Effects of Writing-Related Contingencies on Both Quality of Writing and Multiple-Choice Exam Performance in Large College Courses

Katherine R. Krohn, Megan R. Parker, Lisa N. Foster, Kathleen B. Aspiranti, Daniel F. McCleary, & Robert L. Williams

Abstract

Students \((N = 158)\) in three sections of an undergraduate educational psychology course equivalent in content and assessment procedures completed five-min writing quizzes over assigned subject matter at the beginning of most class sessions. The study compared the effects of three separate writing contingencies on writing scores and multiple-choice exam performance. Overall, writing for daily credit and writing for random credit that could produce the same credit as the daily credit contingency had similar effects on both writing and multiple-choice exam scores, with both conditions producing better results than credit for only one randomly selected writing per unit. Correlations between writing and exam scores in all sections proved large to very large \((r_s = .57 \text{ to } .74)\).

Keyword descriptors: writing, multiple-choice, college, credit, contingencies

Most undergraduates likely take some courses that use multiple-choice exams as a major source of course credit. Instructors may be especially inclined to use multiple-choice exams in large courses because of ease and efficiency of grading (Hautau et al., 2006b). Nonetheless, many students report difficulty in taking multiple-choice exams, claiming that they could do better on essay exams. Consequently, discovering how to heighten student performance on multiple-choice exams is an important issue at the college level (Wallace & Williams, 2003). One possibility for improving multiple-choice exam scores may be the use of daily writing activities related to concepts included on the multiple-choice exams.

Past research demonstrates that brief essay quizzes may improve performance on a variety of exam formats (e.g., short answer, essay, fill-in-the-blank, and multiple-choice). Padilla-Walker (2006) found that brief, extra-credit daily quizzes on assigned reading material predicted performance on major exams (short-answer and essay) better than did gender, self-reported college GPA, and self-reported ACT scores. Daniel and Broida (2004) also reported that completing in-class quizzes over course concepts boosted performance on course exams (multiple-choice and short answer). Narloch, Garbin, and Turnage (2006) showed that prelecture quizzes, compared to no quizzes, produced better performance on both multiple-choice and essay exam items. Additionally, Leeming (2002) found that participating in daily 10-15 min writing activities on course concepts significantly improved performance on a comprehensive final exam that included short essay, fill-in-the-blank, and multiple-choice questions. Similarly, Turner et al. (2006) demonstrated that students required to complete a daily in-class writing activity performed better on the course’s multiple-choice exams than students without daily writing.

Although some research supports using daily essay quizzes to boost major exam performance, maximizing the impact of these quizzes is not without logistical challenges. For example, grading the quizzes could be labor intensive for the instructor. The writing activities may only consume a small percentage of class time, but the time required to grade and record the scores may detract from instructor time needed to organize and prepare for class. Leeming’s (2002) daily quizzes required about 10 to 20% of class time and an hour of instructor grading time each day. Hautau et al. (2006a) and Turner et al. (2006) reduced instructor time for grading quizzes by grading quizzes only on randomly selected days rather than on all days. The quizzes took 6 to 7% of total class time and required about 1 min per student for instructor grading. Given that students in these studies did not know what day’s quizzes would be
randomly selected for grading, the researchers expected the quizzes to have much the same impact on student performance as would daily grading and crediting of quizzes. However, the results proved mixed regarding this expectation, creating additional questions as to how best to maximize student performance on writing quizzes and multiple-choice exams without requiring an inordinate amount of instructor time.

Recent attempts to clarify the conditions under which daily writing could efficiently promote writing and exam scores have been reported by Hautau et al. (2006a; 2006b). Hautau et al. (2006a) required students to analyze in writing pairs of concepts that could be found within instructor notes available to students. The class website identified two to five pairs of concepts each day that students could be asked to address on the next quiz. At the following class period, the instructor randomly selected one of those pairs and instructed students to identify the concepts’ commonalities, differences, and the effect of one on the other. When the final writing activity had been completed for each unit, a student randomly chose one day’s writing activities to count for course credit. This randomized selection produced better writing performance than writing for no credit. Also, the randomized-credit condition led to better exam scores than did an equivalent oral discussion of issues addressed under the written-quiz arrangements.

