The Relation between GPA and Exam Performance During Interteaching and Lecture

Bryan K. Saville, Derek Pope, Jacob Truelove, & Jacob Williams

James Madison University

Abstract

A growing number of studies have shown that interteaching produces better student-learning outcomes than more traditional teaching methods such as lecture. But to date, no studies have directly examined whether interteaching is differentially effective for students who typically earn either good or bad grades in their courses. In this study, we alternated between interteaching and lecture several times during the semester and measured the exam performance of students with low, moderate, and high grade point averages. We found that interteaching improved the exam scores for students in each of these groups but that the improvement was most noticeable for students with low and moderate GPAs. Thus, interteaching may be especially useful for students who typically earn lower grades in their courses.

Keywords

interteaching, grade point average, exam scores, behavioral education

Interteaching is a multi-component teaching method that has its roots in the behavior-analytic tradition, a tradition in which most behavior is considered to be largely (but not solely) a function of its consequences presented in context (e.g., Hineline, 1980; Skinner, 1953). A typical interteaching session proceeds as follows (for more detail, see Boyce & Hineline, 2002; Saville, Lambert & Robertson, 2011). The instructor constructs a preparation (prep) guide consisting of various questions, the purpose of which is to guide students through a reading assignment. The instructor distributes the prep guides several days early, and students answer the questions before coming to class. Once in class, students hear a brief lecture that covers certain prep-guide questions from the previous class period (see below). After the lecture, students form pairs and spend the rest of the class session discussing their answers to the prep-guide questions they completed for class. During the discussions, the instructor moves among the pairs, answering questions and guiding students' discussions. Once students have completed the discussions, they complete a record sheet, on which they list, among other information, any prep-guide questions they would like the instructor to review. The instructor then uses the information from the record sheets to prepare a lecture that targets the prep-guide questions that students had trouble understanding. The lecture begins the next class period and precedes discussion of the next prep guide.

Bryan K. Saville, Derek Pope, Jacob Truelove, and Jacob Williams, James Madison University. Derek Pope is now a doctoral student in the Department of Psychology at Auburn University.

Correspondence concerning this article should be addressed to Bryan K. Saville, Department of Psychology, James Madison University, 91 E. Grace Street – MSC 7704, Harrisonburg, VA. Email: savillbk@jmu.edu

Since Boyce and Hineline's (2002) introduction of interteaching, researchers have begun to examine its efficacy relative to more traditional teaching methods. In general, researchers have found that interteaching improves student-learning outcomes when compared to more traditional teaching methods (Arntzen & Hoium, 2010; Saville, Zinn, & Elliott, 2005; Saville, Zinn, Lawrence, Barron, & Andre, 2008; Saville, Zinn, Neef, Van Norman, & Ferreri, 2006; Scoboria & Pascual-Leone, 2009). For example, in the first experimental analysis of interteaching, Saville et al. (2005) found in a lab-based study that students in an interteaching condition performed significantly better on a quiz taken after a 1-week delay than students in either a lecture condition or a reading condition. In two subsequent studies, Saville et al. (2006) found that students in a graduate-level special education course (Study 1) and students in two sections of an undergraduate psychological research methods course (Study 2) earned higher exam grades, on average, following interteaching sessions than they did following lectures. In addition, a majority of students in the Saville et al. (2006) studies reported that they preferred interteaching to lecture (see also Goto & Schneider, 2009; Scoboria & Pascual-Leone, 2009).

Although the preceding results suggest that interteaching may improve student-learning outcomes on average (i.e., at the class level), researchers have not systematically examined whether it works better for students who typically earn good or bad grades in their courses (although see Saville et al., 2006, Study 2, for some individual performance data). For example, it may be that students who typically do well in lecture-based courses might benefit even more from the structured format that interteaching introduces. Conversely, interteaching may be more beneficial for students who tend to perform poorly in traditionally taught courses, which often do not produce optimal study habits (see,

e.g., Michael, 1991), or for "middle-of-the-road" students who might need just a bit more structure to improve their performance.

