Behavior Analysis: Research and Practice
2017, Vol. 17, No. 2, 129–141

2372-9414/17/$12.00 http://dx.doi.org/10.1037/bar0000073

An Evaluation of Readiness Assessment Tests in a College Classroom: Exam Performance, Attendance, and Participation

Megan R. Heinicke
California State University, Sacramento

Carrie K. Zuckerman
Auburn University

Danielle A. Cravalho
California State University, Sacramento

This study evaluated the use of frequent online assessments due prior to lecture, known as readiness assessment tests (RATs), in 2 sections of an upper-division psychology course. We compared the efficacy of RATs on students' exam performance, in-class participation, and attendance using a nonequivalent control group design. We also measured students' self-report of studying and preference for RATs using a satisfaction survey. Results indicated significantly higher average unit exam grades, a higher level of student attendance, and more reports of active study habits for students exposed to RATs compared with the control group, but we did not find significant differences in student participation. Students also reported a preference for RATs over frequent in-class quizzes. Overall, our results support that RATs may be an effective and preferred assessment strategy to improve students' overall exam grades and promote active study habits. Recommendations for assessment in higher education and future research are discussed.

Keywords: attendance, college students, exam scores, participation, readiness assessment tests

Academic success in college requires a combination of active study habits such as completing assigned readings before class, taking effective notes during lectures, and studying course materials regularly (Credé & Kuncel, 2008; Lei, 2015). Many students entering college are not proficient in effective study skills (Karpicke, Butler, & Roediger, 2009; Lammers, Onwuegbuzie, & Slate, 2001). This lack of preparedness may be caused by a discontinuity between high school and college course demands (MacKenzie, 2009). For example, college students are expected to learn primarily from assigned readings and lectures and are expected to spend much more time outside of the classroom studying course materials (Michael, 1991; Thomas, Bol, & Warkentin, 1991). Studying may be maintained by aversive control, in that it relieves the anxiety of not being prepared for an upcoming exam (Michael, 1991). As such, a sharp increase in studying often occurs as the exam approaches (e.g., Jarmolowicz, Hayashi, & St. Peter Pipkin, 2010; Perrin et al., 2011). Distributed practice (studying across time) is highly effective, compared with studying once before an exam, but may take more time and effort (Dunlosky, 2013). Additionally, environmental variables that may interfere with consistent studying are plentiful in college (e.g., access to leisure media), and these variables often have more immediate consequences (Panek, 2014).

This article was published Online First February 27, 2017.

Megan R. Heinicke, Department of Psychology, California State University, Sacramento; Carrie K. Zuckerman, Department of Psychology, Auburn University; Danielle A. Cravalho, Department of Psychology, California State University, Sacramento.

Danielle Cravalho is now at Department of Psychology, University of California, Riverside.

Data appearing in this article have been presented by the authors in symposia at the 42nd Annual Convention of the Association for Behavior Analysis and the 34th Annual Western Regional Conference for the California Association for Behavior Analysis. We thank the research assistants who served as data collectors: Amanda Chastain, Martha Cisneros, Emily Darcey, Danielle Geierman, Joanna Hurtt, Kyla McLaughlin, Vicki Moreno, and Kaylin Stromberg.

Correspondence concerning this article should be addressed to Megan R. Heinicke, Department of Psychology, California State University, 6000 J Street, Sacramento, CA 95819-6007. E-mail: megan.heinicke@csus.edu

Although instructors may have little control over competing variables outside the classroom, carefully designed course contingencies can make a positive impact on student study habits (e.g., Johnson, Perrin, Salo, Deschaine, & Johnson, 2016; Perrin et al., 2011). Instructors have control over the grading structure of a course, which can be an important motivational factor for students. One way in which instructors can use this motivational factor to promote effective study habits is through frequent testing (e.g., Lyle & Crawford, 2011). Distributed practice, in the form of self-testing or practice tests, is a highly effective strategy to promote retention of course content and enhance retrieval on exams (Dunlosky, 2013; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger & Butler, 2011). However, researchers have found that students do not engage in self-testing (and often repeatedly read over their notes or assigned readings) when left to self-regulate their learning (Dunlosky, 2013; Karpicke et al., 2009). Instructors implementing frequent tests create contingencies to promote effective study habits when tests impact overall course grades and require a thorough understanding of the material (Lyle & Crawford, 2011; McDaniel, Wildman, & Anderson, 2012; Michael, 1991).

Frequent testing procedures have been empirically evaluated in many college classrooms with positive results (Bangert-Drowns, Kulik, & Kulik, 1991; Duty, 1982; Gaynor & Millham, 1976; Kuo & Simon, 2009; Martin & Srikameswaran, 1974; Roediger & Karpicke, 2006). For example, Connor-Greene (2000) sought to compare students' self-report of studying between classes experiencing daily quizzes versus four infrequent exams. The majority of students reported that they completed their readings in both testing conditions. However, 92% of the students who were given daily quizzes reported completing the reading by the assigned date (i.e., before attending class) compared with 16% of students with four scheduled tests. In addition, several studies have shown that students are more likely to attend classes with frequent quizzing or testing schedules (Hovell, Williams, & Semb, 1979; Pennebaker, Gosling, & Ferrell, 2013; Wilder, Flood, & Stromsnes, 2001). Overall, past research has demonstrated frequent testing contingencies may increase learning, the timing of studying, and class attendance, and students have reported positive attitudes toward instruction when frequent testing schedules were implemented (Bangert-Drowns et al., 1991; Kuo & Simon, 2009; Nguyen & McDaniel, 2015; Roediger & Karpicke, 2006).

