International Journal for the Scholarship of Teaching and Learning

Volume 11 | Number 2 Article 3

July 2017

Effects of Quizzing Methodology on Student Outcomes: Reading Compliance, Retention, and Perceptions

Carey Bernini Dowling
University of Mississippi Main Campus, cbdowlin@olemiss.edu

Recommended Citation
Dowling, Carey Bernini (2017) "Effects of Quizzing Methodology on Student Outcomes: Reading Compliance, Retention, and Perceptions," *International Journal for the Scholarship of Teaching and Learning*: Vol. 11: No. 2, Article 3.
Available at: https://doi.org/10.20429/ijsotl.2017.110203

Effects of Quizzing Methodology on Student Outcomes: Reading Compliance, Retention, and Perceptions

Abstract
This study set out to replicate and extend research on students' reading compliance and examine the impact of daily quizzing methodology on students' reading compliance and retention. 98 students in two sections of Abnormal Psychology participated (mean age = 21.5, *SD* = 3.35; 72.4% Caucasian). Using a multiple baseline quasi-experimental design the daily quizzing methodology was changed at different points in the semester from Clicker questions to Clicker questions plus random written quizzes. The classes did not differ significantly on predictors of success and only differed significantly on one demographic variable. 77.6% of students failed Sappington et al.'s (2002) objective measure of reading compliance and the majority lied about their reading compliance. There was mixed evidence for the impact of quizzing methodology on learning outcomes. Daily quizzing appears to be effective, but adding written quizzes may not improve learning outcomes enough to justify increased grading time.

Keywords
Quizzing methods, reading compliance

Cover Page Footnote
I would like to express special thanks to the research assistants involved with this project: Billy Rush, Teresa Davis, Mia Kloth, and Brittney Stone. I would also like to express sincere appreciation to Conor Dowling and the reviewers for their feedback on previous versions of this manuscript.

Effects of quizzing methodology on student outcomes: Reading compliance, retention, and perceptions

Carey Bernini Dowling

Department of Psychology, University of Mississippi, University, MS 38677, USA

(Received 29 October 2015; Accepted 28 November 2016)

This study set out to replicate and extend research on students' reading compliance and examine the impact of daily quizzing methodology on students' reading compliance and retention. 98 students in two sections of Abnormal Psychology participated (mean age = 21.5, *SD* = 3.35; 72.4% Caucasian). Using a multiple baseline quasi-experimental design the daily quizzing methodology was changed at different points in the semester from Clicker questions to Clicker questions plus random written quizzes. The classes did not differ significantly on predictors of success and only differed significantly on one demographic variable. 77.6% of students failed Sappington et al.'s (2002) objective measure of reading compliance and the majority lied about their reading compliance. There was mixed evidence for the impact of quizzing methodology on learning outcomes. Daily quizzing appears to be effective, but adding written quizzes may not improve learning outcomes enough to justify increased grading time.

INTRODUCTION

An undergraduate college education in psychology has multiple desired learning goals (APA, 2012). In order for students to meet these goals, it is necessary for them to actively participate in their education. As educators who desire to help students succeed in college we must understand what predicts their success and what we can do to help them succeed.

One of the first ways students can actively participate in their education is to prepare for their classes by completing reading assignments. Research suggests this preparation is important because it is associated with overall class performance (Sappington, Kinsey, & Munsayac, 2002) and students report lack of preparation for class is a barrier to their class participation (Karp & Yoels, 1976). However, recent research suggests that a majority of college students do not complete reading assignments prior to coming to class (Burchfield & Sappington, 2000; Clump, Bauer, & Bradley, 2004; Connor-Greene, 2000; Sappington et al., 2002). Sappington et al. (2002) found only 22% of students passed their objective measure of reading compliance. Unfortunately, this trend of lack of preparation for class might be increasing (Burchfield & Sappington, 2000). Yet, it is possible that students' reading compliance varies by the testing schedule of the course, with students reporting they are more prepared for classes with daily quizzing than classes with exams only (Connor-Greene, 2000).

If students' reading compliance is declining and consistently at levels below 30%, it is important to determine effective strategies for increasing and maintaining student reading compliance across the semester. Multiple strategies have been implemented to increase student reading compliance and course performance, such as completion of out-of-class assignments that require reading (Carkenord, 1994; Ryan, 2006), daily written quizzes (Connor-Greene, 2000), and randomized reading quizzes (Ruscio, 2001).

Although reading is not required to complete in-class quizzes, quizzes may be an effective means of improving reading compliance (Connor-Green, 2000; Ruscio, 2001) while also improving course performance. Quizzing has been found to positively impact exam grades when done in a manner to simulate basic research on the testing effect (see Nguyen & McDaniel, 2015). Research on the testing effect suggests that testing itself and testing with feedback are powerful means to improve the learning of material (Butler, Karpicke, & Roediger, 2008; Roediger, Agarwal, McDaniel, & McDermott, 2011; Roediger & Karpicke, 2006). Immediate feedback after testing allows the learner to correct erroneous knowledge as well as correct metacognitive errors regarding low confidence in correct answers (Butler et al., 2008). Therefore, it is not surprising that previous research has found utilizing student response systems (SRS) during class to quiz and provide immediate feedback to students improves students' course and examination performance (Brady, Seli, & Rosenthal, 2013; Hall, Collier, Thomas, & Hilgers, 2005; Morling, McAuliffe, Cohen, & DiLorenzo, 2008) and increases course engagement and motivation (Hall et al., 2005).