The current study builds most directly on the Hautau et al. (2006b) study, which included three writing contingencies: writing for daily credit, writing for random credit, and writing for no credit. Daily credit entailed credit for each day’s writing activity; random credit entailed credit for only one randomly selected day each unit; and no credit involved no credit for any writing activity. The writing-for-daily-credit contingency produced the highest writing scores, followed by the writing-for-random-credit, and finally the writing-for-no-credit contingency. Plus, multiple-choice exam scores proved higher under the daily contingency than under either the random-credit or no-credit contingency. Although the study found writing for daily credit superior to writing for random credit, a question remained as to whether the difference resulted from a differential reinforcement schedule (continuous versus variable) or the different amounts of total credit possible. Students in the random-credit condition could earn only one-fourth the writing credit of students in the daily-credit condition.

Thus, the current study further examined the effectiveness of a random-credit contingency that could produce the same amount of writing credit as daily contingencies. The research evaluated the effects of three types of writing contingencies: (a) credit for every day’s writing quizzes, (b) credit for one randomly selected day’s writing quizzes multiplied by the number of writing quizzes included in each course unit, and (c) credit for only one randomly selected day’s writing quiz per unit. Mathematically, contingencies a and b permitted students to earn the same amount of credit per unit. In addition to evaluating the effects of these contingencies on writing and multiple-choice exam performance, we correlated writing and exam performance under each contingency to determine if writing performance was more predictive of exam performance under some contingencies than others.

Method

Participants

Students (N = 158) enrolled in three sections of the same undergraduate educational psychology course participated in the study. Approximately 25% of the students were males, and 75% were females. Academic classifications of the students were 4% freshmen, 41%
sophomores, 32% juniors, 17% seniors, 3% graduate students, and 3% unreported academic classification. The average reported GPA across sections was 3.21 (SD = 0.54).

General Procedures

All sections had exactly the same subject matter, assessment procedures, organization, and schedule. Also, graduate teaching assistants (GTAs) who had at least one year’s experience with the course taught all sections. All GTAs were trained and supervised by the same senior instructor. We also randomly assigned the three contingencies to the three intact sections. Thus, the principal differences between the sections, other than the randomly-assigned contingencies, were the specific instructors and students included in the various sections.

During the second class meeting, students completed a course pre-test consisting of 50 multiple-choice items similar in design and content to unit-exam items. The function of the pre-test was to determine students’ initial course knowledge, permitting the use of pre-test scores as a covariate to further equalize non-treatment characteristics of the sections. No course credit was attached to the completion of the course pre-test.

The course included five units: Unit A—Physical Development, Unit B—Cognitive Development, Unit C—Social Development, Unit D—Psychological Development, and Unit E—Values Development. Students had journal articles, instructor notes, a video, and study questions for each unit. The three sections met on Mondays, Wednesdays, and Fridays for 50-minute periods. Each unit consisted of seven class periods, with the same sequence of class events in each unit: the first day included a video related to a topic in the unit; the second through the fifth days included a five-minute quiz based on a pre-announced portion of the study questions related to the instructor notes; the sixth day involved a short essay quiz over a question from the reading materials and the scoring of a practice exam completed outside of class; and the last day of each unit consisted of a comprehensive, multiple-choice exam over the instructor notes, video, and reading materials for the unit.

The principal class activity targeted in the current study was the five-min quiz based on the instructor notes that occurred on days 2 through 5 in each unit. On these days, the instructor began class by projecting a randomly chosen quiz question via power-point. The 16 chosen questions for the course as a whole came from the students’ study-guide questions over the instructor notes (135 questions in all). Students knew in advance the range of questions from which the selected question would be taken each day. (See Appendix for samples of the quiz questions.) After five min of writing time, students submitted their written answers and the instructor then presented correct answers to the quiz questions via power-point. All answers had been prepared by the senior professor who supervised all sections of the course.

In addition to the power-point feedback presenting acceptable quiz answers, graduate teaching assistants (GTAs) graded the quizzes and gave individual written feedback for all quizzes the following class session (including quizzes not selected for credit). The quizzes were rated on a scale of 0-3: 0 = no credit (no accurate information provided), 1 = poor (some accurate information provided but some critical information omitted or misrepresented), 2 = good (all information provided is accurate but some critical information omitted), and 3 = excellent (the answer is complete and entirely accurate). The course syllabus described all rating criteria.