Despite positive outcomes, intensive testing schedules have been shown to have a few disadvantages for students and instructors. First, students with little to no experience with frequent testing schedules may report a punitive feeling toward having to prepare for multiple quizzes or exams and may display greater feelings of anxiety in the classroom (e.g., Padilla-Walker, 2006). Second, instructors implementing frequent testing schedules may experience a negative testing effect if they opt to sample items provided by textbook publishing companies (e.g., quiz bank questions, supplemental quizzes) rather than developing their own quiz items that relate to later exam items (Nguyen & McDaniel, 2015). Instructors may choose this option to limit the amount of time and resources needed to implement a frequent testing schedule. However, recent research tentatively suggests that adopting supplemental textbook materials may not lead to student gains if there is an unsystematic association between quiz and exam items (Nguyen & McDaniel, 2015; Wooldridge, Bugg, McDaniel, & Liu, 2014). Finally, frequent in-class quizzes may take away from the allotted time students spend in class, leaving the instructor with less time to cover course material and facilitate in-class interactions. To overcome this disadvantage of frequent testing, instructors could save valuable class time by posting quizzes online for students to complete outside of class. One approach that uses this method is Just in Time Teaching (JiTT; Novak, Patterson, Gavrin, & Christian, 1999).

JiTT includes frequent online Readiness Assessment Tests (RATs) to ensure students complete assigned readings before class (Novak et al., 1999). RATs are typically due 2 hr before class begins and contain questions from the reading material that will be covered in the upcoming lecture. The instructor reviews student RAT responses before class and adapts the lecture based on student responses (Novak et al., 1999). Benedict and Anderton (2004) evaluated JiTT and found that students earned higher final exam scores and reported greater satisfaction with the course compared with stu-

dents who completed in-class reading quizzes. A greater number of students also reported reading before class when JiTT was added to a lecture course (Howard, 2004). Although JiTT promotes appropriate study habits, provides the instructor with information on which areas of the material to focus on during lecture, and minimizes the use of class time for assessment, it may be unrealistic for some instructors to adopt JiTT. Instructors implementing JiTT must devote time immediately before class (typically during the 2 hr beforehand) to review RAT responses and incorporate them into the lecture. This time commitment may be impractical if the instructor teaches an early morning class, has back-to-back classes, or has other time conflicts before class.

Instructors may not have teaching schedules that allow them to use all components of JiTT, but the RAT component of JiTT may still be a beneficial assessment strategy on its own. Weinstein and Wu (2009) compared the use of RATs alone with frequent online quizzes in a psychology course with four major content units. Students were required to complete RATs prior to class in the first and third units and completed online quizzes at the end of each week during the second and fourth units. The authors found that students scored higher on unit exams when they completed RATs in place of online quizzes. Students also reported that RATs enhanced their ability to participate in lectures, increased the amount of readings they completed, and were preferred over frequent quizzes. Although this investigation demonstrated promising results for the use of RATs, two limitations of this investigation should be noted. First, Weinstein and Wu compared their assessment procedures across units of varying difficulty in a single college classroom. Therefore, there is a possibility that students performed better in units for which RATs were used because the content in those units was less difficult than the content covered in units for which online quizzes were administered. Second, the authors did not directly measure pertinent student behaviors such as attendance and student participation. It may be the case that RATs increase these behaviors as well as exam scores.

The purpose of this investigation was to extend the work of Weinstein and Wu (2009) by evaluating the effects of RATs on students' academic success by comparing performance across two sections of a college course taught by the same instructor. In addition to evaluating students' exam scores, we directly assessed the effects of RATs on students' participation in the classroom and attendance. We also gathered students' self-report of reading, ability to follow along in class, preparation for class and unit exams, and preference for the assessment strategy.

Method

Participants and Course Structure

Students in two sections of an undergraduate Psychology of Exceptional Children course at a large Western state university participated in the current study. Both sections met in the morning (Section A at 8:00 a.m., Section B at 10:00 a.m.) for 50 min three times per week. Course content was divided into six units, each centered on a general topic, and subtopics within a unit typically spanned two class meetings. The instructor for both sections (the first author) used identical Microsoft PowerPoint slides and started and ended the lecture at the same point in the slides in an effort to cover the same content across sections each day. In addition, the instructor posted guided notes (as described by Barbetta & Skaruppa, 1995) online for both sections at least 24 hr before each lecture, but did not require students to use this note-taking strategy.

Students were not aware that they were participating in a research study until the end of the semester. On the last day of class, the instructor debriefed the students and gave them the option to provide informed consent and complete a participant information and satisfaction survey. The instructor was not aware of which students provided consent until after the semester concluded. The university's institutional review board approved all procedures used.

Section A. A total of 85 students were enrolled in Section A. One student withdrew from the course during the sixth week of the semester. Of the 84 students who completed the course, 75 students (61 females, 14 males) gave consent and completed the participant information and satisfaction survey. Students ranged from 18 to 60 years of age ($M = 24.0$, $SD = 6.7$). Section A was mostly comprised of juniors

(61.3%) and seniors (25.3%), and the majority (81.3%) of students were psychology majors. Students were enrolled in an average of 13.1 credit hours ($SD = 3.3$) and had an average grade point average (GPA) of 3.1 ($SD = 0.4$). The majority (81.3%) of students reported being employed during the semester, and most students were working between 11 to 30 hr per week.

Section B. A total of 86 students were enrolled in Section B, and one student withdrew from the course during the 10th week of the semester. Of the 85 students who completed the course, 79 students (59 females, 20 males) gave consent and completed the survey. Students ranged from 19 to 49 years of age ($M = 23.7$, $SD = 5.0$). Section B was also mostly comprised of juniors (46.8%) and seniors (49.4%), and the majority (89.9%) of students were psychology majors. Students were enrolled in an average of 13.2 credit hours ($SD = 2.1$) and had an average GPA of 3.2 ($SD = 0.4$). The majority (72.2%) of students reported being employed during the semester, and most students were working between 11 to 30 hr per week.