Although SRS and written quizzing have shown positive benefits for students, these methods are not without concerns. First, there are multiple time demands on professors that may make grading of written quizzes impractical, especially in large sections. Additionally, multiple time demands are a large source of stress for faculty (Gmelch, Lovrich, & Wilke, 1984), so it is especially important to examine if assessments that require grading confer enough of a benefit to justify the grading time. Second, while utilizing SRS during class reduces (or eliminates) grading time, it is easier for students to guess the correct answer even if they have not read the material, thus potentially reinforcing students who did not read and perpetuating their perception that they can succeed without coming to class prepared. A third concern with utilizing both forms of quizzing has to do with potential negative ramifications on student evaluations. Individuals responsible for evaluating teaching effectiveness rate student evaluation scores and written comments among the top three most important measures to use for evaluating teaching effectiveness (Shao, Anderson, & Newsome, 2007). Thus, it is pragmatic for professors to be concerned about poor student evaluations.

Given the multiple time demands for professors as well as concerns over poor student evaluations, it is beneficial for professors to determine the best methods to simultaneously achieve multiple goals (encouraging students' reading compliance, engagement with the material, and learning of the material; avoiding an unduly difficult grading load; and avoiding unfavorable student evaluations). Therefore, I set out to determine whether a combination of the use of SRS with pop written quizzes would achieve all of these goals. This study utilized daily SRS quizzes, which require minimal

https://doi.org/10.20429/ijsotl.2017.110203

grading time, and daily SRS quizzes plus written quizzes in only 25% of class sessions, which increase grading load but not excessively.

Limited research on the impact of active learning strategies has included the impact on students' reading behaviors (Carkenord, 1994; Connor-Greene, 2000; Morling et al., 2008; Ryan, 2006). I am unaware of any studies that report the impact of quizzing on objective measures of whether, and how thoroughly, students read on a daily basis. Given the theoretical importance of reading assigned readings on time and the evidence that suggests it improves class participation and performance (Karp & Yoels, 1976; Sappington et al., 2002) it is important to determine whether active learning strategies, such as quizzing, also impact how frequently students read on time and how thoroughly they read assigned readings.

It is possible that previous research has rarely reported student reading behaviors because student reports of reading are likely to be invalid (see Sappington et al., 2002). Thus, the first aim of the study was to explore the validity of students' self-reports regarding how thoroughly they read assigned readings to determine if they could be a valid dependent variable. Sappington et al. (2002) utilized an objective measure of student reading compliance utilizing a dichotomous "yes/no" option for students to report whether they read the entire syllabus. Thus, students who had skimmed the entire syllabus or read most of the syllabus were forced to decide whether they felt what they did counted as "reading the entire syllabus" and potentially increased the chances of students engaging in self-enhancement bias. Thus, I set out to determine if when students were given multiple options regarding their reading compliance allowing them to choose options such as "read all, read at least some, skimmed all, did not look at any," whether they might show less self-enhancement bias and more valid responses. I hypothesized that, similar to Sappington et al. (2002), students would show evidence of a self-enhancement bias and a majority would lie on their self-reported reading compliance, but that students who failed the objective measure of reading the entire syllabus would report lower levels of reading compliance than students who passed the objective measure.

Previous research on predictors of students' success in college classes has found that students' performance goals (Elliot & Church, 1997; Elliot & Murayama, 2008), intrinsic motivation (Clark, Middleton, Nguyen, & Zwick, 2014), conscientiousness, intelligence, and SAT performance (Conard, 2006; Kappe & van der Flier, 2012), all predict students' success in college classes. However, I am unaware of any studies examining the effectiveness of teaching methodology that also examine whether groups of students in experimental conditions vary significantly on these or other variables of potential importance such as intrinsic motivation for the course and expectations of competence in the course. Therefore, the second aim of the present study is to determine if the students in each quasi-experimental condition differed significantly on any of these other potential predictors of students' success in college classes. Students registered for classes independently, but it was assumed that there would not be significant differences between students who chose to sign up for an 8 AM versus 9 AM section meeting on the same days taught by the same professor.

The third, and final, aim of this study was to examine student engagement with the material and class, students' understanding and retention of the material, and student evaluations of the methods and the class overall. A mixed within-groups, between-groups design was utilized to be able to differentiate potential differences that may occur naturally across the semester from those that may have occurred as a result of different quizzing methodologies. The first quizzing method was daily SRS reading comprehension quizzes, which required an upfront time commitment to create the questions but no grading time (referred to after this as "Clicker only"). The second method (implemented on a staggered timeline in each section) was continuing the daily clicker reading comprehension quizzes and adding a policy that during 25% of the remaining class sessions their quiz grade would be based on a written quiz instead of their clicker answers for that day (referred to after this as "Clicker plus written"). It was hypothesized that students would report reading more when their class was quizzed using the Clicker plus written method than with the Clicker only method. In order to compare students' retention of the material taught while students were quizzed with different quizzing methodologies, class sections were compared on identical assessments of the material. It was hypothesized the section who had learned the material covered when being quizzed using the Clicker plus written method would have higher grades than the section being quizzed with the Clicker only method on assessments of that portion of material.

