

Preliminary Analysis Of Interteaching's Frequent Examinations Component In The Community College Classroom

Theresa A. Felderman, Bismarck State College, USA

ABSTRACT

Interteaching has shown to be an effective alternative to traditional lecture in a number of studies, but thorough analyses of its components, including frequent exams, is limited. Research suggests that increasing the frequency of exams may improve student learning. This study assessed the effectiveness of interteaching's frequent exams component on student exam scores in two introductory psychology course sections at a community college. Six unit exams were given to Section 1 and 12 to Section 2. Both sections received identical pretests and posttests. Although statistically insignificant, Section 2 consistently scored better or equal to Section 1 on all exams. Section 2 also produced a greater difference between pretest and posttest scores. Although not reaching traditionally significant levels, the results suggest that students given frequent exams may achieve higher exam scores and increase retention of overall course material.

Keywords: Interteaching; Frequent Examinations; Community College Instruction; Education

INTRODUCTION

Interteaching is an instructional program comprised of a number of components including advanced completion of preparation (prep) guides, group discussions, reinforcing contingencies, and frequent exams. The basic interteaching process begins with assigned reading and a complementary instructor-generated prep guide for student completion. In the following class session, the students form dyadic partnerships to complete the "interteach." The interteach is defined as a "mutually probing, mutually informing conversation between two people" (Boyce & Hineline, 2002, p. 221). Students discuss their independently completed prep guides during the interteach, while the instructor moves around the classroom answering questions and helping the students with understanding the prep guide concepts. Immediately following the interteach, students complete written evaluation of the interteaching session identifying difficult concepts and topics they would like explained by the instructor. From these interteaching records, the instructor develops a clarifying lecture given the following class meeting. The final component is the frequent exam as the primary measure of learning in the interteaching classroom (for a more detailed description of interteaching, see Boyce & Hineline, 2002; Saville, Lambert, & Robertson, 2011). Interteaching component analyses research has focused on the inclusion of prep guides (Saville, Zinn, & Serdikoff, 2008), reinforcing properties of quality points (Saville & Zinn, 2009), and the types of prep guides used, specifically instructor-created prep guides or prep guides where students generate their own questions (Cannella-Malone, Axe, & Parker, 2009). These studies have begun to uncover the underlying mechanisms making interteaching effective (Saville et al., 2011). Nonetheless, many questions remain unanswered regarding the variables that play key roles in the process (Saville et al., 2011). The current interteaching process consists of five to six exams throughout the semester and considered frequent compared to the traditional midterm and final exam schedule, used by many college instructors (Boyce & Hineline, 2002). Interteaching, as a method of college instruction, uses frequent exams and relies on the premise that frequent exams enhance student learning (Boyce & Hineline, 2002), yet thorough component analysis is lacking. To understand interteaching, it is essential to understand its components, including frequent exams (Saville et al., 2011).

Educators have long struggled with the role of testing in the educational system, specifically measurement, assessment, types of testing, the purpose of testing, increasing retention, and many variables that impact testing procedures, and ultimately how the role of testing is related to student learning (Peckham & Roe, 1977; Roediger & Karpicke, 2006a). One such variable extensively explored is frequent exams and the effects increased testing has on student achievement (Basol & Johanson, 2009). The question of whether students perform better on frequent exams throughout the semester has long been debated and researched (Basol & Johanson, 2009; Peckham & Roe, 1977). Nonetheless the research is inconclusive. Some research points to enhanced academic performance (Dustin, 1971; Fulkerson & Martin, 1981; Gaynor & Millham, 1976; Keys, 1934; Kika, McLaughlin, & Dixon, 1992; Kling, McCorkle, Miller, & Reardon, 2005; Miller, 1987; Turney, 1931). Testing can be powerful and can improve long-term retention of knowledge (Roediger & Karpicke, 2006b). A phenomenon in which taking a test can improve future performance and retention is referred to as the *testing effect* (Roediger & Karpicke, 2006a). The testing effect suggests that exams can be a powerful instructional tool to improve and assess learning (Roediger & Karpicke, 2006a). Although there are several studies supporting increasing the frequency of testing, several other studies uncovered little or no relationship between frequent exams and increased student exam performance (Beaulieu & Utecht, 1987; Beaulieu & Zar, 1986; Schunert, 1951; Selakovich, 1962; Standlee & Popham, 1960). With the variable research regarding the frequency of exams, as well as the limited component analyses in the current interteaching research (Cannella-Malone et al., 2009; Saville & Zinn, 2009; Saville et al., 2008), the effectiveness of the frequent exams interteaching component requires thorough analysis. The purpose of the present study was to determine whether the number of exams given throughout the semester during the interteaching process affects student exam scores.