Experimental Design

A nonequivalent control-group design (Martella, Nelson, & Marchand-Martella, 1999) was used to evaluate the effects of RATs on students' unit exam scores, participation, and attendance. Prior to the start of the semester, a research assistant flipped a coin to select the sections that would serve as the control (Section A) and experimental (Section B) groups.

Preexperimental Procedure (Pretest)

A pretest was used to compare students' preexisting knowledge of the course content between Sections A and B because students self-selected their section assignments during enrollment. The pretest was administered online during the first week of class and consisted of 30 multiple-choice and true–false items. An independent samples t test revealed no statistically significant differences between the sections on the pretest, $t(151) = 0.40$, $p > .05$. We also used information from the participant information and satisfaction survey to evaluate group equivalence and found no differences between groups in terms of age, $t(152) = -.24$, $p > .05$, GPA, $t(143) = 1.80$, $p > .05$, and the total number of units or credit hours in which students were enrolled, $t(149) = 0.40$, $p > .05$.

Experimental Procedures

Students in both sections were exposed to the same course content and took an in-class exam worth 50 points at the end of each unit. Unit exams consisted of 38 multiple-choice and true–false items worth 1 point each and two essay questions worth 6 points each. Students were also offered the option of taking a comprehensive, remedial final exam to replace a low unit exam score. This exam consisted of 50 multiple-choice and true–false items worth 1 point each. The instructor also assigned three homework assignments that included the pretest described above (worth 10 points), a preference assessment worksheet (worth 20 points), and a short reaction paper to a guest speaker (worth 40 points). Finally, students in the experimental group were required to complete RATs online prior to class (described below), whereas students in the control group had no RAT assignments.

Experimental group (Section B). Students in Section B were given three opportunities to complete a RAT per unit (with the exception of Unit 4, for which there were four RATs). The number of RATs matched the number of readings per unit. Each RAT was worth a total of 5 points and consisted of three fill-in-the-blank definitional statements (each worth 1 point) and one reflective short-answer essay (e.g., "Describe the major point of the article," "Describe an area of the reading that was most intriguing/challenging to comprehend and why") worth 2 points. All RAT questions were developed by the instructor and related to the assigned readings. All RATs were due before the beginning of class and were posted online approximately 72 hours before the due date. Students were only required to complete two RATs per unit, for a total of 60 points. However, students could complete the third RAT in each unit for an extra credit point if they scored a minimum of 4 points on the third RAT. Students could earn 2 extra credit points in Unit 4, as there were four readings in that particular unit. That is, students in Section B could earn a total of 7 potential extra credit points over the course of the semester, whereas students in Section A were not offered extra credit opportunities. Extra credit

points were added to students' raw grades at the end of the semester (i.e., not to unit exam scores) and were not included in the dependent measures for this investigation. All RATs were graded based on both the accuracy and completeness.

Participant information and satisfaction survey. Students were given the option of completing a survey on the last day of class. The participant information portion of the survey consisted of 13 multiple-choice and fill-in-the-blank items used to gather demographic information, and the participant satisfaction portion consisted of either four (Section A) or 11 (Section B) Likert-scale questions used to gather students' perceptions of reading before class, their ability to follow discussions, participation, and feeling prepared for exams. Students in Section B were asked additional questions regarding their perceptions of and preference for RATs. All participant satisfaction survey questions are listed in full in Table 1.

Data Collection

All class meetings were video recorded. Per the university's institutional review board instructions, the video camera was placed in the front of the classroom and captured only the instructor (i.e., images of students were not captured). Students were told that videos were being used to evaluate the instructor's teaching style.

Unit exams and readiness assessment tests. The primary dependent measure was students' unit exam scores. Teaching assistants graded all exams and RATs using a detailed grading rubric developed by the instructor. The instructor introduced all rubrics during weekly teaching-assistant meetings, in which she modeled how to assign points using examples of correct and incorrect student responses.

Student participation. Research assistants viewed the first 30 min of each lecture via video and tallied (mainly using the audio of the recordings) student participation. Data were collected in this manner because it would have been too difficult to collect real-time data, given the large number of students in the classroom. An instance included when students (a) responded to questions posed by the instructor, or (b) asked questions of the instructor by either calling out or being called on by the instructor. Questions or comments unrelated to the class topic for the day (e.g., when an assignment was due) were not included in the total.

Instructor prompts. The instructor posed several standard-lecture, open-ended questions in an oral manner during each class. The instructor included asterisks as stimulus prompts on the PowerPoint slides in an effort to prompt her to ask the same questions to control for the number of questions presented to each group. Event recording was used to collect data (either in vivo by teaching assistants or via video by

Table 1
Percentage of Students Agreeing or Strongly Agreeing to Satisfaction Survey Questions

Question	Section A (n = 75)	Section B (n = 79)
1. I read the assigned material before attending class	13.3%	62.0%
2. I was able to follow class discussions	92.0%	94.9%
3. I felt prepared to participate in class	57.3%	77.2%
4. I felt prepared for unit exams	64.0%	79.7%
5. I perform better on exams when I complete RATs		59.5%
6. I like using RATs		53.2%
7. Other instructors should use RATs		59.5%
8. I spent more time listening and thinking about concepts in class on days when I completed a RAT		65.8%
9. I did not like being required to complete RATs		20.3%
10. Overall, I had a positive reaction to RATs		64.6%
11. I would prefer RATs to in-class reading quizzes		77.2%

Note. Ratings were based on a Likert-type scale from 1 (*strongly agree*) to 5 (*strongly disagree*). Students in both sections completed Questions 1 to 4, and only students in the experimental group (Section B) completed Questions 5 to 11. RAT = readiness assessment test.

research assistants) on the number of instructor prompts in the first 30 min of each lecture. Any questions not related to class material (e.g., asking students to repeat back when an assignment was due) were not included in the total.