METHOD

Participants

All students enrolled in my Spring 2014 Abnormal Psychology sections (taught at 8 AM and 9 AM MWF) were recruited for this study (56 students per section at the start of the semester). Although 101 students originally consented to participate in the study, only 98 of those students completed the course. Of the 52 students who consented in the 8 AM section, 49 completed the pre-packet and 51 completed the post-packet. Of the 46 students who consented in the 9 AM section, 45 completed the pre-packet and 42 completed the post-packet.

Participants mean age was 21.5 years (SD = 3.35), they were primarily juniors (46.8%) or seniors in college (40.4%), Caucasian (72.4%), single (96.6%), living off campus (75.5%), and either a psychology major (23.7%) or minor (47.4%). The majority (55%) reported their fathers obtained a bachelors' degree or higher and 44.2% reported the same for their mothers. Participants' mean self-reported high school GPA was a 3.69 (SD = .40) and self-reported current GPA was 3.20 (SD = .54). Based on the results of Chi-square tests for independence and independent t-tests, the only participant demographic characteristic that varied significantly by section was the percent living in each setting (on-campus vs. off-campus), X^2 (4, n = 94) = 11.16, p = .03; with 82.7% living off campus in the 8 AM section and 60.9% living off campus in the 9 AM section. (Percent of students with mothers and fathers who completed a college education approached significance, p < .10, with higher percentages in the 9 AM section).

Procedure

During the fourth class, I explained the study and protections put in place to reduce the possibility of coercion. Following the announcement, students were given the consent form and a pre-numbered assessment packet that was linked to their name in a password-protected file. Students were asked to read the consent form, complete the questionnaires if they consented to participate in the study, and then place their packet (complete or incomplete) in a manila envelope. I then left the room and the envelope was sealed by the research assistant and was kept sealed until after final grades had been submitted. Utilizing the password protected file and the manila envelope was done to protect students and keep me blind to percentages participating in each course until the end of data collection. Students who were absent were given individual envelopes with a copy of the consent form and the packet and asked to return the envelope sealed with the packet completed if they consented and blank if they did not.

To replicate and extend Sappington et al.'s (2002) results, the following line appeared near the bottom of the syllabus: "Students who have read this far in the syllabus will receive 1 point added to their final average if they e-mail me at […] with the subject line: 'Psy 311, Section 3 Syllabus bonus' by the time Homework #1 is due." Students were asked to report their compliance with reading the syllabus first at the end of homework #1 and again using their clickers at the start of class reviewing the syllabus/homework #1. Students were asked for the e-mail address they use the most on their homework assignment to verify they had access to an e-mail address.

All students experienced the course as if the study was not being conducted. However, the quizzing method was modified according to a multiple-baseline quasi-experimental design (the 9 AM section was randomly chosen prior to the semester to receive the manipulation first). At the start of the semester both sections participated in daily multiple-choice clicker quizzes (referred to as the Clicker only method). Following Exam 1, the 9 AM section continued daily clicker quizzes but now had a 25% chance of having a written quiz to start the class, which would replace their clicker quiz points for the day (referred to as the Clicker plus written method)[1]. Halfway through new material coverage for Exam 3, the 8 AM section also began being quizzed with the Clicker plus written method. The class lectures and clicker questions were identical with the exception of variations in student responses to my questions and student questions prompting varying responses from me.

All students were given the post-assessment packet with their ID number at the start of the final examination and asked to complete the packet of questionnaires (if participating) at the end of the final exam along with a course evaluation (which was not part of this study), thereby ensuring that I was unaware if they were completing the final, the post-assessment packet, or the course evaluation. All students were told to place the post-assessment packet in a manila envelope regardless of whether they completed it or not and I sealed the envelope at the end of the final exam period.

Students completing the pre- and post-packets were entered into a raffle for one of two $10 Amazon gift cards.

Quizzes. Students' quiz grades accounted for 15% of their final grade in the course. Throughout the semester for both sections, every non-exam class day included five clicker multiple-choice quiz questions embedded in the class plan for the day. Clicker questions were primarily designed to test their reading compliance and open class discussion of a topic while simultaneously providing feedback to the students and I. Immediately following each question a histogram of class responses appeared and then a correct answer indicator appeared on the slide. I utilized the immediate feedback to adjust the depth of coverage needed on a topic. For instance, following a question such as "Which of the following is definitely present in ALL anxiety disorders?", if more than 80% of the students answered correctly, I would quickly review the answer, explain (or ask students to explain) why the other answers were incorrect, when appropriate give a brief lecture related to that topic, and move on to the next planned topic or activity. If very few students got the answer correct and it was a key point, I would do all of the same things but would review the answer and concept in more detail before moving on. Clicker questions were also designed to assess comprehension of the assigned reading (e.g., "Conor has a diagnosis of a Specific Phobia, which of the following is NOT a possible trigger?"). The relevant concepts addressed by the question would then be reviewed in class. Depending on length of the clicker questions, clicker quizzes took five to ten minutes to administer throughout the class.

When random written quizzes were added to the class design, they were also designed to test reading compliance (e.g., "Who was the case study in the reading about?") and assess comprehension of the assigned reading (e.g., "What is one of the effective treatments for substance-use disorders that arises from the psychological model?"). Students were given approximately ten minutes to complete the written quizzes. Students received feedback on written quizzes the following class but answers were not reviewed because the material had been covered immediately after the written quiz.