Graduate teaching assistants graded all quizzes for their own section and an additional 20% of the quizzes from one of the other two sections. Before any of the quizzes were graded, the GTA in each section (the primary rater) randomly chose 20% of the quizzes to be photocopied and given to the secondary rater. We used Unit-A quizzes to develop inter-rater reliability and, thus, did not include the
Unit-A data in the database for the study. The overall correlation between the ratings of the primary and secondary raters for Units B-E was .82.

**Dependent Variables**

*Quiz performance.* Past research on in-class writing activities indicated that an average of a student’s in-class writing scores was a better predictor of multiple-choice exam performance than was the total amount of points accumulated from the writing scores (Turner et al., 2006). For this reason, we used the average quiz rating per student for each unit as our measure of quiz performance. All quiz scores, not just those counted for credit, were included in the computation of average scores. However, missed quizzes were not included in the calculation of average scores.

*Exam performance.* Students took a 50-item, multiple-choice exam at the end of each unit in the course. A previous study using these exams reported high (.87) internal consistency for the exam items (Turner et al., 2006b). The exams included most of the concepts embedded in the study questions over the instructor notes in Units B-E. Approximately two-thirds of the exam questions were based on instructor-notes concepts, and the remaining items reflected concepts in reading materials and videotapes. Thus, the coverage represented in the exams was substantially greater than the coverage represented in the quiz questions.

**Experimental Contingencies**

The experimental conditions included the following credit contingencies for the four instructor-notes quizzes in each unit: daily credit (daily feedback and credit for all quizzes), daily quiz with random small credit (daily feedback but credit for only one randomly selected quiz), and daily quiz with random large credit (daily feedback and credit for one randomly selected quiz multiplied by the number of days quizzes were given that unit). On all days, a lead instructor (advanced GTA) and a first-year GTA present in the classroom checked and confirmed correct application of treatment logistics.

*Writing for daily credit (WDC).* In this section of the course, students completed a quiz on days 2 through 5 of each unit and received credit for each day’s quiz. Each quiz score was recorded on the course website’s grade-book accessible to students. If a student was not present for a quiz and had not made prior arrangements with the GTA to take the missed quiz another time, the student lost the possible quiz credit for that day. In this contingency, 12 credit points were possible for the instructor-notes quizzes in each unit.

*Writing for random small credit (WRSC).* Students in this condition also had quizzes on days 2 through 5 of each unit. However, instead of receiving credit for every quiz in each unit, students only received credit for one randomly selected day’s quiz per unit. On day 5 of each unit immediately after the quiz had been completed, the instructor asked a student to randomly choose one of four folded cards, each representing one of the four quiz days that unit. When the student had chosen a card, the student displayed it to show the class which quiz would count for course credit. Therefore, although all quizzes were scored, only one day’s quiz scores were posted per unit in the course grade-book. If a student was not present on the selected quiz day and had not made prior arrangements with the instructor, the student received no quiz credit for that unit. On the other hand, if a student had only attended the randomly selected day in the unit, the student received the quiz credit attained for that day. This contingency provided a potential of 3 points credit for the selected quiz day in each unit.

*Writing for random large credit (WRLC).* Students in this condition also had quizzes on days 2 through 5 of each unit. Instead of receiving credit for every quiz in each unit, students in this section received credit for one randomly selected day’s quiz multiplied by 4 (number of writing days in the unit),
creating a possible 12-points credit per unit for the quizzes (total credit equivalent to the potential credit in the daily-credit condition). The procedure for selecting the day to credit for the unit was the same as that used in the WRSC condition. If a student was not present on the selected quiz day and had not made prior arrangements with the instructor to make up missed work, the student received no quiz credit for that unit. On the other hand, if a student had only attended the randomly selected day in the unit, the student would receive the quiz credit attained for that day multiplied by 4.

Research Design

The pre-test score was used as a covariate for both the writing and exam scores. Thus, we used a one-way analysis of covariance to determine the effects of the three writing contingencies on writing and exam scores each semester. All writing and exam means reported in the study were adjusted for the effects of the pre-test covariate. In addition to comparisons of means across sections, writing scores were correlated with exam scores in each section of the course.

