Did students who took more notes score better or worse on the exam? The quantity of student notes, based on the MIPAR software, was recorded as a percentage of handwriting on the page. The lowest percentage was 2.9 and the highest was 14.6, as seen in Figure 1. The student with the least amount of notes (out of those who wrote any notes) scored a 79.5% on the exam and the student with the most notes earned a 62%. A regression analysis, where the three students who did not make notes were included as zero densities, showed that the density of students' notes did not have an association with their final exam grade (F = .98, df = (1,17), p = .33). Overall, quantity did not correlate to quality in terms of notes and final exam scores.

Definitions and Examples

Students in this course were provided with lecture notes through the Blackboard learning management system. The course notes specifically labelled each definition, and there were a total of 65 definitions. Between the notes and homework, there were many more examples than definitions. However, more students wrote definitions and formulas on their notes than examples or exercises. The average number of definitions/formulas on students' notes was 22.9 compared to 13.7 examples (see Table 1 for a full list of measurements).

Given the small sample size (n = 16), the nonparametric Wilcoxon Signed-Rank test is more appropriate than a matched pairs t-test for comparing the number of definitions and examples. The Wilcoxon Signed-Rank test takes into account the magnitude and signs of the differences between the definition and example counts. In this case, there is evidence that the median number of definitions/formulas is significantly greater than the median number of examples ($\boldsymbol{T}^{t} = 101.5$, $\boldsymbol{p} = .042$). Although this statistic indicates a preference for definitions and formulas, the question remains as to whether students correctly use those definitions. I return to this question later.

Table 1. Count of Density, Definitions, Examples, and Mistakes on Students' Note Sheets

	Density	Definitions	Examples	Mistakes
1.	3.39	26	3	0
2.	8.29	26	1	9
3.	5.42	13	34	3
4.	2.90	8	0	3
5.	8.30	11	36	2
6.	6.52	10	17	9
7.	4.36	7	16	1
8.	12.84	33	35	0
9.	7.65	26	11	2
10.	5.02	15	8	1
11.	7.73	20	2	2
12.	14.55	33	16	8
13.	13.24	52	2	2
14.	3.89	31	17	4
15.	10.29	27	6	2
16.	7.73	29	15	0
Average:	8.11	22.94	13.69	3.0

Mistakes

Finally, I turn to the number of mistakes, or incorrect statements, that appeared in the notes. Out of the 16 notes examined, only three had no mistakes. One of these (student 3 in Table 1) had very few notes written and therefore less room for error. That student scored a 92.9% on the exam. The other two students with mistake-free notes had scores of 91.7% and 37.5%, respectively. There were four students who scored higher than a 90% (in the A range) on the exam, and I can see that two of those were able to write completely correct notes. On the other hand, the third student with no mistakes had the lowest score on the final exam. Apparently, this student was able to copy verbatim from the course notes and/or homework but did not truly understand the material.

The average number of mistakes on a student's notes, including those students who had none, was 3.0. There were students with as many as nine incorrect statements or formulas. One such error, which occurred on multiple cheat sheets, was with the formula for the fundamental matrix of a Markov chain problem. The formula is $\boldsymbol{T} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}$; students forgot the inverse notation (-1) in the exponent position. The number of students with mistakes, and the quantity of those mistakes, was surprising given that they were not asked to create the notes from memory and were expected to copy them from the course materials.

Multivariable regression analysis using some or all of the number of definitions, examples, and/or mistakes (or ratio of mistakes to density) as explanatory variables and final exam scores as the response did not give significant results, perhaps due to the small sample size.

Comparing Student Notes and Exam Solutions

First, I consider whether students are successful at choosing the topics for their notes and at using the notes. Students were given a list of final exam topics, not all of which ended up on the exam. On average, the 16 students whose notes were examined referenced 7.5 of the 11 exam topics on their notes. They were more likely to include notes on recent material rather than early topics. Including notes, even a formula, for a topic did not imply that they applied those notes to the test. There were four instances from four different students where a problem was left completely blank despite the fact that the student had written notes related to the question. For example, when asked about the "independence" of two events in a probability question, they did not write the equation to check independence that they had in their notes. On the other hand, one student was able to get partial credit for simply recopying a formula even though she did not follow up with any further work.