Quiz questions (both types) were intentionally basic, recognition questions designed to be difficult to answer without reading the assigned reading but not so difficult they required students to do more than read the entire assigned reading actively prior to class. I occasionally used clicker quiz questions that resembled exam questions for practice, but only after we had reviewed the relevant concept in class. Thus, exam questions were not directly tied to quiz questions, but they did resemble class activities. Exam questions were designed to primarily evaluate students' understanding of the material and ability to apply knowledge gained (e.g., by correctly stating which disorder a vignette most closely characterized). A portion of exam questions also assessed recall and recognition of important facts.

Measures

Demographic questionnaire. The demographic questionnaire created for this study requested information on major demographic characteristics such as participants' age and year in school.

Predictors of Student Success. Four questionnaires were given to students to evaluate whether the sections differed significantly on variables found in previous studies to significantly predict student success. First, the Achievement Goal Questionnaire-Revised (AGQ-R; Elliot & Murayama, 2008) was included in the pre-assessment packet to measure mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance goals (Cronbach's alphas for this study were .85, .80, .86, and .71, respectively.) Second, the average score of two items by Elliot and Church (1997) were utilized to measure how well students expected to do in the course at pre-assessment (Cronbach's α = .98). Third, eight items from Elliot and Church (1997) were utilized to measure students' intrinsic motivation for the course at pre- and post-assessment. The items were revised to be future tense for the pre-assessment and past tense for the post-assessment (Cronbach's

https://doi.org/10.20429/ijsotl.2017.110203

α = .93 pre and .91 post). Fourth, the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003) was included in the post-assessment to measure students' Big 5 personality characteristics: Conscientiousness, Agreeableness, Emotional Stability, Extraversion, and Openness to new experiences (Cronbach's alphas for this study were .51, .37, .70, .76, and .46, respectively).

Global Reading Behavior. Students self-reported global reading behaviors were examined using two questions written by Connor-Greene (2000). The first item measures the frequency with which students report they read the assigned reading by the due date ranging from 1 (*always*) to 5 (*never*). The second item measures when they typically began reading the assignments with response options such as *before class, several days before test*, and *did not read*. The items were modified slightly to ask about reading behavior in previous classes on the pre-assessment and in the present course on the post-assessment.

Evaluation of strategies and course. Six items (based off similar items used by Connor-Greene, 2000) were written for the post-assessment to evaluate students' perceptions of the two quizzing methods and the impact of the methods on their behaviors. The measure evaluated how much they read, time spent preparing for classes, thoroughness of reading, class participation, and understanding of course material by quizzing method as well as which quizzing method they believed would lead to the most learning.

Objective syllabus reading check. To obtain an objective measure of whether students read the syllabus, they were expected to follow the directions embedded in the syllabus regarding e-mailing me for a 1% bonus on their final grade. Students who sent the e-mail by the time Homework #1 was due were coded *passed*, and all others coded as *failed*.

Self-reported syllabus reading level responses 1 and 2. Students were asked twice to respond to the multiple choice prompt: "I ___ of the syllabus. (a) read all, (b) read almost all, (c) read at least some, (d) skimmed all, (e) skimmed some, (f) skimmed headers, (g) did not look at any". Response #1 occurred at the end of the first homework assignment (completed online) and response #2 occurred at the beginning of the class reviewing the syllabus/Homework #1 (completed using their clickers in-class).

Daily clicker reading checks. At the start of each class students were asked to respond utilizing their clicker to the multiple choice item: "I ___ of the assigned reading." with the same response options as the self-reported syllabus reading level responses.

Retention assessments. Exam grades across the semester were utilized to assess student retention of the material. Exam 1 and the first segment of the Final Exam assessed material covered when both sections were quizzed using the Clicker only method. Exam 2, the first half of Exam 3, and the second segment of the Final Exam assessed material covered when only the 9 AM section was quizzed using the Clicker plus written method. The second half of Exam 3, Exam 4, and the third segment of the Final Exam assessed material covered when both sections were quizzed using the Clicker plus written method. Exams were comprised of 60% multiple choice questions and 40% short-answer or fill-in-the-blank questions. Exams were designed to primarily assess mastery and comprehension of the material with some questions assessing recognition and recall.

RESULTS

Assessments of Reading Compliance and Behavior

Syllabus reading. Of the 98 students participating in this study, only 22 passed the objective syllabus reading check (22.4%). However, all students completing the first homework assignment provided a valid e-mail address, thus verifying they were able to send an e-mail if they read the entire syllabus carefully and were motivated to receive the 1% bonus on their final grade.

Not all students completed each of the self-reported syllabus reading level responses, so the remaining results are based only on those who responded. The majority of students who passed the objective syllabus reading check reported on both their homework (90.5%) and in-class (100%) that they read all of the syllabus. However, the majority of students who failed the objective syllabus reading check also reported on both their homework (81.8%) and in-class (70.1%) that they read all of the syllabus.

For the first self-reported syllabus reading level check, there was not a significant difference in students' self-reported reading levels by group (passed/failed objective syllabus reading check), X^2 (2, N = 87) = 2.51, p = .29. However, for the second self-reported syllabus reading level check, there was a significant one-tailed, medium-sized effect of group on students' self-reported reading levels, X^2 (2, N = 88) = 8.11, p = .05, Cramer's φ = .30, such that 100% of the students who passed the objective syllabus reading check reported they "read all" of the syllabus whereas only 70.1% of those who failed did (the remaining students chose four other options).