Student attendance. Approximately 10 min into each lecture, a teaching assistant tallied the number of students in the classroom. Any students who entered the classroom after the tally had been completed were not included in the total. The total was then divided by the number of students enrolled in the section to generate a percentage of attendance for each lecture.

Interobserver Agreement

Unit exams and readiness assessment tests. Teaching assistants photocopied a random selection of approximately 30% of the students' RATs or exams for a second teaching assistant (who was blind to the experimental and control groups) to grade. This practice was used for 100% of RATs and exams to ensure students were graded fairly and accurately. Interobserver agreement (IOA) was calculated using the point-by-point method by dividing the number of agreements by the total number of agreements plus disagreements and converting the ratio to a percentage (Cooper, Heron, & Heward, 2007). Mean IOA across all unit exams was 97.4% (range = 95.9% to 98.5%) for Section A and 97.2% (range = 94.6% to 98.9%) for Section B. Mean IOA across all RATs was 98.7% (range = 82.6% to 100%) for Section B.

Student participation. A second research assistant independently collected data on student participation for 100% of lectures for both sections. IOA was calculated using the frequency ratio method by dividing the smaller total by the larger total and converting the ratio to a percentage (Kazdin, 2011). Mean IOA was 98% (range = 81.8% to 100%) for Section A and 98.1% (range = 83.3% to 100%) for Section B.

Instructor prompts. A second teaching or research assistant independently collected data on instructor prompts for 92.9% and 65.5% of lectures for Sections A and B, respectively. IOA was calculated using the frequency ratio method. Mean IOA was 96.3% (range = 80% to 100%) for Section A and 95.2% (range = 80% to 100%) for Section B.

Student attendance. A second teaching assistant independently counted the number of students in the classroom for 96.8% and 69% of lectures for Sections A and B, respectively. IOA was calculated using the frequency ratio method. Mean IOA was 100% for Section A and 99.4% (range = 97% to 100%) for Section B.

Effect Size Calculations

We selected Cohen's *d* (Cohen, 1988) as a measure of effect size (ES) to estimate the practical significance of RATs on unit exam scores, participation, and attendance. We calculated *d* by dividing the mean difference between groups by the pooled standard deviation and interpreted scores by using Cohen's (1988) suggested benchmarks of 0.2, 0.5, and 0.8 to indicate small, medium, and large effects.

Results

Unit Exams and Readiness Assessment Tests

Table 2 summarizes unit exam performance for Sections A and B. Students in Section B (exposed

Table 2
Comparison of Performance on Unit Exams for Sections A and B

	Section A (control)		Section B (experimental)					
Exam	*M*	*SD*	*M*	*SD*	Diff (B − A)	*t*(152)	*p*	ES
Unit 1	73.81	18.89	79.29	14.01	+5.48	2.05	.04	.33
Unit 2	69.69	25.74	76.82	21.15	+7.13	1.88	.06	.30
Unit 3	75.79	14.62	78.09	12.65	+2.30	1.05	.30	.17
Unit 4	77.29	21.03	81.80	14.36	+4.50	1.56	.12	.25
Unit 5	71.95	15.59	74.89	15.81	+2.94	1.16	.25	.19
Unit 6	80.39	12.52	82.56	9.74	+2.17	1.20	.23	.19

Note. Diff = difference score; ES = effect size.

to RATs) performed significantly better than the students in Section A (the control group) on the first unit exam. Although average exam scores were higher for Section B across all unit exams, independent samples t tests revealed that these differences were not significant for any exam other than Unit 1. The ESs for the first, second, and fourth unit exams are considered small effect sizes according to Cohen's (1988) criteria, whereas the ESs for the remaining exams fell below criteria for practical significance.

When comparing the distribution of average unit exam scores at the end of the semester across sections (see Figure 1), a larger proportion of students in Section B received As, Bs, and Cs (12.7%, 38.0%, and 32.9%, respectively) than students in Section A (6.7%, 32.0%, and 28.0%, respectively). Also, a larger proportion of students in Section A received Ds or failing grades (24.0% and 9.3%%, respectively) than students in Section B (12.7% and 3.8%, respectively). An independent samples t test revealed that students had a significantly higher unit exam score average in Section B (M = 78.85, SD = 9.82) than students in Section A (M = 74.82, SD = 10.94), $t(152) = 2.41$, p = .02, with an ES of .39, indicating a small to medium effect (Cohen, 1988). Also, 62.7% of students in Section A self-elected to take the optional remedial final exam compared with 41.8% of students in Section B.

The number of RATs completed by students in Section B ranged from five to 19, and the majority of students (81.0%) completed all 12 required RATs. In addition, 91.1% of students completed at least one additional RAT for extra credit, and seven students completed all 19 possible RATs. We also found a moderate positive correlation between the number RATs completed and students' overall unit exam average in Section B, $r(77) = .31$, $p < .05$.

Student Participation and Instructor Prompts

As seen in Figure 2, the number of student responses per class was quite variable for both Section A (M = 10.57, SD = 5.43) and Section B (M = 10.50, SD = 6.36) across the semester. We found no significant difference in the number of student responses between groups, $t(54) = -.04$, $p > .05$, and the ES for student participation (0.01) fell below Cohen's (1988) criteria for practical significance. We also compared the number of instructor prompts provided in Section A (M = 5.0, SD = 3.27) with the prompts provided in Section B (M = 5.3, SD = 3.70), and we found no statistically significant difference between the groups, $t(54)$ = 0.19, $p > .05$, suggesting that the instructor provided a similar amount of participation opportunities across sections.

