Crossword Puzzles and Knowledge Retention
Elisa R. Torres, University of Mississippi Medical Center, etorres@umc.edu
P. Renée Williams, University of Mississippi Medical Center
Wondwosen Kassahun-Yimer, University of Mississippi Medical Center
Xiaoshan Zhu Gordy, University of Mississippi Medical Center

Abstract. Students perceive crossword puzzles as enjoyable. In addition to students’ perceptions, crossword puzzles actually improve knowledge retention. However, crossword puzzles increased exam scores for some students but not others. Recommendations have been made for students to create puzzles for their classmates to complete with the rationale that students are encouraged to research and understand the material in order to write meaningful clues for the puzzle. While students enjoy creating their own crossword puzzles, the association between students creating crossword puzzles and knowledge retention is unknown. The purpose of this project was to determine if creating crossword puzzles and completing peers’ crossword puzzles were associated with improved knowledge retention indicated by higher scores on quizzes. Students in a research course from two institutions across three semesters had the option each week to upload a blank puzzle they created prior to completing each other’s puzzles and taking a quiz. Quiz scores were compared between those who did versus did not create their own puzzles and complete their peers’ puzzles. Results varied by institution and programs as well as the same program within the same institution but different semesters. Results highlight the importance of moving beyond student perceptions and towards assessing knowledge retention while taking into consideration institution, program, and semester.

Keywords: crossword puzzles; knowledge retention; active learning; educational activities; educational measurement

Active learning is defined as “anything that involves students doing things and thinking about the things they are doing” (Bonwell & Eison, 1991, p. 19). In contrast to traditional teacher-centered learning that often positions students as passive recipients of instructional content (Gregory, 2002), active learning requires students to become active participants who must make decisions and solve problems (Franklin et al., 2003).

Crossword puzzles are an example of active learning. Crossword puzzles require the ability to remember factual knowledge and provide foundational knowledge, the basic understanding that is necessary for other kinds of learning (Fink, 2013). Thus, crossword puzzles would meet a course objective of demonstrating foundational knowledge, which is at the remembering level of Anderson’s revision of Bloom’s
Taxonomy of Educational Objectives (Anderson & Krathwohl, 2001; Krathwohl, 2002).

Crossword puzzles improve knowledge retention (Abuelo et al., 2016; Gaikwad & Tankhiwale, 2012; Murphy et al., 2016; Nirmal et al., 2020; Orawiwatnakul, 2013; Patrick et al., 2018; Singh Matreja et al., 2021). However, crossword puzzles increased exam scores for some students but not others (Davis et al., 2009).

Recommendations have been made for students to create puzzles for their classmates to complete with the rationale being that students are encouraged to research and understand the material in order to write meaningful clues for the puzzle (Davis et al., 2009). While students enjoy creating their own crossword puzzles (Coticone, 2013), the association between students creating crossword puzzles and knowledge retention is unknown.

Crossword puzzles can be cumulative or noncumulative. Cumulative learning is an assignment that requires students to draw on information they learned earlier in the semester (Lang, 2016). For example, a weekly cumulative crossword puzzle assesses content in the course up to that point as opposed to a noncumulative weekly crossword puzzle that only assesses content for that week. Students who have cumulative assessments retain more information than students who have noncumulative assessments (Beagley & Capaldi, 2016; Khanna et al., 2013; Lawrence et al., 2013). The association between students creating and completing cumulative crossword puzzles with knowledge retention is unknown.

Creating and completing weekly cumulative crossword puzzles may be a method to address the issue that some students can struggle with a semester-long scaffolding project (Vandiver & Walsh, 2010) because by the time they get towards the end of the semester, they have forgotten what they learned early in the semester. One way to address this issue may be to have weekly cumulative assignments throughout the semester. The purpose of this quality improvement project was to determine if creating and completing weekly cumulative crossword puzzles were associated with improved knowledge retention as indicated by higher scores on corresponding weekly cumulative quizzes.