METHODS

Participants and Setting

Fifty-two undergraduate students enrolled in two sections of an introductory psychology course at Bismarck State College participated in this study. Each section was comprised of 26 participants (Section 1 = 8 males, 18 females; Section 2 = 7 males, 19 females). Due to the inability to randomly assign participants to groups, the first day of class self-reported demographic information was gathered including (a) sex, (b) age, (c) year in school, (d) cumulative grade point average higher than 2.0, (e) employment status, (f) previous psychology courses taken, and (g) involvement in unrelated school activities. The data analyses showed no statistically significant differences between sections on any of the demographic variables. Both sections attended each class session in the same classroom, with the same instructor throughout the semester.

Materials and Procedure

The students of two course sections met for 50 min, 3 days per week for the total of 16 weeks throughout the semester. Section 1 met at 9:00 a.m. and Section 2 at 10:00 a.m. The instructor taught both sections using Boyce and Hineline's (2002) interteaching process. A Nonequivalent Groups Design model (NEGD) was used to assess the effects of the number of exams given throughout the semester on overall exam scores (see Table 1).

Table 1: Nonequivalent Groups Design Model

Section	Order of Variable Presentation		
	Pretest Variable	Independent Variable	Posttest Variable
Section 1	Pretest	6 Exams	Posttest
Section 2	Pretest	12 Exams	Posttest

Note. Interteaching was the method of instruction used throughout the semester in both sections.

Section 1 and Section 2 were exposed to the same interteaching process, lectures, textbook, and supplementary handouts. The PowerPoint® presentations were developed from the completed interteaching records. The prep guides were available to both sections, created by the instructor prior to the interteaching sessions, and consisted of essay and short answer questions based on the course objectives. Students also were able to view the presentations, prep guides, and course grades on the instructor's course companion website. Both sections were involved in the interteach and lectures on identical days, but Section 2 met 6 additional days throughout the semester.

to take the six additional unit exams. Section 1 took only six unit exams, whereas Section 2 took 12, thus Section 1 did not meet on 6 designated days during the semester. No additional assignments were given to Section 1 during the designated non-meeting days.

A pretest was given to both sections during the first class meeting. The 40-point pretest consisted of a variety of questions similar to that of the final exam or posttest. Each section's pretest was graded, but did not factor into total points for the course. However, correct answers earned on the pretest were added as bonus points at the end of the semester. Bonus points did not change any of the students' individual unit exam scores, including the posttest scores, but were intended to motivate students to complete the pretest thoroughly and accurately.

Section 1 was given six exams worth 40 points each and Section 2 was administered 12 exams worth 20 points. Section 1's exams covered two chapters and Section 2's exams covered one chapter's material. The same chapters were reviewed in both sections throughout the semester. Identical questions were used on all exams, but were given based on the material covered to that point. At the end of the semester during final exam week, the two sections received the same 40-point cumulative final (posttest) consisting of true-or-false, short answer, and multiple-choice questions similar to the questions on the unit exams and the pretest. A study guide was provided in both sections to assist with preparation for the posttest (cumulative final exam).

Interobserver Agreement

Two independent observers graded 30 % of the exams and the level of interobserver agreement was calculated by taking "the number of agreements (i.e., the number of times each grader obtained the same exam score on each question) divided by the number of agreements and disagreements and multiplied by 100" (Saville et al., 2006, p. 53). IOA scores across the six exams ranged from 75 to 92%, with a mean score of 82%.

RESULTS

No significant differences were found between sections on all demographic measures: (a) sex, $\chi^2(1, N = 52) = .09, p = .76$; (b) age, $\chi^2(1, N = 52) = 1.09, p = .30$; (c) year in school, $\chi^2(1, N = 52) = 2.56, p = .11$; (d) cumulative grade point average, $\chi^2(1, N = 52) = .35, p = .55$; (e) employment status, $\chi^2(1, N = 52) = .51, p = .65$; (f) previous psychology courses taken, including high school, $\chi^2(1, N = 52) = .32, p = .58$; (g) involved in unrelated school activities, $\chi^2(1, N = 52) = .50, p = .48$.