Student Attendance

As seen in Figure 3, attendance decreased during approximately the first half of the semester in both sections followed by relatively stable atten-

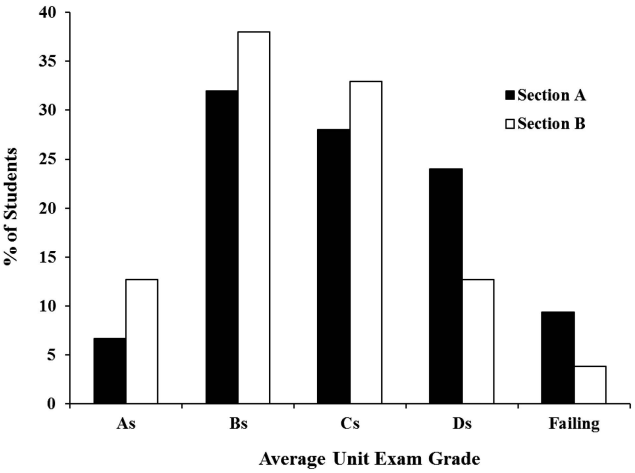

Figure 1. Distribution of end-of-the-semester average unit exam grades for students in Sections A and B.

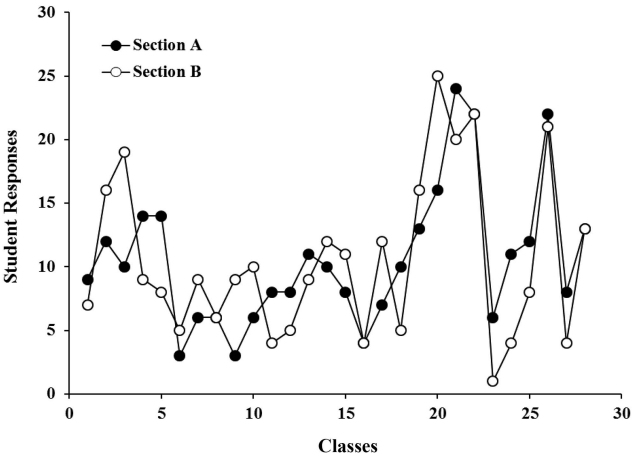

Figure 2. Number of student responses per lecture in Sections A and B over the course of the semester.

dance for the remainder of the semester. When student attendance was compared across sections, an independent samples *t* test revealed that a significantly higher percentage of students in Section B ($M = 80.55$, $SD = 9.33$) attended class than students in Section A ($M = 75.14$, $SD = 8.41$), $t(56) = 2.32$, $p = .02$. The ES for attendance was 0.61, which is considered a medium effect according to Cohen's (1988) criteria.

Participant Satisfaction Survey

The percentages of students in each section who agreed or strongly agreed with each of the questions in the participation satisfaction survey are presented in Table 1. In general, the majority of students in Section B had a positive reaction to RATs and reported spending more time listening and thinking about concepts on days when they completed RATs.

Discussion

Past research suggests that using RATs in college classrooms may have many advantages for students, such as encouraging reading before class and preparing for exams (Benedict & An-

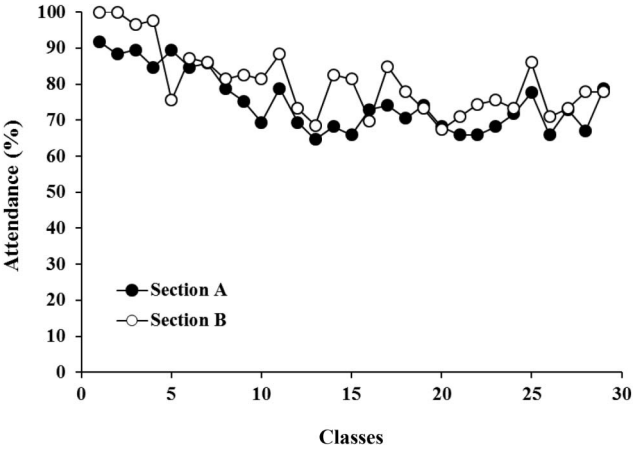

Figure 3. Percentage of students attending lectures in Sections A and B over the course of the semester.

derton, 2004; Howard, 2004; Weinstein & Wu, 2009). The purpose of the current study was to expand this line of research by evaluating the efficacy of RATs on students' (a) exam performance, (b) participation, (c) attendance, and (d) self-report of active study habits across two sections of a college course. We also measured students' reported preference for RATs in the experimental group using a satisfaction survey. We found that students required to complete RATs performed better on all unit exams compared with our control group. Those differences were statistically significant for the first unit exam and practically significant (according to our ES indices) for the first, second, and fourth exams. Our results were similar to those of Weinstein and Wu (2009), who found significant differences between only the first and second unit exam scores when comparing RATs with weekly online quizzes. However, it should be noted that Weinstein and Wu compared two assessment strategies across units of varying difficulty in one classroom, whereas we evaluated the effectiveness of RATs across two sections of the same course taught by the same instructor.

Although we only found a statistically significant difference in exam performance for the first unit exam, students appeared to benefit from the cumulative effects of RATs. We found that students exposed to RATs had significantly higher overall average unit exam grades (i.e., what many professors teaching large college courses use as a final course grade). Benedict and Anderton (2004) found similar results when comparing the effects of JiTT with five-item in-class quizzes on students' final cumulative exam scores. Furthermore, students in Section B were more likely to receive a passing course grade and were less likely to take an optional remedial final exam than students in Section A. That is, more students in the experimental group were able to allocate their resources (e.g., study time) to other courses during finals week. Taken together, these results suggest that students in the experimental group experienced both a statistically and socially significant impact of RATs on their overall exam grade.