Results

Differences in Writing Scores across Contingencies

Figure 1 displays the adjusted writing means across contingencies and Units B-E. This figure shows that the WDC writing means remained high across all units, whereas the WRSC means remained lower than means for the other contingencies. The WRLC group started lower than the WDC group but improved to the same level as the latter group for the last 2 units in the course. A significant main effect was obtained for contingency across the combined units, $F(2, 152) = 25.62, p < .001$. Pairwise comparisons revealed no significant differences between adjusted writing means in the WDC (2.42) and WRLC (2.29) groups. However, significant differences were found between the WDC and WRSC ($m = 1.82$) groups ($p < .001$), as well as between the WRLC and WRSC groups ($p < .001$).

![Figure 1. Adjusted writing means across units.](image)

Differences in Exam Scores across Contingencies

Figure 2 indicates considerable variation in the pattern of adjusted exam means across units for the 3 contingencies. Nonetheless, WDC exam means appear higher than the other means for most units. Main effects were obtained for contingency, $F(2, 152) = 3.10, p < .05$, for the combined units. Pairwise comparisons revealed a significant difference ($p < .05$) between the WDC ($m = 41.71$) and WRSC ($m =$...
39.57) exam means. No significant differences were found between the WDC and WRLC (\(m = 40.10\)) means, or between the WRLC and WRSC means.

![Figure 2. Adjusted exam means across units.](image)

**Correlations between Writing Scores and Exam Performance**

Pearson product-moment correlations were computed to determine how well writing scores predicted exam performance. All correlations were in the moderate to strong range. The strongest correlation for the combined units was obtained in the WRSC condition \(r = .74\), closely followed by the WRLC condition \(r = .70\), and then by the WDC condition \(r = .57\). This pattern generally persisted across units with the exception of Unit D, where the magnitude of the correlations converged across contingencies (see Table 1).

**Table 1**

*Correlations between Writing Scores and Exam Performance across Contingencies and Course Units*

<table>
<thead>
<tr>
<th>Course Units</th>
<th>Unit B</th>
<th>Unit C</th>
<th>Unit D</th>
<th>Unit E</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contingency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WDC</td>
<td>.23</td>
<td>.24</td>
<td>.43**</td>
<td>.29*</td>
<td>.57**</td>
</tr>
<tr>
<td>WRLC</td>
<td>.42**</td>
<td>.60**</td>
<td>.41**</td>
<td>.39**</td>
<td>.70**</td>
</tr>
<tr>
<td>WRSC</td>
<td>.66**</td>
<td>.62**</td>
<td>.42**</td>
<td>.54**</td>
<td>.74**</td>
</tr>
</tbody>
</table>

Note. **Correlations significant at the .01 level. *Correlations significant at the .05 level.
Discussion

The primary purpose of this study was to determine if randomly selecting a few daily quizzes to credit would improve exam performance to the same degree as daily quizzes with all counting toward course credit. Hautau et al. (2006b) previously concluded that writing for daily credit was superior to writing for random credit; however, this difference might have been attributable to the different reinforcement schedules or the different amounts of potential credit under the two contingencies. Current findings support the latter possibility: random writing credit can lead to exam performance statistically comparable to daily writing credit when the potential credit is equivalent under the two contingencies. When credit on an intermittent schedule is sufficiently large, students will likely deem the quizzes to be important, keep up with assigned readings, and study for the quizzes, which will consequently contribute to exam performance.

Differential Effectiveness of Random Credit

Hautau et al. (2006b) showed that even a small amount of credit for quizzes leads to better exam performance than no credit awarded for quizzes. Therefore, although the students likely benefited from the WRSC condition, perhaps students worked even harder to prepare for the quizzes in the WRLC condition because each quiz potentially had greater impact on their grade. On the other hand, the weakest contingency (WRSC) was the most predictive of exam scores; whereas the strongest contingency (WDC) was the least predictive of exam scores even though all writing products counted toward credit. One might speculate that students who maximized the small amount of credit available for quiz performance sometime during each unit (WRSC) would also prepare well for the large amount of credit available on the unit exams; whereas more students in the WDC condition may have invested considerable time in preparing for the quizzes to the possible detriment of time necessary to prepare for the unit exams. Thus, quizzes may become less predictive of exam scores as the regularity of quiz credit increases.