A series of independent-samples *t* tests were conducted to determine statistically significant differences between Section 1 and Section 2 on all exams given throughout the semester, including the pretest and posttest (i.e., cumulative final). Both Figure 1 and Figure 2 show Section 2 scoring better or equal to Section 1 on all mean percentage exam scores, though none reached traditional levels of significance. Figure 1 shows the mean percentage exam scores for six unit exams taken by both sections throughout the semester. The 12 unit exam scores for Section 2 were combined by units to correspond with the six exams given to Section 1.

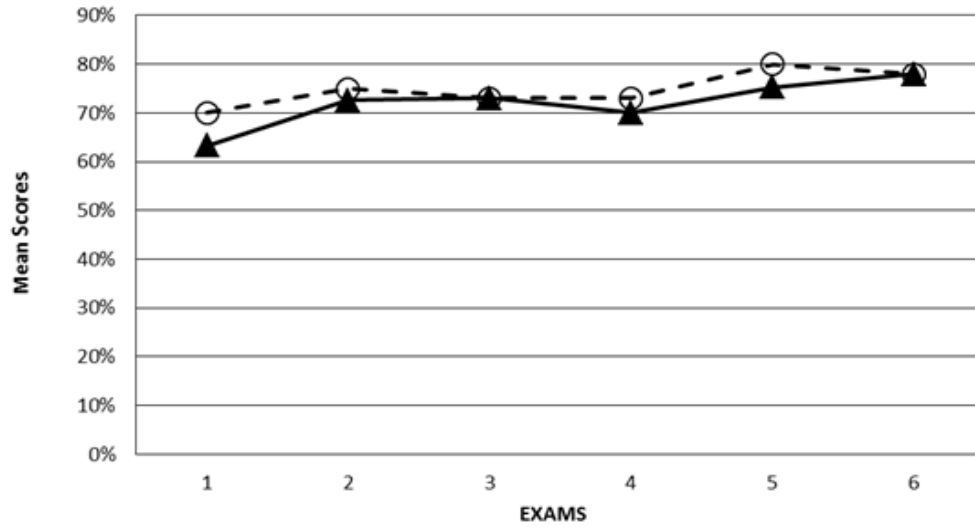


Figure 1. Mean percentage exam scores for six unit exams.
 Filled triangles with solid line = Section 1 mean percentage exam scores.
 Open circles with dashed line = Section 2 mean percentage exam scores.

The student exams from Section 1 were divided into questions from each of the 12 units that Section 2 was tested and the mean percentage exam scores for each of the 12 units are displayed in Figure 2. The students in Section 2 scored higher than Section 1 on the pretest, five of 12 unit exams (1, 2, 4, 8, and 10), and the posttest, but none of the differences reached traditional levels of significance (all $ps > .10$). The mean percentage exam scores of the remaining seven exams (3, 5, 6, 7, 9, 11, and 12) were the same for both sections, thus resulting in no statistically significant differences (all $ps > .65$). Section 2 scored an average of 2.8 percentage points higher on all exams when combined into the six exam scores of Section 1 and 2.9 percentage points higher when Section 1's exam scores were separated into the same 12 unit exams of Section 2.