To the authors' knowledge, the current study is the first to evaluate the effects of RATs on participation using a direct measure rather than self-report (e.g., Weinstein & Wu, 2009); however, we did not find differences in participation. The number of questions an instructor poses during lectures could reasonably contribute to the frequency of participation (e.g., Shabani & Carr, 2004). In the current study, the instructor included written cues to ask the class questions on her PowerPoint slides to control for this potential confound. However, it may be the case that other variables outside of the instructor's control led to equivalent participation across sections. For example, there may have been more students in the control group who had a genuine interest in (or practical experience with) the subject matter, which may have motivated certain students to participate more often. Anecdotally, the instructor did observe a few students who participated multiple times during each class in Section A compared with a greater variety of students who would participate (but less often) in Section B. Researchers interested in the topic of encouraging student participation might consider including a more sensitive measure that controls for potential "conversation monopolizers" by only including the number of students who participated in a given lecture in addition to an overall frequency count. Instructors might also consider measuring student responding via an electronic classroom response system, commonly referred to as clickers, when teaching large class sizes (see Keough, 2012, for a recent review of clickers in the classroom).

Earlier research suggests that frequent testing schedules can have a positive impact on student attendance (e.g., Hovell et al., 1979; Pennebaker et al., 2013; Wilder et al., 2001), and we found comparable results. This finding is noteworthy, as students in the current study were not required to attend class to complete online RATs or receive points toward their grade. We may have observed an increase in attendance in the experimental section because completing RATs increased students' motivation to attend class. That is, attending class was more reinforcing when students were prepared. It should be noted that more students in the experimental group attended class at the beginning of the semester compared with the control group. Therefore, our significant difference between groups could be due to the independent variable as well as potential extraneous variables such as the timing of the class (e.g., Section B met at 10:00 a.m., whereas Section A met at 8:00 a.m.).

Past evaluations of JiTT and RATs have found that students report positive reactions to these teaching and assessment strategies (e.g., Benedict & Anderton, 2004; Howard, 2004; Weinstein & Wu, 2009), and students responded similarly in the current study. In addition, 77.2% of students experiencing RATs reported they would prefer RATs to in-class reading quizzes. This proportion is larger than what has been reported in previous studies (e.g., Weinstein & Wu, 2009). This discrepancy may be due to the fact that students in the Weinstein and Wu (2009) study were asked their preference between online RATs and *online* weekly quizzes rather than *in-class* reading quizzes. In addition, the students in the Weinstein and Wu study were exposed to both assessment strategies and may have been able to report more accurately, whereas students our investigation were never exposed to in-class quizzes.

A few potential limitations of the investigation should be mentioned. First, procedural fidelity was not assessed other than the measurement of instructor prompts. Although the procedure can probably be considered low risk for failure (i.e., the same instructor taught both sections using identical RAT assignments and slides), researchers interested in extending our procedures should conduct procedural fidelity checks to ensure the same content was taught across groups. Second, our lack of participant randomization poses a threat to the internal validity of our investigation. That is, if there were any prior differences between the students in Sections A and B, our results might have been prone to either Type I or Type II error. However, we found no significant differences between the groups in terms of participant characteristics and pretest scores, and our choice of a nonequivalent control group design allowed us to control for the difficulty of content across units. Although random assignment was not feasible for our study, researchers whose courses are web based and include online instruction have the potential to apply random assignment, as suggested by LoSchiavo, Shatz, and Poling (2008).

Third, we may have been able to gather more information from students on our student satisfaction survey if we (a) asked students to nominate assessment strategies that have encouraged active study habits in their past, and (b) administered the survey repeatedly (e.g., Kellum, Carr, & Dozier, 2001) to evaluate whether students' preferences shifted over the semester. Third, we evaluated RATs in comparison with a control group experiencing no RATs, rather than pitting RATs against another assessment strategy. Researchers interested in this topic area might consider conducting comparisons of RATs to (a) online weekly quizzes, (b) frequent in-class quizzes (both scheduled and randomized), (c) in-class writing exercises, or (d) online tools or exercises provided by textbook publishing companies.

Finally, the instructor allowed students in both sections to use guided notes because this active student response technique is preferred by students (Neef, McCord, & Ferreri, 2006) and has been shown to increase student participation (Austin, Lee, Thibeault, Carr, & Bailey, 2002), quiz scores (Williams, Weil, & Porter, 2012), and the quality of student note-taking (Austin, Lee, & Carr, 2004). However, the addition of guided notes may have increased the levels of the dependent variables in this investigation, and therefore compressed differences between groups. Thus, instructors interested in this area of research might consider evaluating RATs with and without guided notes to compare the effects of the assessment strategy on its own versus its combined effects with an active student response procedure. It is also possible that we observed higher levels of our dependent variables in Section B because more students in that section took advantage of guided notes. Instructors should consider measuring students' use of guided notes via self-report on a student satisfaction survey or by collecting students' notes and directly measuring note-taking (e.g., Austin et al., 2004). As with any supplemental supports (e.g., e-mail, office hours, study groups, access to teaching assistants, flash cards), students in both sections had equal access to guided notes and were free to use them at their discretion.

Although we did find RATs to be beneficial for students, instructors should take the time to weigh the practical disadvantages of RATs before rolling out this assessment strategy in their classrooms. Instructors would need to consider the additional time required to create multiple quizzes, manage the web forum, and grade RATs. Although we did not measure the instructor's total preparation time for each section, researchers interested in RATs might consider including this metric in future

investigations. Simultaneously, instructors should be aware of potential technical issues associated with entirely web-based assessments and students' possible inability to complete the assessment if they have limited access to the web. Lastly, this strategy may be difficult to administer with large class sizes. For instance, if instructors do not have additional assistance, such as teaching assistants, administering RATs may be too resource intensive.