Exploratory analyses of daily clicker reading check. Following the analysis indicating a high level of dishonesty on the syllabus reading level checks, exploratory analyses were conducted to determine if the intended dependent variable, daily clicker reading checks, could be trusted as valid. This was done because the daily clicker reading check was an identical clicker reading check question but was given at the start of all classes with a quiz.

The correlation between students' self-reported syllabus reading level responses 1 and 2 was significant (r = .29, p = .008). Further analyses showed that 27.1% of students who failed the objective syllabus reading check were inconsistent in their responses while 10% of students who passed were inconsistent in their responses. Due to the high rates of inaccurate reporting for students who failed the objective syllabus reading check, no analyses were conducted on their daily self-reported reading levels during the semester.

Global reading behavior. Although students' self-report on the global reading behavior questions may also be invalid, it is interesting to note their responses. Wilcoxon signed ranks tests were utilized to compare students' reports of their global reading behavior from previous classes to the present class. Students' reports of how often they completed the reading assignments by the assigned due date did not differ significantly from previous classes to the present class, X = -1.53, p = .13; the majority reported "always" or "almost always" (77.7% in previous classes and 67.7% in the present class). Students' reports of when they typically began their reading assignments did differ significantly from previous classes to the present class, X = -3.26, p = .001. See Table 1 for full data.

Predictors of Student Success

A series of independent means t-tests were performed to determine if students in each section differed significantly on any of the measured potential predictors of students' success in college classes. There were no significant differences found for self-reported current or high school GPA, competency expectations, intrinsic motivation for the course, AGQ-R achievement goals, or TIPI personality factors.

TABLE 1. Percentages of self-reported global reading timing in previous classes and the present class

Typical time began reading the assignment	Before class	Shortly after class	Several days before test	Day or night before test	Did not read them
In previous classes	46.2	10.8	28.0	14.0	1.1
In present class	66.7	5.4	16.1	11.8	0

Note. Statistically significant difference in ranks.

Assessments of retention of material

As shown in Table 2, the sections did not differ significantly in their scores on any assessments in which they had the same quizzing method (i.e., pre-manipulation and post-manipulation in both sections). There were mixed results for the hypothesis that the 9 AM section would score higher when they had the Clicker plus written quizzing method than the 8 AM section. As hypothesized, the 9 AM section scored significantly higher on Exam 2. However, contrary to hypotheses, the 9 AM section did not score significantly higher on the first portion of Exam 3 and did not perform significantly higher on any segments of the final exam, including segments assessing material covered when only the 9 AM section had the Clicker plus written quizzing method.

TABLE 2. Descriptive Statistics and Results by Assessment

	8 AM Section			9 AM Section				
Assessment	n	M	SD	n	M	SD	t	p
Exam 1[a]	52	78.52	16.28	46	83.00	15.19	-1.40	.16
Exam 2[b]	51	75.70	15.30	46	81.72	11.52	-2.20[c]	.03
Exam 3, 1[b]	51	28.44	5.78	44	29.61	5.60	-1.00	.32
Exam 3, 2[d]	51	22.92	7.84	44	24.10	7.19	-.77	.45
Exam 4[d]	52	76.34	15.63	46	80.07	12.51	-1.29	.20
Final[a]	52	8.92	1.58	46	9.46	1.47	-1.72	.09
Final[b]	52	15.25	2.63	46	15.70	2.53	-.85	.40
Final[d]	52	22.02	4.92	46	22.48	4.56	-.48	.63

Note. All p values reported as two-tailed. Exam 1 - 4 grades are based on multiple choice and written items combined. Final exam grades are based only on multiple choice items. Exams 1, 2, 4 grades are percentile scores. Exam 3 and Final scores are raw scores.

[a]Both sections were being quizzed using the Clicker only method. [b]Only the 9 AM section was being quizzed using the Clicker plus written method. [c]Levene's test for equality of variances was significant. [d]Both sections were being quizzed using the Clicker plus written method.

Course evaluations & self-reported engagement in class by quizzing methodology

Overall, students did not prefer the possibility of having a random written quiz during 25% of the classes where they were guaranteed to have a quiz. Despite the fact that students had a 100% guarantee that they would have a daily quiz (they were just not sure what format would be used to determine their grade for the day), written comments on course evaluations and responses to an anonymous mid-semester evaluation reflected that they disliked the random written quizzes. For instance, on an anonymous clicker midterm evaluation question given only in the 9 AM section, 73% reported that they would prefer the Clicker quizzes alone, 16% reported they would prefer the Clicker plus written method, and 11% reported they would prefer to have only a written quiz every chapter.

As can be seen in Table 3, even though students did not prefer the Clicker plus written method there were no significant differences on the intrinsic motivation for the course questionnaire at the end of the semester, which resembled course evaluation questions. Furthermore, their overall scores on the intrinsic motivation questions did not vary significantly by section at post-assessment, $t(91)$ = -.70, p = .49. Additionally, the majority of students (range = 50 – 66.7%) reported their reading amount, time spent preparing for classes, thoroughness of reading, class participation, and understanding of course material was the same with both quizzing methods. Finally, when asked in the post-assessment which quizzing schedule they believed would lead to the most learning, 46.7% of students chose the Clicker plus written method, 30.4% chose the Clicker only method, 17.4% chose the option that the quizzing methodology would not make a difference, and 5.4% had other suggestions.