To date, a handful of studies have examined whether behavior-analytic teaching methods are differentially effective for students who tend to do poorly or well in their courses. In one relevant study, Born, Gledhill, and Davis (1972) assigned students in an undergraduate psychology of learning course to one of four conditions: personalized system of instruction (PSI; Keller, 1968), modified PSI, lecture, or a condition in which students rotated among the three teaching methods (lecture followed by exposure to the other PSI conditions). Born et al. found a main effect for teaching method, with students in the PSI, modified PSI, and rotating conditions scoring significantly higher on their exams and earning higher final course grades than students in the lecture condition. Born et al. also examined students in the rotating condition to see how they compared to the rest of the class when they switched from lecture to the PSI conditions. They found that students from the rotating condition who scored high under the lecture condition (i.e., their scores on the first midterm ranked in the top half of the class) had similar rankings following exposure to the PSI methods. In contrast, students from the rotating condition whose scores on the first midterm fell in the lower half of the class improved their class rankings when they switched to the PSI methods. Born et al. concluded that PSI might be especially effective for students who typically perform poorly in their classes.

In another study on PSI, Born and Whelan (1973) examined in two introductory psychology courses the percentage of unit tests that students with low and high GPAs (the lower and upper quartile of class, respectively) attempted and passed during the semester.[1] Born and Whelan found that the percentage of "passes" for students with high GPAs remained relatively constant or increased only slightly during the semester. Students with low GPAs, however, saw a dramatic increase in their percentage of passes; by the end of the semester, they were passing unit tests at the same rate as the students with high GPAs.

As noted above, interteaching has its roots in behavior analysis and thus shares features with other behaviorally oriented teaching methods, including PSI (see Boyce & Hineline, 2002). In accord, it seems probable that interteaching might also have differential effects on students who typically do poorly or well in their courses. The purpose of the present study, therefore, was to expose students to both interteaching and lecture and see whether those with low and high GPAs performed differently under the two teaching conditions. To gain additional information, we also examined the performance of students with moderate GPAs.

METHOD

PARTICIPANTS

The participants were 73 students (19 men, 54 women) at James Madison University (JMU), the majority of whom were juniors and seniors enrolled in an undergraduate psychology of learning course taught by the first author. The class met for 50 min on Mondays, Wednesdays, and Fridays.

MATERIALS AND PROCEDURE

During the first week of class, we distributed an information sheet on which students provided us with the following demographic information: (a) gender, (b) age, (c) cumulative GPA, (d) year in school, (e) number of psychology courses already completed, (f) number of credit hours taken during the semester, and (g) employment status. Two students (one man, one woman) did not provide us with their cumulative GPAs, so we removed them from the sample. We then separated the remaining 71 students into low- (n = 24; 10 men, 14 women), moderate- (n = 24; 6 men, 18 women), and high-GPA (n = 23; 3 men, 20 women) groups by conducting a tertiary split on their self-reported GPAs. The average GPA for students in each group was as follows: low GPA = 2.86 (out of 4.00), moderate GPA = 3.22, and high GPA = 3.67.

During the semester, we alternated six times between lecture and interteaching, following the general method described by Saville et al. (2006, Study 2). On lecture days (of which there were 16 during the semester), students heard a 45- to 50-min lecture provided by the first author in which he discussed course concepts that were displayed on PowerPoint© slides. In an attempt to control for possible differences in attendance on lecture and interteaching days, we distributed during the lectures a course roster on which students signed their names (on interteaching days, students submitted a record sheet on which they wrote their names; see below). Such a procedure has been shown to increase class attendance and course performance, even if there are no explicit consequences for attending (Shimoff & Catania, 2001).