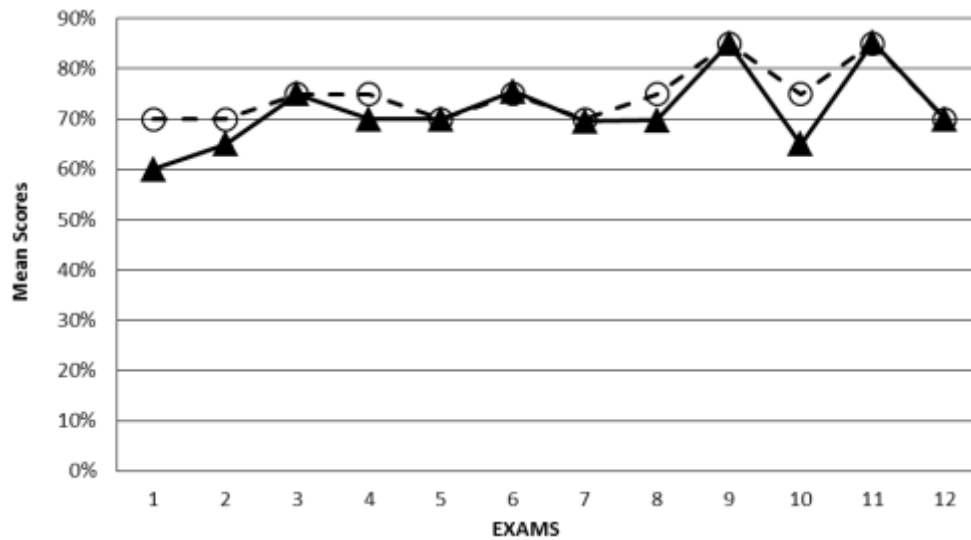


Figure 2. Mean percentage exam scores for twelve unit exams.
 Filled triangles with solid line = Section 1 mean percentage exam scores.
 Open circles with dashed line = Section 2 mean percentage exam scores.

As shown in Figure 3, students in both sections correctly answered a greater percentage of posttest questions correctly compared to the mean pretest exam scores, with a 7.5% average increase in scores on the posttest. Specifically, students in Section 1 answered correctly 69% of the posttest questions compared to 63% of pretest questions. Section 2 answered 73% correctly on the posttest and 64% on the pretest. Section 2 scored four percentage points higher than Section 1 on the posttest compared to one percentage point difference on the pretest. Additionally, Section 2's posttest exam scores were nine percentage points higher than the pretest scores compared to a six percentage point gain by Section 1.

Simple main effect tests were conducted on the posttest to assess differences among sections at low (one *SD* below the mean), medium (mean), and high (one *SD* above the mean) values on the covariate and a *p* value of .017 (.05/3) was required for significance for each of the tests. The simple main effects test was not significant for all values of the covariate $F(1, 49) = .43, p = .51$, suggesting no differences between sections on the posttest as a function of the pretest. Given no significance among means, pairwise comparisons among sections were not examined.

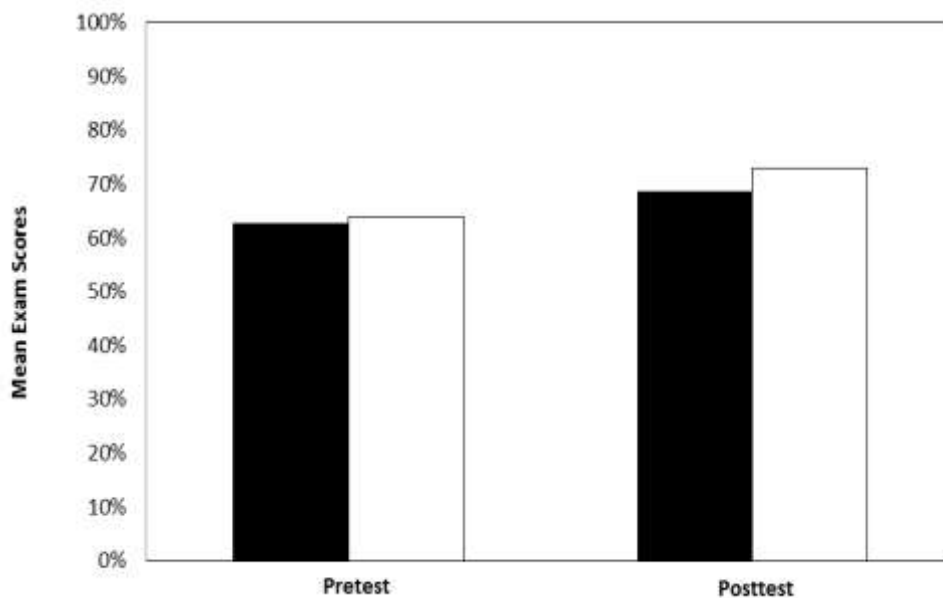


Figure 3. Mean percentage of correctly answered pretest and posttest questions.
Filled bars = Section 1. Open bars = Section 2.

CONCLUSION

Although the results of this study were statistically insignificant, Section 2 scored better or equal to the Section 1 on all exams, including on the posttest, suggesting the differences though small may lead to important practical differences affecting overall course performance. A larger increase from the average pretest to posttest scores was also observed in the students in Section 2 than in Section 1, suggesting that frequent exams may lead to better retention of course material in the community college setting. The results may have practical significance for students and educators. Students may achieve higher exam scores in the frequent exams condition, thus influencing overall course grades. Frequent exams may result in an increase in overall course performance, which may be important to some, if not many students. These results suggest that increasing the number of exams in the interteaching process may help with achieving learning outcomes.