Method

Sample

Participants comprised students from two state institutions across three semesters. The first semester included students in a Doctorate of Nursing Practice (DNP) or Doctor of Philosophy (PhD) program at a university in the Midwest of the United States. Students were enrolled in Nursing Research, a required course in the DNP (optional in the PhD) program that fulfilled the American Association of Colleges of Nursing (2006) Essentials of Doctoral Education for Advanced Nursing Practice: Clinical Scholarship and Analytical Methods for Evidence-Based Practice. This was a 16-week hybrid course where students met three hours face-to-face once a month for four months, with the rest of the course online. Prior to the start of the course,
the university’s education Institutional Review Board (IRB) reviewed and exempted this quality improvement project. The second and third semesters included students enrolled in a Master of Science in Nursing (MSN) program at a university medical center in the Southeast United States. Students were enrolled in Research Design and Methods for Advanced Nursing Practice, a required course that fulfilled the American Association of Colleges of Nursing (2011) Essentials of Master’s Education in Nursing: Essential IV: Translating and Integrating Scholarship into Practice. This was a 16-week online course. Prior to the start of the course, it was determined by the IRB that this quality improvement project was not considered research and therefore would not be reviewed by the IRB according to institutional guidelines. There were no conflicts of interest in either course.

Context

Both courses examined quantitative and qualitative research methods; the interrelationships among theory, research, and practice; the research process; critical evaluation of research findings; and ethical criteria for the protection of human subjects in research. The prerequisites were graduate standing or consent of the instructor. The courses were taught by a PhD-prepared nurse with an active program of research, who designed the current quality improvement project as a participant in the Wisconsin Teaching Fellows and Scholars (WTFS) Program. The WTFS program offered University of Wisconsin faculty and teaching academic staff a unique opportunity to collaborate with other teachers from various disciplines across the University of Wisconsin System. In addition to discussing the pedagogical literature, participants were guided through the process of completing a Scholarship of Teaching and Learning (SoTL) project with input from fellow participants and the program co-directors.

Intervention

Students were assigned two chapters most weeks from the textbook *Nursing Research: Generating and Assessing Evidence for Nursing Practice* (Polit & Beck, 2012). Online quizzes that corresponded with the chapter readings were assigned. All questions came from the test bank of the course textbook available at no cost. Completing weekly quizzes was associated with improved knowledge retention; higher scores on the weekly cumulative quizzes were associated with higher scores on the cumulative final exam (Torres, 2019), providing evidence of predictive validity, a form of criterion validity (Polit & Beck, 2012). Evidence for reliability in the quizzes was found with a Cronbach alpha of .69.

For extra credit and to assist with preparing for the weekly cumulative quizzes, weekly cumulative crossword puzzles that corresponded with the chapter readings were due four days prior to the quizzes. Uploading a blank puzzle and corresponding answer key to the learning management site four days prior to the quiz due date allowed students sufficient time to complete each other's puzzles prior to taking the weekly quiz. Students were provided a link to a site where they could complete the puzzles: *Crossword labs* (https://crosswordlabs.com/). Every
puzzle was to have at least one term from each chapter, with the length of the puzzles increasing each week from at least two terms on the puzzle in the first week to at least 26 terms on the final puzzle at the end of the semester. Students scored 100% if they submitted their puzzles on time and the answers were correct and 0% if their puzzles were not submitted on time or any answer was not correct. Two PhD-prepared faculty checked each submission to ensure adherence to guidelines and entered assigned puzzle grades into a spreadsheet; one faculty member also checked the completeness and accuracy of the data. All of the extra-credit crossword puzzles combined were worth an additional 1% of the final grade in the course. The crossword puzzle assignment was created in response to students having difficulty understanding the content in other assignments at the understanding, applying, analyzing, evaluating, and creating levels.