Instructors who do not have the resources required to implement RATs might consider other strategies to increase active study habits. By including randomized or announced in-class quizzes or writings (Butler, Phillmann, & Smart, 2001; Krohn et al., 2008; Wilder et al., 2001), instructors can provide contingencies for attendance. However, it has been suggested that students dislike required in-class quizzes (Combs, 1976). Extra credit opportunities such as writings (e.g., Carkenord, 1994) and quizzes (e.g., Padilla-Walker, 2006), or creating contingencies in which students are allowed to use their notes during exams only if they attend previous lectures (e.g., Messling & Dermer, 2010), may have the same effect on attendance without the punitive nature of required in-class quizzes. Instructors might also consider implementing active student response strategies (Heward, 1997). One strategy, guided notes, was used in both sections in this study. Another strategy known as "response cards" allows the class to respond to instructor prompts by holding up written or preprinted cards (e.g., Kellum et al., 2001). Instructors of large class sizes may also opt to use a high-tech alternative of response cards in which students respond via clickers (e.g., Fies & Marshall, 2006; Flosason, McGee, & Diener-Ludwig, 2015; Jones, Crandall, Vogler, & Robinson, 2013).

The current study found that the arrangement of contingencies in the RATs section promoted student success by producing higher unit exam scores and levels of attendance. Likely because of higher exam scores and increased attendance, a larger proportion of students who completed RATs earned high grades (A, B, or C) and fewer failed the course. Because some students may encounter college courses that do not actively promote effective study strategies, educators may be able to promote continued academic success by being slightly more systematic in facilitating the transition from high school to college. High school teachers could gradually fade achievement-related activities and introduce contingencies for regular independent study through frequent quizzing. High school teachers could also introduce a document similar to a syllabus that lays out student expectations, test dates, and deadlines to promote self-management. In addition, college instructors could teach study skills either outside or within the context of a course. For example, a general education or elective course in academic strategies (e.g., Tuckman & Kennedy, 2011) or self-management (e.g., Choi & Chung, 2012) could be offered and emphasized early in the curriculum. Finally, instructors teaching lower division courses could incorporate contingencies that promote effective study habits, such as frequent assessment or RATs, into their classrooms and encourage colleagues to do the same in their courses.

References

Austin, J. L., Lee, M., & Carr, J. P. (2004). The effects of guided notes on undergraduate students' recording of lecture content. *Journal of Instructional Psychology, 31,* 314–320.

Austin, J. L., Lee, M. G., Thibeault, M. D., Carr, J. E., & Bailey, J. S. (2002). Effects of guided notes on university students' responding and recall of information. *Journal of Behavioral Education, 11,* 243–254. http://dx.doi.org/10.1023/A:1021110922552

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research, 85,* 89–99. http://dx.doi.org/10.1080/00220671.1991.10702818

Barbetta, P. M., & Skaruppa, C. L. (1995). Looking for a way to improve your behavior analysis lectures? Try guided notes. *The Behavior Analyst, 18,* 155–160.

Benedict, J. O., & Anderton, J. B. (2004). Applying the just-in-time teaching approach to teaching statistics. *Teaching of Psychology, 31,* 197–199.

Butler, A., Phillmann, K., & Smart, L. (2001). Active learning within a lecture: Assessing the impact of short, in-class writing exercises. *Teaching of Psychology, 28,* 257–259. http://dx.doi.org/10.1207/S15328023TOP2804_04

Carkenord, D. M. (1994). Motivating students to read journal articles. *Teaching of Psychology, 21,* 162–164. http://dx.doi.org/10.1177/009862839402100309

Choi, J. H., & Chung, K. M. (2012). Effectiveness of a college-level self-management course on successful behavior change. *Behavior Modification, 36,* 18–36. http://dx.doi.org/10.1177/0145445511418102

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Combs, L. L. (1976). A teaching approach to physical chemistry. *Journal of Research in Science Teaching, 13,* 467–472. http://dx.doi.org/10.1002/tea.3660130511

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

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill-Prentice Hall.

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3,* 425–453. http://dx.doi.org/10.1111/j.1745-6924.2008.00089.x

Dunlosky, J. (2013). Strengthening the student toolbox: Study strategies to boost learning. *American Educator, 37,* 12–21.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Duty, R. C. (1982). Weekly or biweekly quizzes in organic chemistry: Does it make a difference? *Journal of Chemical Education, 59,* 218–219. http://dx.doi.org/10.1021/ed059p218

Fies, C., & Marshall, J. (2006). Classroom response systems: A review of the literature. *Journal of Science Education and Technology, 15,* 101–109. http://dx.doi.org/10.1007/s10956-006-0360-1

Flosason, T. O., McGee, H. M., & Diener-Ludwig, L. (2015). Evaluating impact of small-group discussion on learning utilizing a classroom response system. *Journal of Behavioral Education, 24,* 317–337. http://dx.doi.org/10.1007/s10864-015-9225-0

Gaynor, J., & Millham, J. (1976). Student performance and evaluation under variant teaching and testing methods in a large college course. *Journal of Educational Psychology, 68,* 312–317. http://dx.doi.org/10.1037/0022-0663.68.3.312

Heward, W. L. (1997). Four validated instructional strategies. *Behavior and Social Issues, 7,* 43–51. http://dx.doi.org/10.5210/bsi.v7i1.298

Hovell, M. F., Williams, R. L., & Semb, G. (1979). Analysis of undergraduates' attendance at class meetings with and without grade-related contingencies: A contrast effect. *The Journal of Educational Research, 73,* 50–53. http://dx.doi.org/10.1080/00220671.1979.10885204

Howard, J. R. (2004). Just-in-time teaching in sociology or how I convinced my students to actually read the assignment. *Teaching Sociology, 32,* 385–390. http://dx.doi.org/10.1177/0092055X0403200404

Jarmolowicz, D. P., Hayashi, Y., & St. Peter Pipkin, C. (2010). Temporal patterns of behavior from the scheduling of psychology quizzes. *Journal of Applied Behavior Analysis, 43,* 297–301. http://dx.doi.org/10.1901/jaba.2010.43-297

Johnson, P. E., Jr., Perrin, C. J., Salo, A., Deschaine, E., & Johnson, B. (2016). Use of an explicit rule decreases procrastination in university students. *Journal of Applied Behavior Analysis, 49,* 346–358. http://dx.doi.org/10.1002/jaba.287