A number of limitations should be considered when evaluating the results of this study. First, although demographic information was gathered and no significant differences existed between the sections, it is possible that demographic or other preexisting differences between groups existed, potentially affecting the results. However, the

results of this research were similar to earlier interteaching studies increasing the reliability and minimizing the effects of any preexisting differences.

Second, although the instructor used the same materials and presented lectures based on identical course content, experimenter expectancies may have influenced the results due to the dual role of instructor and researcher. The material covered in the lectures was available in all courses via the course companion website, but the potential for differences during the lectures still existed. Furthermore, an increasing trend in the exam scores across the semester suggests carryover effects. The levels of achievement may have passed over into the achievement levels of the next exam as the exam scores appeared to be continually increasing. However, the data also indicated that exam scores consistently were higher in the frequent exams condition.

Third, although the prospective power analysis to compute required sample size using $\alpha = .05$, a large effect size estimation, and a power of 0.80 results in an estimate of 21 participants required per group and 26 students participated in each section, this sample size may have been too small to produce an effect. A type II error, due to small sample size, may have occurred. This error may have affected the conclusion that increasing the number of exams given during interteaching failed to produce a significant difference in student exam scores.

Lastly, due to the methodology where Section 2 took 12 exams and Section 1 took six exams with identical questions, it is possible the insignificant differences found could have been the result of students in Section 2 sharing exam content with Section 1 over the exams they had already taken. However, this scenario seems unlikely due to the consistent upward trend and higher average exam scores in Section 2. Additionally, although the results provide some evidence that taking 12 exams versus six exams may result in better exam scores, it is impossible to determine whether doubling the number of exams was responsible for the differences or if any number beyond six may improve exam performance.

To address the limitations of this study, future research should examine interteaching's frequent exams component in other settings and with other populations and larger sample size. Replications should include undergraduate and graduate classrooms, varying numbers of students, various course levels, and instructional modalities (e.g., distance education). Also, examination of the frequent exams component in a setting with a higher degree of experimental control, such as a laboratory setting, would be beneficial to establish the validity of the observation.

This study was a preliminary analysis; thus, future research should consider several manipulations of the independent variable to include various numbers of exams (e.g., compare courses with three exams to six exams) in order to assess the effectiveness of the frequent exams component. Additionally, given the positive effect that frequent exams may have on student learning, the effects daily testing may have on exam scores should be assessed. Lastly, although this research focused on interteaching's frequent exams component, student study behavior may have also changed. Researchers should examine the affects frequent exams may have on student studying behavior. Students may modify behavioral repertoires based on the schedule of exams and other contingency arrangements. This will add to the theoretical foundation and understanding of interteaching's frequent exams component.

In sum, these results demonstrate doubling the number of exams produced slightly higher exam scores and retention, suggesting the importance of frequent exams in interteaching. However, this study was a preliminary analysis with inconclusive results, providing the opportunity for additional examination of interteaching's frequent exams component. Additional research is necessary to gain a better understanding of interteaching, potentially leading to modifications in the process enhancing student learning and retention.

AUTHOR INFORMATION

Dr. Theresa Felderman is an Associate Professor of Psychology at Bismarck State College (BSC), Bismarck, ND. She also is an adjunct psychology professor for Minot State University (MSU), Minot, ND. Prior to her employment with BSC, Theresa worked for developmental disability, educational, and medical agencies as a behavioral therapist and consultant. She also has worked as a Vocational Counselor and instructor for Sitting Bull College, Fort Yates, ND and adjunct instructor for the University of Mary, Bismarck, ND.

Theresa earned her doctorate degree in psychology from Northcentral University, Scottsdale, AZ. She also has a master's degree in applied psychology with specialization in behavioral analysis from St. Cloud State University, St. Cloud, MN and an undergraduate degree in social and behavioral science from University of Mary, Bismarck, ND.

Correspondence concerning this article should be addressed to Theresa A. Felderman, Department of Psychology, Bismarck State College, 1500 Edwards Avenue, Schafer Hall 302, Bismarck, ND.