A data trend worth noting in the current study is that students’ performance on writing activities in the WRLC condition improved toward the end of the course. In fact, WRLC writing scores were approximately equivalent to WDC writing scores in the last two units. One possible explanation for this improvement is that students in the WRLC condition realized that much more preparation was necessary for the writing activities after having received a low score multiplied by 4 in previous units. Perhaps students were not fully aware early in the course of the magnified impact low writing scores would have on their course grade. Paradoxically, improved writing did not lead to improved exam scores. Exam scores in the WRLC condition did not increase to the same level as WDC exam scores in the last two units. Again, greater investment in preparing for the daily quizzes may have been at the expense of exam-preparation time.

Overall, the WDC condition did not produce significantly better performance on either the quizzes or unit exams than did the WRLC condition, whereas both conditions proved better than the WRSC condition in promoting quiz performance. This finding is valuable because instructors can improve the quality of student writing by selecting only a few quizzes to grade for each student, so long as the potential credit for each quiz is equivalent to daily writing credit for a unit.

Advantages and Disadvantages of Random Credit

Certain aspects of the randomized writing contingencies used in this study might have unduly
The current system penalized students who were absent on the day randomly selected for writing credit. Students present for every writing day except the one day randomly selected for credit obtained no writing credit for the unit. The penalty was especially heavy for students absent on the writing day when that day’s writing credit was multiplied by the number of writing days in the unit, which resulted in these students’ receiving no writing credit for the unit. Another potential inequity with the randomized large credit contingency was that students present only on the randomly selected writing day in a unit had their credit multiplied by the number of writing days in the unit, exactly the same as for students who had been present every writing day in the unit. Of course, one way for students to avoid potential inequities under the randomized contingencies was to be present each day in a unit.

Future research should compare the effects of other types of random credit that may be more equitable, not disadvantaging some students or excessively rewarding other students. Staying within the framework of grading only one quiz per unit, instructors might explore other ways to increase the potential credit of a randomly selected quiz, such as multiplying students’ writing score by the number of days they actually wrote each unit rather than by the number of potential writing days in that unit. This arrangement would preclude extending students’ writing credit to days they had not actually written. If students were absent for legitimate reasons (e.g., illness, school trip) on the randomly selected writing day and had notified the instructor in advance of their absence, a day could be randomly selected for credit from the days they wrote and then multiplied by the number of days they wrote in the unit.

**Alternative Explanations of the Findings**

Few studies are so tightly designed that extraneous variables can be totally ruled out as potential contributors to outcome differences. In the current study, one potential extraneous variable was the possible difference in the effectiveness of the three instructors who taught the treatment sections. Even though the treatments were randomly assigned to course sections equivalent in most ways other than the treatment, the efficacy of the treatments could have been affected by differential effectiveness of the instructors. All three lead instructors were advanced GTAs in the course and supervised by the same senior professor, but their in-class instructional behavior was not systematically monitored throughout the study.

The absence of systematic observational data on instructor behavior leaves open the possibility that instructor differences might have affected treatment outcomes. Possible differences in instructor behavior included (1) the types of questions posed to students, such as factual versus higher order questions, (2) the clarity of instructor explanations, and (3) the degree of similarity between information discussed in class and identifiable exam content. In particular, the type of questions asked by instructors merits attention due to the complex nature of its potential effects. Higher order questions likely encourage deeper processing of information, enhanced student engagement in class discussion, and greater perception of content relevance in comparison to strictly factual questions. Although an emphasis on factual questions likely detracts from student understanding and retention of the course material, an over-reliance on factual questions reflecting information linked to exam questions could have led to elevated exam performance. Despite the possibility of instructor differences, no anecdotal evidence suggested that the instructors were performing differently on any of the three dimensions noted above.

If indeed instructors differed in their effectiveness in facilitating class discussion, those differences would have only affected exam scores. The questions students answered in the writing activities had never been discussed in class prior to students writing about them. Students in all sections had the same access to potential quiz questions and the instructor notes on which those questions were based. On the other hand, exam scores likely were affected by subsequent class discussion of writing questions and other course issues. Consequently, future research should include direct observation of
instructor behaviors. Specifically, if instructors assume a lecture approach to teaching, the depth and breadth of the content coverage would need to be systematically assessed. In courses that involve an interactive, discussion-based approach, the types of instructor questions should be categorized, perhaps as previously suggested (factual versus higher-order). The observer could subsequently provide the instructors with feedback on the quantity of questions falling in each category to reduce potential discrepancies between instructor questions. To eliminate instructional style almost entirely, the current study could be replicated in three separate classrooms all taught by the same instructor.