A striking feature of the students' notes was how often a student would write the same formula multiple times. This especially occurred with the topic of probability. One student wrote the formula for a conditional probability six times. In fact, half of the 16 notes examined had that formula more than once. This repetition happened with other equations, as well. In some cases the student may not have been aware that they had the same formula. Although I consider $C(x,y)=x!/y!(x-y)!$ and $C(n,r)=n!/r!(n-r)!$ to be equivalent, the student likely did not have the same realization.

Since students were more prone to writing definitions and formulas in their notes, did that practice benefit them on the exam more than writing examples? As previously mentioned, the mere inclusion of a definition did not imply its use. Similarly, having a related example on their notes did not always transfer to students' exam solutions. For example, two students wrote an example for finding the probability that two freshmen, from a group of five freshmen and four seniors, were chosen for a committee. The exam question was about the probability of choosing for a committee no Democrats from a group of 51 Democrats and 49 Republican senators. The students got partial credit because they successfully used the formula for counting the number of ways to choose the members ($C(n,r)$), but then missed points for not dividing to get the probability. It appeared that these students could more easily focus on the counting formula rather than reading through their notes to find the probability example. In general, the application of a technique seen in an example did not transfer to the associated exam question.

CONCLUSION

The online students were originally encouraged to use cheat sheets on their exams to discourage actual cheating, but there was also a hope that creating the notes would be beneficial as they reviewed and summarized course material. Three factors led to a conclusion that students did not generally make gains in terms of studying despite writing down notes. First, there were numerous mistakes on almost all students' notes. Second, they were not sufficiently absorbing the material to notice that they were writing the same formula multiple times. Finally, they were unaware of what was in their notes to the point that many were unable to find examples that matched test questions. There was statistical evidence to indicate that students prefer including definitions rather than examples; indeed, definitions seemed to be more useful when a student included both. The quantity, or density, of a student's notes did not associate with higher scores on the exam. Students, on average, included notes covering 7.5 of the 11 main topics on the exam and usually left out material from the first quarter of the course.

Obviously, the mere allowance of notes does not make them a worthwhile endeavor for students. Organization is probably important but difficult to objectively measure as a researcher. Quantity of notes is another factor, but students need more guidance on how to reduce and emphasize information. For example, instead of including every bit of a ten-step solution, they could write a sentence summarizing what is done in the solution. By synthesizing the information, the example is less likely to get lost in a sea of notes or not be used simply because they do not actually know what happened in the solution. More instruction is needed to make the practice of summarizing material for cheat sheets useful. In the future, I plan to make it clear that quantity is not quality. This conclusion matches that of Ludorf & Clark (2014), who found that the denser a student's notes were, the lower their performance. They also recommend that instructors become more involved with cheat sheet construction.

A next step in this investigation of the use of notes in online courses could be implementing a peer-review process for the notes. Students would have the opportunity to share their notes with classmates and get feedback. Visco et al. (2007) conjectured, at the end of their report, that a peer-group prepared study guide might be more beneficial than self-constructed notes. Hanson, Millington, and Freewood. (2001) similarly recommended including peer-review when designing online assessment. Many of the mistakes and repetition of formulas noted in this paper may have been avoided if students' notes were subject to review. Also, the lapses in organization and neatness could be prevented if they are motivated to prepare notes for their peers ahead of time. Sharing notes encourages a discussion of what topics are important enough to be on the notes.

Future studies could compare students' notes, given such additional instructions and feedback, to the original notes. Alternatively, a controlled experiment could be implemented where sections of the course are randomly assigned to have cheat sheets or not. If the sections are taught the same otherwise, with identical exams, stronger conclusions could be made than this comparison of online summer semester students.