Email: theresa.felderman@bismarckstate.edu

ACKNOWLEDGEMENTS

This paper is based on a dissertation submitted by the author in partial fulfillment of the requirements for the Doctor of Philosophy of Psychology at Northcentral University. Thank you Bettina Shapira, Ph.D., Henry Riegler, Ph.D., and Bryan Saville, Ph.D. for their mentoring and guidance throughout this process. A special thanks to Anita Wirtz, Ph.D. for her assistance with data collection and analysis.

REFERENCES

1. Basol, G. & Johanson, G. (2009). Effectiveness of frequent exams over achievement: A meta-analysis study. *International Journal of Human Sciences* [Online], 6(2), 99-121. Retrieved from <http://www.insanbilimleri.com/en>
2. Beaulieu, R. P., & Utecht, K. M. (1987). Frequently administered formative tests and student achievement. *Journal of Instructional Psychology*, 14(4), 195-200.
3. Beaulieu, R. P., & Zar, M. C. (1986). The effects of examination frequency on student performance. *Journal of Instructional Psychology*, 13(2), 81-89.
4. Boyce, T. E., & Hineline, P. N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangement in the college classroom. *The Behavior Analyst*, 25, 215-226.
5. Cannella-Malone, H. I., Axe, J. B., & Parker, E. D. (2009). Interteach preparation: A comparison of the effects of answering versus generating study guide questions on quiz scores. *Journal of Scholarship and Teaching*, 9(2), 22-35.
6. Dustin, D. (1971). Some effects of exam frequency. *The Psychological Record*, 21, 409-414.
7. Fulkerson, F. E., & Martin, G. (1981). Effects of exam frequency on student performance, evaluations of instructor, and test anxiety. *Teaching of Psychology*, 8(2), 90-93. doi:10.1207/s15328023top0802_7
8. Gaynor, J., & Milham, J. (1976). Student performance and evaluation under variant teaching and testing methods in a large college course. *Journal of Educational Psychology*, 68(3), 312-317. doi:10.1037/0022-663.68.3.312
9. Keys, N. (1934). The influence of learning on retention of weekly as opposed to monthly tests. *The Journal of Educational Psychology*, 25(6), 427-436. doi:10.1037/h0074468
10. Kika, F. M., McLaughlin, T. F., & Dixon, J. (1992). Effects of frequent testing of secondary algebra students. *Journal of Educational Research*, 85(3), 159-162. doi:10.1080/00220671.1992.9944432
11. Kling, N., McCorkle, D., Miller, C., & Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business*, 81(2), 67-72. doi:10.3200/JOEB.81.2.67-72
12. Miller, F. (1987). Test frequency, student performance and teacher evaluation in the basic marketing class. *The Journal of Marketing Education*, 9(2), 14-19.
13. Peckham, P. D., & Roe, M. D. (1977). The effects of frequent testing. *Journal of Research and Development in Education*, 10(3), 40-50.
14. Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181-210.
15. Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.
16. Saville, B. K., Lambert, T., & Robertson, S. (2011). Interteaching: Bringing behavioral education into the 21st century. *The Psychological Record*, 61, 153-166.
17. Saville, B., & Zinn, T. (2009). Interteaching: The effects of quality points on exam scores. *Journal of Applied Behavior Analysis*, 42(2), 369-374. doi:10.1901/jaba.2009.42369.

18. Saville, B., Zinn, T., Neef, N., Van Norma, R., & Ferreri, S. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis*, 39, 49-61.
doi:10.1901/jaba.2006.42-05
19. Saville, B., Zinn, T., & Serdikoff, S. (2008). *Interteaching: The impact of preparation guides*. Unpublished raw data.
20. Schunert, J. (1951). The association of mathematical achievement with certain factors resident in the teacher, in the teaching, in the pupil, and in the school. *Journal of Experimental Education*, 19, 219-238.
21. Selakovich, D. (1962). An experiment attempting to determine the effectiveness of frequent testing as an aid to learning in beginning college courses in American Government. *The Journal of Educational Research*, 5, 178-180.
22. Standlee, L. S., & Popham, W. J. (1960). Quizzes' contribution to learning. *Journal of Educational Psychology*, 51, 322-325.
23. Turney, A. H. (1931). The effect of frequent short objective tests upon the achievement of college students in educational psychology. *School & Society*, 33, 760-762.