We presume that the quality of student answers to quiz questions was affected by student preparation for class. Otherwise, what would account for students’ doing better writing under some writing contingencies than others? Although student scores on the quizzes indirectly indicated the quality and extent of student preparation for class, we have no independent measures of how much time students devoted to study out of class or what study strategies they might have used. Presumably, those who did best on the writing quizzes prepared written answers to all potential quiz questions prior to coming to class and then rehearsed those answers in some fashion shortly before class. We could have asked students to keep a log of how much time they spent studying for the quizzes outside of class and what study strategies they used in preparing for the quizzes. Nonetheless, even this approach would not have insured the accuracy of students’ homework report.

Despite limitations in the study because of failure to collect some potentially illuminating data regarding related variables, we believe the findings point to the possibility of improving exam scores by including brief writing assignments in class. Although daily writing activities could be labor intensive to score, the efficacy of writing assignments can be preserved by having students write frequently but then randomly choosing only certain days for grading and crediting the writing products. Thus, we believe that the findings substantially support the claim that the combination of frequent writing and random selection of a few days to grade writing products can be both an effective and efficient way to maximize the quality of student writing and its potential linkage to multiple-choice exam scores.

References


**Appendix, Next Page!**
Appendix

Sample Questions of Daily Writing Activities

Unit A—Physical Development
What are the major differences in girls’ and boys’ patterns of drug use?
What have been the effects of health and physical education in the schools?
What are the likely effects of yo-yo dieting on body composition and weight change?

Unit B—Cognitive Development
Elaborate on the educational implications of Piaget’s theory.
What is the relationship between age and synaptic development in the brain?
What possible testing accommodations are legally required for students with officially diagnosed learning disabilities?

Unit C—Social Development
Describe the logistical specifics and purpose of Classwide Peer Tutoring.
Describe the relationship of social skills to age.
Identify the major elements of an authoritative parenting style.

Unit D—Psychological Development
What do the concepts of locus of control and self-efficacy have in common, how are they different, and how are they linked?
What do positive and negative reinforcement have in common and how are they different?
Contrast the following techniques for weakening behavior: punishment, extinction, and response cost.

Unit E—Moral Development
Describe the four conditions employed in Albert Bandura’s laboratory research on the modeling of aggression and the principal findings of the research.
What percentage of college students have observed others cheating, reported observed instances of cheating, and have been caught cheating by the instructor?
Describe the nature and the effects of college students’ beliefs regarding the sustainability of the world’s natural resources.
Author Contact Information

Katherine R. Krohn, B. A.
Department of Educational Psychology and Counseling
The University of Tennessee
Knoxville, TN 37996-3452
e-mail kkrohn@utk.edu
phone 865-974-4169
fax 865-974-0135

Megan R. Parker, B. A.
Department of Educational Psychology and Counseling
The University of Tennessee
Knoxville, TN 37996-3452
e-mail mparker9@utk.edu
phone 865-974-4169
fax 865-974-0135

Lisa N. Foster, B. A.
Department of Educational Psychology and Counseling
The University of Tennessee
Knoxville, TN 37996-3452
e-mail lisafoster11@gmail.com
phone 865-974-4169
fax 865-974-0135

Kathleen B. Aspiranti, B. A.
Department of Educational Psychology and Counseling
The University of Tennessee
Knoxville, TN 37996-3452
e-mail kaspir82@utk.edu
phone 865-974-4169
fax 865-974-0135

Daniel F. McCleary, B. A.
Department of Educational Psychology and Counseling
The University of Tennessee
Knoxville, TN 37996-3452
e-mail dmcclear@utk.edu
phone 865-974-4169
fax 865-974-0135

Corresponding Author:
Robert L. Williams, Ph. D.
Department of Educational Psychology and Counseling
The University of Tennessee
Knoxville, TN 37996-3452
e-mail bobwilliams@utk.edu
phone 865-974-6625
fax 865-974-0135