Under the interteaching condition (which also occurred on 16 class days), students completed an instructor-prepared prep guide prior to each class period. Each prep guide consisted of eight to 12 items, some of which contained two or more related questions, and required students to define, analyze, and apply course concepts.[2] At the start of each class period, the instructor gave a 15- to 20-min lecture that targeted those prep-guide items from the previous class period that students found confusing (see below). Following the lecture, students formed pairs and spent the remaining class time discussing the prep guides they had completed for the day. During the discussions, the instructor and two undergraduate teaching assistants (TAs) moved among the groups, answering questions and guiding the discussions. Once students had finished their discussions, they completed a record sheet on which they listed their partner's name, how long the discussion lasted, how well the discussion went (on a 7-point scale), reasons why it went poorly or well, which prep-guide items were difficult to understand and why,

1 In PSI courses, students must "master" each unit of material before they can move on to the next unit. At the end of each unit, they take a test, which can be taken as many times as needed to meet the "mastery" criteria (e.g., 90%).

2 The following are examples of (a) definitional, (b) analysis, and (c) application questions that appeared on the prep guides: (a) What is a schedule of reinforcement? (b) With regard to operant conditioning and the three-term contingency, do rules typically function as discriminative stimuli, behavior, or consequences? (c) Imagine you are a store manager who wants to get his or her employees to approach more customers. How might you use the Premack principle to increase the likelihood that your employees will approach more customers? Readers interested in seeing the prep guides used in this study can contact the first author. For another sample prep guide, see Boyce and Hineline (2002, Appendix A).

Table 1. Demographic information for low-, moderate-, and high-GPA students[1,2]

	Low	Moderate	High
n	24	24	23
Sex			
Men	10	6	3
Women	14	18	20
Age	20.8	20.4	20.2
Grade Point Average	2.86	3.22	3.67
Year in School			
Sophomore	0	2	4
Junior	16	17	16
Senior	8	4	3
Number of Psychology Courses	8.1	7.0	7.2
Credit Hours	14.5	16.3	15.8
Employment Status			
Yes	11	8	7
No	13	15	16

[1] Although we conducted statistical analyses on each of these measures, we do not report the details here. Other than GPA, the groups were not significantly different (all ps > .05) on any of the measures.
[2] A few students did not report certain demographic information. Thus, the totals sometimes do not equal the total number of students in each group.

and which items the instructor should review. The instructor then used the information provided on the record sheets to construct a brief lecture that began the next class period and preceded the next pair discussion.

After each unit of information (usually five or six lectures or interteaching sessions), students took a 50-point exam. Each exam consisted of approximately 20 items, most of which were short-answer or fill-in-the-blank questions. The items were based on concepts covered in the lectures (during the lecture condition) or on the prep guides (during the interteaching condition) and required students to solve problems, apply information, and show higher-level comprehension of the material. There were six exams during the semester, three of which followed lectures (Exams 1, 3, and 5), and three of which followed interteaching (Exams 2, 4, and 6).

To ensure reliability in grading, the TAs graded the exams. For each exam, one TA graded all of the exams; the other TA graded a subset of 20 (28%) randomly chosen exams. We calculated interobserver agreement by dividing the number of exam items on which the TAs agreed by the total number of items (agreements and disagreements) and multiplying by 100%. In cases where the TAs disagreed, they discussed the item and came to agreement on the final grade. The mean agreement score across all six exams was 92% (range = 84-98%).

RESULTS

First, we compared the low-, moderate-, and high-GPA groups on their self-reported demographic information. In general, the students in the different groups were similar to one another on each measure (see Table 1). We also examined attendance across the semester. Under the lecture condition, students attended 92% of the classes, on average; under the interteaching condition, students attended 93% of the classes, on average. Thus, it is unlikely that preexisting demographic differences or differences in attendance under the two teaching conditions contributed significantly to our results.