Statistical Analysis

Reliability was assessed with the parallel-forms procedure (Waltz et al., 2005) to determine if uploading a blank puzzle and corresponding answer key to the learning management site four days prior to the quiz due date was comparable to completing each other’s puzzles prior to taking the weekly quiz. \( P_o \) is the proportion of observed agreements in classifications in both puzzles (100% if submitted on time and correctly, and 0% if not submitted, not submitted on time, or not submitted correctly). \( K \) is the proportion of persons consistently classified in the same category with both puzzles beyond expected by chance. Kruskal-Willis was performed for the entire sample comparing weekly quiz scores between those who did versus did not complete the corresponding weekly puzzle correctly to determine the predictive validity of the puzzle scores on quiz scores or whether the observed quiz outcomes were associated with the crossword puzzle intervention. The same analyses were performed stratified by semester using Mann-Whitney U test with a Bonferonni correction (\( p < .0167 \)) to account for multiple analyses and reduce the risk of Type I error since context matters (Ogrinc et al., 2019), such as internal elements (modality: hybrid vs. 100% online), external elements (institution), and characteristics of the individuals (masters vs. doctoral students).

Results

The first semester comprised 14 students in a DNP program and 2 in a PhD program with all students completing the course. The second semester was a year after the first semester and comprised 20 students in a MSN program at a separate institution from the first semester with all students completing the course. The third semester immediately followed the second semester and comprised 53 students in the same MSN program as the previous semester, with 51 students completing the course. The total sample size was 87. The parallel-form reliability found \( P_o \) ranged from 0.75 to 1.00, while \( K \) ranged from -0.56 to -0.00 (see Table 1).
Table 1

Parallel Form Reliability of Creating versus Completing Crossword Puzzles

<table>
<thead>
<tr>
<th>Puzzle</th>
<th>Semester</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>P&lt;sub&gt;0&lt;/sub&gt;</td>
<td>1.00</td>
<td>0.94</td>
<td>0.88</td>
<td>0.88</td>
<td>0.94</td>
<td>0.88</td>
<td>0.94</td>
<td>0.88</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.94</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td>K</td>
<td>-0.56</td>
<td>-0.53</td>
<td>-0.50</td>
<td>-0.36</td>
<td>-0.34</td>
<td>-0.36</td>
<td>-0.25</td>
<td>-0.36</td>
<td>-0.02</td>
<td>-0.06</td>
<td>-0.02</td>
<td>-0.16</td>
<td>-0.06</td>
<td>-0.16</td>
<td></td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>P&lt;sub&gt;0&lt;/sub&gt;</td>
<td>0.95</td>
<td>1.00</td>
<td>1.00</td>
<td>0.95</td>
<td>0.95</td>
<td>0.75</td>
<td>0.85</td>
<td>0.90</td>
<td>0.75</td>
<td>0.90</td>
<td>0.85</td>
<td>0.80</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>K</td>
<td>-0.11</td>
<td>-0.49</td>
<td>-0.81</td>
<td>-0.47</td>
<td>-0.47</td>
<td>-0.35</td>
<td>-0.43</td>
<td>-0.34</td>
<td>-0.31</td>
<td>-0.34</td>
<td>-0.25</td>
<td>-0.20</td>
<td>-0.18</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>P&lt;sub&gt;0&lt;/sub&gt;</td>
<td>0.88</td>
<td>0.94</td>
<td>0.96</td>
<td>0.86</td>
<td>0.90</td>
<td>0.90</td>
<td>0.94</td>
<td>0.88</td>
<td>0.96</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
<td>0.90</td>
<td>0.94</td>
</tr>
<tr>
<td>K</td>
<td>-0.23</td>
<td>-0.24</td>
<td>-0.24</td>
<td>-0.14</td>
<td>-0.11</td>
<td>-0.09</td>
<td>-0.06</td>
<td>-0.13</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.00</td>
<td>-0.04</td>
<td>-0.10</td>
<td>-0.06</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen in Table 2, when all the semesters were combined, results from the Kruskal-Willis test found those who completed puzzles 5, 6, 7 and 13 correctly had statistically significantly higher scores on the corresponding quizzes compared to those who did not complete those puzzles. The same analyses were performed stratified by semester using the Mann-Whitney U test. In the first semester (n = 16), those who completed puzzle 13 correctly had significantly higher scores on the corresponding quiz compared to those who did not (97.6 vs. 90.0, p = .022), although this was not statistically significant with a Bonferroni correction. In the second semester (n = 20), those who completed puzzles 6 (93.3. vs. 80.3, p = .039) and 11 (92.5 vs. 82.0, p = .002) had significantly higher scores on the corresponding quiz, although only week 11 was statistically significant after Bonferroni corrections. In the third semester (n = 53), those who completed puzzle 7 had significantly higher scores on the corresponding quiz (92.9 vs. 89.0, p = .019), although this was not statistically significant with a Bonferroni correction.