REFERENCES

de Raadt, M. (2012). Student created cheat-sheets in examinations: Impact on student outcomes. In Proceedings of the Fourteenth Australasian Computing Education Conference (Vol. 123, pp. 71–76). Melbourne, Australia.

Dickson, K. L., & Bauer, J. J. (2008). Do students learn course material during crib sheet construction? Teaching of Psychology, 35(2), 117–120. doi:10.1080/00986280801978343

Dorsel, T. N., & Cundiff, G. W. (1979). The cheat-sheet: Efficient Coding Device or Indispensable Crutch? The Journal of Experimental Education, 48(1), 39–42. doi:10.1080/00220973.1979.11011712

Gharib, A., Phillips, W., & Mathew, N. (2012). Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety. Psychology Research 2(8), 469–478. Retrieved from https://eric.ed.gov/?id=ED537423

Hamouda, S., & Shaffer, C. A. (2016). Crib sheets and exam performance in a data structures course. Computer Science Education, 26(1), 1–26. doi:10.1080/08993408.2016.1140427

Hanson, J., Millington, C., & Freewood, M. (2001). Developing a methodology for online feedback and assessment. Leicestershire, UK: Loughborough University. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/1812

Harmon, O. R., & Lambrinos, J. (2008). Are online exams an invitation to cheat? The Journal of Economic Education, 39(2), 116–125. doi:10.3200/JECE.39.2.116–125

Heberling, M. (2002). Maintaining academic integrity in online education. Online Journal of Distance Learning Administration, 5(2).

Hindman, D. C. (1980). Crib notes in the classroom: Cheaters never win. Teaching of Psychology, 7(3), 166–168. doi:10.1207/s15328023top0703_10

Kennedy, K., Nowak, S., Raghuraman, R., Thomas, J., & Davis, S. F. (2000). Academic dishonesty and distance learning: Student and faculty views. College Student Journal, 34(2), 309–314.

King, C. G., Guyette, R. W., & Piotrowski, C. (2009). Online exams and cheating: An empirical analysis of business students' views. Journal of Educators Online, 6(1). Retrieved from https://eric.ed.gov/?id=EJ904058

Ludorf, M. R., & Clark, S. O. (2014). Help sheet content predicts test performance. In W. S. Altman, L. Stein, & J. R. Stowell (Eds.), Essays from Excellence in Teaching Volume XIII (pp. 60–63). Society for the Teaching of Psychology.

Nicolaides, A. (20180326). The pursuit of ethical practice in distance higher education assessment in a Business Leadership School. Global Journal of Economics and Business Administration, 3. doi:10.28933/gjeba-2018-01-1802

Nye, P. A., Crooks, T. J., Powley, M., & Tripp, G. (1984). Student note-taking related to university examination performance. Higher Education, 13(1), 85–97. doi:10.1007/BF00136532

Song, Y., Guo, Y., & Thuente, D. (2016). A quantitative case study on students' strategy for using authorized cheat-sheets. In 2016 IEEE Frontiers in Education Conference (FIE) (pp. 1–9). doi:10.1109/FIE.2016.7757656

Trigwell, K. (1987). The crib card examination system. Assessment & Evaluation in Higher Education, 12(1), 56–65. doi:10.1080/0260293870120106

Visco, D., Swaminathan, S., Zagumny, L., & Anthony, H. (2007). Interpreting student-constructed study guides. In Proceedings of the 114th Annual ASEE Conference & Exposition (pp. 1–9). Honolulu, HI.

Wachsman, Y. (2002). Should cheat sheets be used as study aids in economics tests? Economics Bulletin, AccessEcon, 1(1), 1–11.

Wingo, N. P., Ivankova, N. V., & Moss, J. A. (2017). Faculty perceptions about teaching online: Exploring the literature using the technology acceptance model as an organizing framework. Online Learning, 21(1), 15–35. doi:10.24059/olj.v21i1.761