Jones, S. J., Crandall, J., Vogler, J. S., & Robinson, D. D. (2013). Classroom response systems facilitate student accountability, readiness, and learning. *Journal of Educational Computing Research, 49,* 155–171. http://dx.doi.org/10.2190/EC.49.2.b

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17,* 471–479. http://dx.doi.org/10.1080/09658210802647009

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Kellum, K. K., Carr, J. E., & Dozier, C. L. (2001). Response-card instruction and student learning in a college classroom. *Teaching of Psychology, 28,* 101–104. http://dx.doi.org/10.1207/S15328023TOP2802_06

Keough, S. M. (2012). Clickers in the classroom: A review and a replication. *Journal of Management Education, 36,* 822–847. http://dx.doi.org/10.1177/1052562912454808

Krohn, K. R., Parker, M. R., Foster, L. N., Aspiranti, K. B., McCleary, D. F., & Williams, R. L. (2008). Effects of writing-related contingencies on both quality of writing and multiple-choice exam performance in large college courses. *The Behavior Analyst Today, 9,* 184–195. http://dx.doi.org/10.1037/h0100658

Kuo, T., & Simon, A. (2009). How many tests do we really need? *College Teaching, 57,* 156–160. http://dx.doi.org/10.3200/CTCH.57.3.156-160

Lammers, W. J., Onwuegbuzie, A. J., & Slate, J. R. (2001). Academic success as a function of the gender, class, age, study habits, and employment of college students. *Research in the Schools, 8,* 71–81.

Lei, S. A. (2015). Variation in study patterns among college students: A review of literature. *College Student Journal, 49,* 195–198.

LoSchiavo, F. M., Shatz, M. A., & Poling, D. A. (2008). Strengthening the scholarship of teaching

and learning via experimentation. *Teaching of Psychology, 35,* 301–304. http://dx.doi.org/10.1080/00986280802377164

Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology, 38,* 94–97. http://dx.doi.org/10.1177/0098628311401587

MacKenzie, A. H. (2009). Preparing high school students for college science classes. *The American Biology Teacher, 71,* 6–7. http://dx.doi.org/10.2307/27669354

Martella, R. C., Nelson, R., & Marchand-Martella, N. E. (1999). *Research methods: Learning to become a critical research consumer*. Needham Heights, MA: Allyn & Bacon.

Martin, R. R., & Srikameswaran, K. (1974). Correlation between frequent testing and student performance. *Journal of Chemical Education, 51,* 485–486. http://dx.doi.org/10.1021/ed051p485

McDaniel, M. A., Wildman, K., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory & Cognition, 1,* 18–26. http://dx.doi.org/10.1016/j.jarmac.2011.10.001

Messling, P. A., III, & Dermer, M. L. (2010). Increasing students' attendance at lecture and preparation for lecture by allowing students to use their notes during tests. *The Behavior Analyst Today, 10,* 381–390. http://dx.doi.org/10.1037/h0100678

Michael, J. (1991). A behavioral perspective on college teaching. *The Behavior Analyst, 14,* 229–239.

Neef, N. A., McCord, B. E., & Ferreri, S. J. (2006). Effects of guided notes versus completed notes during lectures on college students' quiz performance. *Journal of Applied Behavior Analysis, 39,* 123–130. http://dx.doi.org/10.1901/jaba.2006.94-04

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, 42,* 87–92. http://dx.doi.org/10.1177/0098628314562685

Novak, G. M., Patterson, E. T., Gavrin, A. D., & Christian, W. (1999). *Just-in-time teaching: Blending active learning with web technology*. Upper Saddle River, NJ: Prentice Hall.

Padilla-Walker, L. M. (2006). The impact of daily extra credit quizzes on exam performance. *Teaching of Psychology, 33,* 236–239. http://dx.doi.org/10.1207/s15328023top3304_4

Panek, E. (2014). Left to their own devices: College students' "guilty pleasure" media use and time management. *Communication Research, 41,* 561–577. http://dx.doi.org/10.1177/0093650213499657

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PLoS ONE, 8,* e79774. http://dx.doi.org/10.1371/journal.pone.0079774

Perrin, C. J., Miller, N., Haberlin, A. T., Ivy, J. W., Meindl, J. N., & Neef, N. A. (2011). Measuring and reducing college students' procrastination. *Journal of Applied Behavior Analysis, 44,* 463–474. http://dx.doi.org/10.1901/jaba.2011.44-463

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Shabani, D. B., & Carr, J. E. (2004). An evaluation of response cards as an adjunct to standard instruction in university classrooms: A systematic replication and extension. *North American Journal of Psychology, 6,* 85–100.

Thomas, J. W., Bol, L., & Warkentin, R. W. (1991). Antecedents of college students' study deficiencies: The relationship between course features and students' study activities. *Higher Education, 22,* 275–296. http://dx.doi.org/10.1007/BF00132292

Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *Journal of Experimental Education, 79,* 478–504. http://dx.doi.org/10.1080/00220973.2010.512318

Weinstein, S. E., & Wu, S. (2009). Readiness assessment tests versus frequent quizzes: Student preferences. *International Journal on Teaching and Learning in Higher Education, 21,* 181–186.

Wilder, D. A., Flood, W. A., & Stromsnes, W. (2001). The use of random extra credit quizzes to increase student attendance. *Journal of Instructional Psychology, 28,* 117–120.

Williams, W. L., Weil, T. M., & Porter, J. C. (2012). The relative effects of traditional lectures and guided notes lectures on university student test scores. *The Behavior Analyst Today, 13,* 12–16. http://dx.doi.org/10.1037/h0100713

Wooldridge, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory & Cognition, 3,* 214–221. http://dx.doi.org/10.1016/j.jarmac.2014.07.001

Received July 25, 2016
Revision received November 14, 2016
Accepted November 17, 2016 ■