Table 2

Mean Quiz Scores Based on Puzzle Completion

<table>
<thead>
<tr>
<th></th>
<th>Total (n = 87)</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Semester (n = 16)</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Semester (n = 20)</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; Semester (n = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Completed assignment</td>
<td>Did not complete assignment</td>
<td>Completed assignment</td>
<td>Did not complete assignment</td>
</tr>
<tr>
<td>Quiz 1</td>
<td>93.1</td>
<td>91.7</td>
<td>96.4</td>
<td>100</td>
</tr>
<tr>
<td>n = 51</td>
<td>n = 36</td>
<td>n = 14</td>
<td>n = 2</td>
<td>n = 6</td>
</tr>
<tr>
<td>Quiz</td>
<td>p-value</td>
<td>Quiz 2</td>
<td>Quiz 3</td>
<td>Quiz 4</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>98.5</td>
<td>93.4</td>
<td>95.7</td>
</tr>
<tr>
<td></td>
<td>.77</td>
<td>97.7</td>
<td>97.6</td>
<td>96.4</td>
</tr>
<tr>
<td>p-value</td>
<td>.64</td>
<td>100</td>
<td>91.7</td>
<td>95.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 65</td>
<td>n = 66</td>
<td>n = 49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 22</td>
<td>n = 13</td>
<td>n = 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 3</td>
<td>n = 4</td>
<td>n = 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 17</td>
<td>n = 19</td>
<td>n = 16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 35</td>
<td>n = 4</td>
<td>n = 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.93</td>
<td>100</td>
<td>95.8</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.90</td>
<td>100</td>
<td>90.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.78</td>
<td>.97</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>n = 35</td>
<td>n = 1</td>
</tr>
</tbody>
</table>
No changes were made in the puzzle assignment between semesters due to small sample sizes. There was no missing data. An unintended consequence was a dramatic increase in faculty workload when the number of students increased from 20 in the fall to 53 in the spring without an increase in the number of faculty in the course.

Discussion

Creating cumulative crossword puzzles and completing peers’ cumulative puzzles were associated with higher scores on most corresponding quizzes. These results are consistent with previous literature that found students who have cumulative assessments retain more information than students who have noncumulative assessments (Beagley & Capaldi, 2016; Khanna et al., 2013; Lawrence et al., 2013). A strength of the current quality improvement project is that it incorporated recommendations for students to create puzzles for their classmates to complete, encouraging students to research and understand the material in order to write meaningful clues for the puzzle (Davis et al., 2009).

The current results that found crossword puzzles were associated with higher scores on most corresponding quizzes coincide with previous studies that found crossword puzzles improve knowledge retention (Abuelo et al., 2016; Murphy et al., 2016; Nirmal et al., 2020; Orawiwatnakul, 2013; Patrick et al., 2018; Shawahna & Jaber, 2020; Singh Matreja et al., 2021; Zamani et al., 2021). However, there was not a statistically significant association between most puzzles and their corresponding quizzes. Similar results have been found where completing crossword puzzles was not always associated with improved knowledge retention (Sumanasekera et al., 2020), and student question generation was not associated with statistically significant improvement in achievements (Aflalo, 2018). In the current analysis, regardless of semester, most of the quiz scores were in the 80s and 90s, resulting in a ceiling effect or a reduction in the amount of upward change that is detectable (Polit & Beck, 2012). Our recommendation is to only incorporate crossword puzzles for those quizzes with habitually low scores, such as below 80%, for at least two semesters in the same course at the same institution. The current results move beyond students’ subjective perceptions of crossword puzzles and towards objectively assessing knowledge retention.
Reliability of the Crossword Puzzle Assignment