DISCUSSION

Assessments of Reading Compliance and Behavior

Syllabus reading. Despite all students participating in the study having the capacity to earn 1% on their final grade for sending an e-mail after reading the syllabus carefully enough to read the instructions to send the e-mail, only 22.4% did so. This result almost exactly replicates Sappington et al.'s (2002) result of 22% reading compliance and extends their results by finding that even when students are given a more nuanced opportunity to be honest in their report of whether they read all of the syllabus, a majority still lie. These results lend further support to recent research suggesting the majority of students do not read assigned readings at all, or at least not thoroughly (Burchfield & Sappington, 2000; Clump et al., 2004; Connor-Greene, 2000; Sappington et al., 2002) and call into question the validity of students' self-reported reading compliance.

There was mixed support for my hypothesis that students who failed the objective measure of reading the entire syllabus (the objective syllabus reading check) would report reading all of the syllabus less than students who passed the objective measure. Although lower percentages of students who failed the objective syllabus reading check reported reading all of the syllabus on the homework and in class than students who passed the objective syllabus reading check, this difference was only statistically significant for the in-class responses. It is possible that students who passed the objective syllabus reading check were more honest in their responses at both time points and those who reported not reading the entire syllabus on their homework read the remainder of the syllabus prior to class. It is also possible that students who did not read the entire syllabus felt more compelled to be honest in class compared to on their homework assignment (approximately 19% did decrease their reported level of reading from the homework to the in-class check). It is interesting to think about why students might lie on these self-reports of their reading levels, even when given the opportunity to

https://doi.org/10.20429/ijsotl.2017.110203

give an entirely accurate response and why they might change their answers to be more valid when answering with a clicker in class, but future research is needed to examine the potential reasons behind these behaviors.

Providing further evidence of the lack of validity of student self-reported reading levels, the correlation between their self-reported syllabus reading level responses 1 and 2 was only .29. Although this correlation is statistically significant at p = .008, practically speaking it does not provide much confidence in students' self-reports. Over 25% of students who failed the objective syllabus reading check and 10% of students who passed the objective syllabus reading check changed their response from one report to the next.

TABLE 3. Descriptive Statistics and Results on Course Evaluation Questions by Section

	Percent choosing "Agree a little/ moderately/or strongly"		
	8 AM section (n = 51)	9 AM section (n = 42)	X^2
"I enjoyed this class very much."	88.24	88.10	5.42
"I think this class was interesting."	100	95.24	2.98
"I think this class was fun."	76.47	80.95	6.65
"I'm glad I took this class."	82.35	83.33	1.34
"I intend to recommend this class to others."	74.51	78.57	3.28
	Percent choosing "Disagree a little/moderately/or strongly"		
	8 AM section (n = 51)	9 AM section (n = 42)	X^2
"I think this class was a waste of my time."	88.24	83.33	3.99
"I think this class was boring."	80.39	71.43	5.49
"I didn't like this class at all."	88.24	85.71	4.02

Note. Revised version of items utilized by Elliot and Church (1997) to measure students' intrinsic motivation for the course at the end of the semester. All X^2 are not significant.

Daily clicker reading check. Due to the multiple indicators that students' self-reported reading levels are invalid, no analyses were conducted on their daily self-reported reading levels during the semester. Despite the barriers to obtaining valid self-reported reading levels it is important to attempt to increase students' reading compliance so that they can obtain the maximum benefit from their college courses. Future research will have to contend with these barriers when evaluating interventions aimed at increasing reading compliance levels.

Global reading behavior. If the students' self-reported global reading behavior on their pre- and post-assessments can be trusted (which is questionable), a majority of students reported that they completed the reading assignments by the assigned dates for both this class and previous classes. Therefore, adding quizzing to a class may not have an impact on how frequently students read (or at least say they read) assignments by the "assigned date" compared to their other classes. However, consistent with previous research (Connor-Greene, 2000; Morling et al., 2008), quizzing may have an impact on when students read their assigned readings by encouraging them to read the material assigned for each class prior to that class. Students' reports of when they typically read the assigned reading differed significantly between previous classes and the present class, with a majority of students reporting they began reading the assigned reading prior to class in the present class only. Thus, students may perceive the "assigned date" for reading assignments as the first time they will be tested on that material, rather than the date listed on the syllabus. Therefore, daily quizzes may be effective for encouraging students to read prior to class even if they do not read the entire assigned reading prior to class.

Predictors of student success

The sections did not differ significantly on any of the predictors of student success, thus increasing confidence in the likelihood that any differences found on assessments of retention of material by section were because of the manipulation rather than pre-existing differences on these predictors.

Assessments of retention of material

Overall, there was mixed support regarding whether the quizzing methodology made a significant impact on students' retention of the material. Based on Exams 1 and 2 it appeared that the manipulation in the 9 AM section did cause students to perform better on the assessment of material covered when they were quizzed using the Clicker plus written quizzing methodology. However, none of the other predicted differences were found between the sections' performance on the remaining assessments of their retention of the material tested.

Given the mixed support for the hypotheses regarding the assessments, it is unclear at this point whether the change in the quizzing methodology caused any increases in short- or long-term retention of the material. One possible reason for the mixed results is that the quizzing methodologies did not differ in a manner necessary to promote significantly higher test scores. As Nguyen and McDaniel (2015) discuss, similarity between quiz and test items appears necessary to obtain the benefits of the testing effect. Given the primary goal of quizzes in this study was to increase reading prior to class and open up relevant class discussions, both written and clicker questions were intentionally focused on basic knowledge students should remember from the reading rather than assessing student understanding. Therefore, neither quizzing methodology primarily utilized quiz questions identical or very similar to test items as recommended by Nguyen and McDaniel (2015) to obtain the testing effect. Thus, it remains a possibility that there was no testing effect in this study and the differences found on Exam 2 are the result of some other factor.