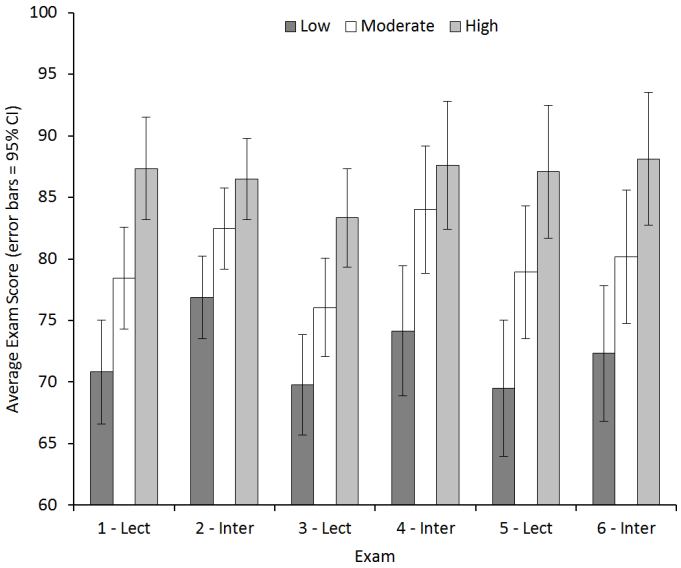

Figure 1. Shows the average exam scores and 95% confidence intervals for students in the low- (dark gray), moderate- (white), and high-GPA (light gray) groups. Students heard lectures (Lect) prior to Exams 1, 3, and 5, and participated in interteaching (Inter) prior to Exams 2, 4, and 6.

Next, we examined students' overall scores on each of the six exams. As shown in Figure 1, students typically performed better on exams following interteaching (Exams 2, 4, and 6) than they did on exams following lectures (Exams 1, 3, and 5). Specifically, across the three interteaching exams, the average score for the entire class was 81.0%, whereas the average score on the three lecture exams was 77.8%, a difference of just over three percentage points.

The pattern of improved exam scores under interteaching was also evident to varying degrees within each of the three GPA groups (Figure 1). Students in the high-GPA group had average exam scores that were 1.5 percentage points higher under interteaching (87.4%) than under lecture (85.9%). With the exception of Exam 3, which followed lecture and was slightly lower than the other exam scores, students in the high-GPA group tended to perform relatively well regardless of the teaching method in effect. Students with low and moderate GPAs also had average exam scores that were higher under interteaching (74.7% and 82.2%, respectively) than under lecture (70.0% and 77.5%, respectively). In contrast with the high-GPA group, though, the low- and moderate-GPA groups had scores that varied much more with changes from lecture to interteaching. Specifically, students in the low- and moderate-GPA groups had exam scores that increased almost five percentage points, on average, following interteaching sessions.

Across the semester, these differences translated into fewer students earning low letter grades and more students earning high letter grades under the interteaching condition (see Figure 2). Specifically, there were approximately 50% fewer Fs and Ds

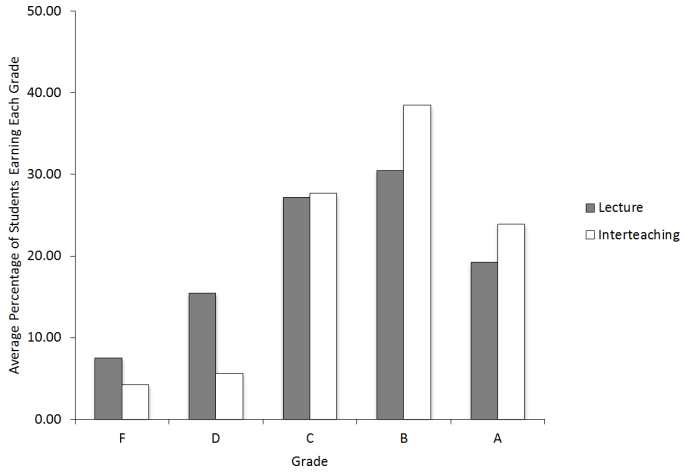

Figure 2. Shows the distribution of grades under the lecture and interteaching conditions. The gray bars represent the average percentage of students receiving each grade on the three lecture exams (Exams 1, 3, and 5), and the white bars represent the average percentage of students receiving each grade on the three interteaching exams (Exams 2, 4, and 6).

and 25% more Bs and As on the interteaching exams than on the lecture exams.