Results from the $P_0$ indicated a good level of agreement in classifying learners as 100% where both puzzles were submitted on time and were correct, versus 0% if not submitted, not submitted on time, or not correct. However, results from the $K$ indicated uploading a blank puzzle and corresponding answer key to the learning management site was not consistent with completing each other’s puzzles prior to taking the weekly quiz. There were several reasons for this discrepancy. Some created a puzzle but did not complete their peer’s puzzle even though both were required in order to earn full credit. Similarly, some completed their peer’s puzzle but did not create their own. Some did not upload a corresponding answer key. Some uploaded a puzzle from a previous week by mistake, and others did not upload their puzzles by the due date. Some provided a nonfunctional link to their puzzles. Thus, results of the criterion-referenced parallel-forms procedure indicate agreement in correctly completing both puzzles. However, not surprisingly, creating one’s own puzzles is not consistent with completing peers’ puzzles. The effectiveness of creating versus completing crossword puzzles should be examined further.

Strengths and Limitations of the Crossword Puzzle Assignment

A major strength of the current work is its addition to a body of literature objectively assessing knowledge retention resulting from crossword puzzles in a variety of disciplines such as English (Orawiwatnakul, 2013) and health professions (Abdulmajed et al., 2015), including pharmacology (Patrick et al., 2018), nursing (Shawahna & Jaber, 2020), dentistry (Nirmal et al., 2020), and sociology (Davis et al., 2009). To our knowledge, the current project is the first to examine crossword puzzles in graduate education and in a hybrid course.

There are several limitations. The results are limited to the students in the current quality improvement project. Efforts were made to minimize this limitation by including students across three semesters, across three programs (MSN, DNP and PhD), across two instructional modalities (online and hybrid), and across two institutions. In addition, this quality improvement project had a quasi-experimental design as there was an intervention without randomization, thereby threatening internal validity (Polit & Beck, 2012). Efforts were made to minimize this limitation by including a control group, i.e. those who did not complete the puzzle assignment. Finally, the sample size was small with deviations from normality necessitating non-parametric statistics, which are not as powerful as parametric statistics (Polit & Beck, 2012). Efforts were made to minimize this limitation by including a Bonferroni correction to adjust for multiple testing.

Conclusion

Students perceive crossword puzzles as an enjoyable, creative, and innovative way to learn (Abuelo et al., 2016; Franklin et al., 2003; Gaikwad & Tankhiwale, 2012; Singh Matreja et al., 2021; Sumanasekera et al., 2020; Zamani et al., 2021). Based on the current results, our recommendation is to only incorporate crossword
puzzles for those quizzes with habitually low scores for at least two semesters in the same course at the same institution. An important implication is the importance of moving beyond students’ subjective perceptions and towards objectively assessing knowledge retention. The current work should be replicated in additional disciplines beyond nursing; undergraduate and graduate education; and face-to-face, hybrid, and online courses.

Acknowledgements

I would like to gratefully acknowledge the Wisconsin Teaching Fellows and Scholars Program, especially Drs. Cyndi Kernahan and David Voelker for their guidance and feedback.

Conflicts of Interests

The authors declare that there are no conflicts of interest regarding the publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Clinical and Translational Science Award program through the NIH National Center for Advancing Translational Sciences [grant numbers UL1TR000427, KL2TR000428] and the Mississippi Center for Clinical and Translational Research [grant number 5U54GM115428]. The funding sources had no role in the project design, collection, analysis, interpretation of data, writing of the report, or the decision to submit for publication. The content is solely the responsibility of the author and does not necessarily represent the official views of the NIH.

References


https://www.aacnnursing.org/Portals/42/Publications/DNPEssentials.pdf


https://www.tandfonline.com/doi/abs/10.1080/10511970.2016.1194343

https://eric.ed.gov/?id=ED336049

https://eric.ed.gov/?id=EJ1011749

https://uncw.edu/jet/articles/vol9_3/davis.pdf


https://eric.ed.gov/?id=EJ673909


Journal of Effective Teaching in Higher Education, vol. 5, no. 1


Singh Matreja, P., Kaur, J., & Yadav, L. (2021). Acceptability of the use of...
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8273532/


https://journals.sagepub.com/doi/10.1177/1469787409355877

https://www.thefreelibrary.com/Measurement+in+Nursing+and+Health+Research%2C+3rd+ed-a0143215378