The differences in Exam 2 and no other assessments could be a result of changes in student behaviors across the semester. For instance, it is possible that students began to study differently for the exams once they began being quizzed with the Clicker plus written method. Given that all students studied for Exam 3, 4, and the Final under this method, that would make it unlikely to find significant differences between the groups for any exams after Exam 2. It is also possible that the students in the 9 AM section did indeed begin coming to class more prepared than the 8 AM section once their quizzing methodology changed, leading to higher Exam 2 scores, but aspects of the quizzing methodology (or the semester) led to the students becoming less prepared as the semester went on. For instance, due to the written quiz only being given in 25% of the remaining classes it is possible that students felt less motivated to be prepared for the written quiz as they had completed more because they (inaccurately) felt as though the probability of having a written quiz decreased for each remaining class. It is also possible that students felt more overwhelmed by other responsibilities as the semester went on and were thus less prepared in general (as one student reported on the post-assessment). Finally, it is possible that some students became discouraged by the written quizzes because, unlike the Clicker quizzes, it was impossible to guess on them and they felt they were too difficult to succeed on without studying (which was hinted at in students' written comments on the post-assessment).

Course Evaluations & Self-reported engagement in class by quizzing methodology

Given the current importance of student evaluations for assessing professors' teaching effectiveness (Shao et al., 2007), it is important to note that the student evaluations of the course as a whole did not differ by section (even though the 9 AM section had the Clicker plus written quizzing methodology for 75.6% of the semester and the 8AM section only had the Clicker plus written methodology for 39% of the semester). This was found despite the fact that written comments on course evaluations and the mid-semester evaluations in the 9AM section suggested many students would prefer the Clicker only quizzing methodology. Overall, student evaluations of the course remained strong across sections, with the majority responding in the desired direction on opinions of the class. Perhaps one reason there were no significant differences in course evaluations by section is that the majority of students did not report their behavior differed significantly by quizzing methodology (i.e., percent of reading completed, thoroughness of their reading, time spent preparing for classes, class participation, and understanding of course material was reported to be the same with both quizzing methods for the majority of students). While the modal response to the question regarding which quizzing method would lead to the most learning was the Clicker plus written method, the majority of students chose other options. Taken together I believe these results suggest that professors who want to incorporate daily quizzing in their course do not need to be excessively concerned about the impact on their course evaluations based on type of methodology. Additionally, it appears that students do not perceive a major impact of quizzing methodology on their study related behaviors.

STRENGTHS, LIMITATIONS, AND FUTURE DIRECTIONS

One major strength of the current study is it was able to answer the important question of whether students enrolled in the sections involved differed significantly on other variables that may impact student performance other than the quasi-experimental manipulation. Although only one of the hypothesized differences in exam grades was found, we can be relatively sure that the difference was not due to pre-existing differences between the sections. Future research on the impact of teaching methods should evaluate comparison groups for potentially important pre-existing differences to ensure results are due to the manipulation only.

An additional strength of the current study is the fact that there were minimal differences in the sections other than the quizzing methodology because both sections were taught the same semester, by the same professor, only one hour apart with identical assessments. Furthermore, the sections had identical class plans for all but the eight classes in which one section received the written quiz at the start of class. During these classes the class plan was identical for the time remaining after the written quiz, but that section experienced a more rushed version of the class plan. One potential important difference between the sections and a limitation of the study which could not be prevented (or easily evaluated) was that I was likely able to provide better class sessions in the 9 AM section due to the practice received by teaching that same class at 8 AM. The fact that the 9 AM section had exam scores 2.5 – 6 points higher than the 8 AM section despite no significant differences in their predictors of student success suggests this may have happened.

Even though the multiple baseline quasi-experimental methodology employed in the present study allowed for a potential replication of results (by evaluating whether the quizzing methodology caused an increase in exam performance for both sections), it did not allow for a more clear-cut differentiation between the quizzing methodologies. Future research should employ the more standard two condition quasi-experimental design for teaching method manipulations.

CONCLUSIONS

The current study adds to two major aspects of the scholarship of teaching and learning literature. First, it adds to the literature suggesting that students do not fully read assigned readings and a majority will lie about their reading level. This finding made it impossible to run an analysis that would be considered valid regarding whether quizzing methodology impacted students' reading levels on a day-to-day basis. Second, it adds to the literature on the benefits of incorporating daily quizzing and begins to evaluate potential differences in daily quizzing methodology. Given the overall results of this study, it appears that while daily quizzes are beneficial for students' long-term retention of the material it does not appear that the additional grading time required by written quizzes is warranted for obtaining outcomes not already obtained by the use of daily clicker reading comprehension quizzes.