DISCUSSION

The purpose of this study was to examine how low-, moderate-, and high-GPA students performed under two different teaching conditions: lecture, the most common pedagogical method used in college classrooms (Benjamin, 2002), and interteaching, a new teaching method with its roots in behavior analysis (Boyce & Hineline, 2002). In sum, we observed that exam grades were higher, on average, for each group following interteaching sessions. We also observed that students with low and moderate GPAs had grades that varied more and were consistently higher (about five percentage points, on average) under the interteaching condition. Finally, we found no differences in average attendance under the two teaching conditions.

Our observation that students in general had higher exam scores following interteaching sessions is consistent with earlier research. Saville et al. (2005), for example, found that students who participated in a mock interteaching session earned quiz scores 1 week later that were significantly higher than students who were exposed to the same material via a traditional lecture. Similarly, Saville et al. (2006, Study 2) compared interteaching to lecture in two sections of an undergraduate research methods course and found that interteaching produced exam scores that were about 10 percentage points higher, on average, than exam scores following lectures. Finally, Scoboria and Pascual-Leone (2009) compared interteaching and lecture in two large sections of an undergraduate abnormal psychology course and observed that students who participated in interteaching sessions scored higher on measures of critical and analytical thinking than students who heard lectures over the same material. Together, these results suggest that interteaching generally has a positive impact on student-learning outcomes when compared to more traditional lecture-based teaching methods.

We also found that students in all three GPA groups tended to earn higher exam grades after they had participated in interteaching sessions. Thus, interteaching seems to improve average exam scores (cf. Saville et al., 2006) by positively impacting the performance of students who fall at all points along the academic-achievement continuum.

There were, however, group differences in how much exam performance changed as a function of teaching method. Specifically, the difference in exam scores under interteaching and lecture was less noticeable for students with high GPAs. This may be for two reasons. First, because these students were earning high grades to begin with, there may have been little room for improvement (i.e., there was a ceiling effect). Second, it may be that students who typically perform well in their courses are already emitting the types of behaviors interteaching affects—behaviors such as studying before class, discussing course material with classmates, and asking their instructors for clarification, each of which likely has a positive effect on student-learning outcomes. Although we did not query our students about their specific activities under the different teaching conditions, if high-GPA students in our study were engaging in similar behaviors during lectures, it may not be surprising that their exam grades did not vary considerably. At the very least, though, our results show that interteaching does not harm the performance of students who typically earn high grades in their courses and that it might even boost their performance a bit.

Conversely, we found that students in the low- and moderate-GPA conditions had interteaching exam scores that were almost five percentage points higher—equivalent to half a letter grade—than their lecture exam scores. Most likely, students who do not typically perform well in their classes (i.e., low-GPA students) and students who sometimes perform well in their classes (i.e., moderate-GPA students) are not consistently engaging in the types of behaviors that result in positive academic outcomes. When exposed to the contingencies of interteaching, however, these students begin to emit behaviors that promote relatively higher performance. Although interteaching may not turn poorly performing students into students who consistently earn high grades, it nevertheless may be most useful (in terms of relative improvement) for students who typically do not perform exceptionally well in their courses.

More broadly, our results coincide with previous studies that examined the course performance of low- and high-GPA students under different behavior-analytic teaching methods (Born et al., 1972; Born & Whelan, 1973). In general, these studies found that behavioral teaching methods had the biggest impact on students who had performed poorly in their previous courses.