NOTES

[1]The change in quizzing methodology was announced during the first class following Exam #1 (class #11) for the 9 AM section. For the 9 AM section, 25% of remaining classes meant they would have a written quiz in six of the remaining 24 classes covering new material. In reality, five of the written quiz classes were randomly chosen and the sixth quiz was intentionally the last class covering new material to ensure that students would always think it was a possibility to have a written quiz. Students were informed that the reading for that day and the following day would appear on the next quiz to ensure that they were reading with the possibility of a written quiz for all remaining material for the semester. At the end of the fourth class covering new material for the third exam (class #25), the 8 AM section was informed the same thing as the 9 AM section, but this meant they would only have two randomly-chosen classes and the final class covering new material as written quiz days for their section. I also posted and e-mailed an announcement through Blackboard informing students of the change in the quiz policy following the relevant class. I originally posted the announcement for the 9 AM section to the 8 AM section but removed the announcement immediately and sent an e-mail saying "Please ignore the e-mail that you just received, that was intended for another class." No students from the 8 AM section

https://doi.org/10.20429/ijsotl.2017.110203

indicated they received the e-mail or were concerned the e-mail had been intended for them, nor did they indicate they remembered it when I announced the changed policy in their section.

ACKNOWLEDGEMENTS

I would like to express special thanks to the research assistants involved with this project: Billy Rush, Teresa Davis, Mia Kloth, and Brittney Stone. I would also like to express sincere appreciation to Conor Dowling and the reviewers for their feedback on previous versions of this manuscript.

REFERENCES

American Psychological Association. (2013). *APA guidelines for the undergraduate psychology major:Version 2.0*. Retrieved from http://www.apa.org/ed/precollege/undergrad/index.aspx

Brady, M., Seli, H., & Rosenthal, J. (2013). "Clickers" and metacognition: A quasi-experimental comparative study about metacognitive self-regulation and use of electronic feedback devices. *Computers & Education, 65*, 56-63. doi: http://dx.doi.org/10.1016/j.compedu.2013.02.001

Burchfield, C. M. & Sappington, J. (2000). Compliance with required reading assignments. *Teaching of Psychology, 27*, 58 – 60.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918 – 928. doi: 10.1037/0278-7393.34.4.918

Carkenord, D. M. (1994). Motivating students to read journal articles. *Teaching of Psychology, 21*, 162 – 164. doi: 10.1207/s15328023top2103_8

Clark, M. H., Middleton, S. C., Nguyen, D., & Zwick, L. K. (2014). Mediating relationships between academic motivation, academic integration and academic performance. *Learning and Individual Differences, 33*, 30 – 38. doi: 10.1016/j.lindif.2014.04.007

Clump, M. A., Bauer, H., Bradley, C. (2004). The extent to which psychology students read textbooks: A multiple class analysis of reading across the psychology curriculum. *Journal of Instructional Psychology, 31*, 227 – 232.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*, 339 – 346. doi: 10.1016/j.lindif.2014.04.007

Connor-Greene, P. A. (2000). Assessing and promoting student learning: Blurring the line between teaching and testing. *Teaching of Psychology, 27*, 84 – 88. doi: 10.1207/S15328023TOP2702_01

Elliot, A. J. & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218 -232. doi: 10.1037/0022-3514.72.1.218

Elliot, A. J. & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*, 613 – 628. doi: 10.1037/0022-0663.100.3.613

Gmelch, W. H., Lovrich, N. P., & Wilke, P. K. (1984). Sources of stress in academe: A national perspective. *Research in Higher Education, 20*, 477 – 490. doi: 10.1007/BF00974924

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Psychology, 37*, 504 – 528. doi: 10.1016/S0092-6566(03)00046-1

Hall, R. H., Collier, H. L., Thomas, M. L., & Hilgers, M. G. (2005). A student response system for increasing engagement, motivation, and learning in high enrollment lectures. *Proceedings of the Eleventh Americas Conference on Information Systems, USA, 1,* 1792 – 1797. Retrieved from http://lite.mst.edu/media/research/ctel/documents/hall_et_al_srs_amcis_proceedings.pdf

Kappe, R. & van der Flier, H. (2012). Predicting academic success in higher education: What's more important than being smart? *European Journal of Psychology of Education, 27*, 605 – 619. doi: 10.1007/s10212-011-0099-9

Karp, D. A. & Yoels, W. C. (1976). The college classroom: Some observations on the meanings of student participation. *Sociology & Social Research, 60*, 421 – 439.

Morling, B., McAuliffe, M., Cohen, L. & DiLorenzo, T. M. (2008). Efficacy of personal response systems ("Clickers") in large, introductory psychology classes. *Teaching of Psychology, 35*, 45 – 50. doi: 10.1080/00986280701818516

Nguyen, K. & McDaniel, M. A. (2015). Using quizzing to assist student learning in the classroom: The good, the bad, and the ugly. *Teaching of Psychology, 41*, 87 – 92. doi: 10.1177/0098628314562685

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*, 382 – 395. doi: 10.1037/a0026252

Roediger, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improved long-term retention. *Psychological Science, 17*, 249 – 255.

Ruscio, J. (2001). Administering quizzes at random to increase students' reading. *Teaching of Psychology, 28*, 204 – 206. doi: 10.1207/S15328023TOP2803_08

Ryan, T. E. (2006). Motivating novice students to read their textbooks. *Journal of Instructional Psychology, 33*, 135-140.

Sappington, J., Kinsey, K., Munsayac, K. (2002). Two studies of reading compliance among college students. *Teaching of Psychology, 29,* 272 – 274.

Shao, L. P., Anderson, L. P. & Newsome, M. (2007). Evaluating teaching effectiveness: where we are and where we should be. *Assessment & Evaluation in Higher Education, 32*, 355 – 371. doi: 10.1080/02602930600801886