Finally, we observed that average attendance under the two teaching conditions was similar (i.e., over 90%). Thus, although recording student attendance has been shown to increase both attendance and course performance (Shimoff & Catania, 2001), it is unlikely that students in our study did better under the interteaching condition simply because they attended class more often. Nevertheless, it would be interesting to replicate our study and see what happened to students' course performance if the contingencies for attending lectures were removed. It may be that interteaching exerts part of its positive effect by increasing attendance, which then results in better course performance.

Although our observations suggest that interteaching is an effective teaching method, especially for students with lower

GPAs, our study lacked some experimental controls. We alternated treatments several times and noted fairly consistent changes with each manipulation, but it is important to remember that the course material was also changing along with our manipulations. Thus, any changes in exam performance we observed might have been impacted by the different material that students were studying along the way or by the particular sequence of treatments we used in our research design. Future research could address this issue at least two ways: by including a counterbalanced control condition (e.g., across two semesters or across two sections of the same class during the same semester, see Saville et al., 2006, Study 2) or by presenting the treatment conditions in a random (or quasi-random) order. In addition, it is possible that students' self-reported demographic information was not entirely accurate. One way to address this concern would be by acquiring (if possible) official university records. Regardless, we hope that continued analysis will further elucidate the extent to which interteaching improves performance compared to more traditional teaching methods.

REFERENCES

Arntzen, E., & Hoium, K. (2010). On the effectiveness of interteaching. *The Behavior Analyst Today, 11,* 155-159.

Benjamin, L. T., Jr. (2002). Lecturing. In S. F. Davis & W. Buskist (Eds.). *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 57-67). Mahwah, NJ: Erlbaum.

Born, D. G., Gledhill, S. M., & Davis, M. L. (1972). Examination performance in lecture-discussion and personalized instruction courses. *Journal of Applied Behavior Analysis, 5,* 33-43.

Born, D. G., & Whelan, P. (1973). Some characteristics of student performance in PSI and lecture courses. *The Psychological Record, 23,* 145-152.

Boyce, T. E., & Hineline, P. N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst, 25,* 215-226.

Goto, K., & Schneider, J. (2009). Interteaching: An innovative approach to facilitate university student learning in the field of nutrition. *Journal of Nutrition Education and Behavior, 41,* 303-304.

Hineline, P. N. (1980). The language of behavior analysis: Its community, its function, and its limitations. *Behaviorism*, *8*, 67-86.

Keller, F. S. (1968). Good-bye teacher… *Journal of Applied Behavior Analysis, 1,* 79-89.

Michael, J. (1991). Behavioral perspectives on college teaching. *The Behavior Analyst, 14,* 229-239.

Saville, B. K., Lambert, T., & Robertson, S. (2011). Interteaching: Bringing behavioral education into the 21st century. *The Psychological Record, 61,* 153-166.

Saville, B. K., Zinn, T. E, & Elliott, M. P. (2005). Interteaching vs. traditional methods of instruction: A preliminary analysis. *Teaching of Psychology, 32,* 161-163.

Saville, B. K., Zinn, T. E., Lawrence, N. K., Barron, K. E., & Andre, J. (2008). Teaching critical thinking in statistics and research methods. In D. S. Dunn, J. S. Halonen, & R. A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 149-160). Malden, MA: Wiley-Blackwell.

Saville, B. K., Zinn, T. E., Neef, N. A., Van Norman, R., & Ferreri, S. J. (2006). A comparison of interteaching and lecture in the college classroom. *Journal of Applied Behavior Analysis, 39,* 49-61.

Scoboria, A., & Pascual-Leone, A. (2009). An 'interteaching' informed approach to instructing large undergraduate classes. *Journal of the Scholarship of Teaching and Learning, 9,* 29-37.

Shimoff, E., & Catania, C. A. (2001). Effects of recording attendance on grades in introductory psychology. *Teaching of Psychology, 28,* 192-195.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.