

INTERTEACHING: THE EFFECTS OF QUALITY POINTS ON EXAM SCORES

BRYAN K. SAVILLE AND TRACY E. ZINN

JAMES MADISON UNIVERSITY

Although previous studies have found interteaching to be an effective alternative to traditional methods of instruction, few studies have examined which of its components contribute to its effectiveness. In the current study, we examined whether manipulating quality points had an effect on our students' exam scores. In two sections of an undergraduate general psychology course, we used interteaching but alternated between quality points and no quality points several times throughout the semester; we also counterbalanced the order of presentation across sections. We found that quality points did not have an effect on exam scores.

DESCRIPTORS: college instruction, exam performance, interteaching, quality points

Interteaching, a method of classroom instruction that has its roots in behavior analysis (Boyce & Hineline, 2002), attempts to capitalize on well-established behavior-analytic principles. But in contrast with earlier behavioral methods of classroom instruction (e.g., Keller, 1968; Skinner, 1968), interteaching may offer more flexibility for instructors. In brief, interteaching requires students to complete a preparation (prep) guide before each class that consists of questions designed to guide them through a specified reading assignment. In class, students form pairs and spend time discussing the material on the prep guide. While students discuss the prep-guide items, the instructor moves around the room, answering students' questions and facilitating their discussions. After the discussions, students fill out a record sheet that the instructor uses to construct a brief clarifying lecture that begins the next class period and precedes the students' pair discussions for that day. Students also receive a small number of participation points (i.e., 10% of a student's course grade) for taking part in the

pair discussions and a small number of quality points when they and their discussion partners both do well on certain exam questions (see below; see also Boyce & Hineline and Saville, Zinn, Neef, Van Norman, & Ferreri, 2006, for a more detailed discussion of interteaching).

To date, two published studies have shown that interteaching may be more effective than traditional methods (e.g., lecture) at improving student learning outcomes. In one study, Saville, Zinn, and Elliott (2005) randomly assigned participants to one of four conditions—interteaching, lecture, reading, or control—and found that participants in the interteaching condition performed significantly better on a short, multiple-choice quiz given 1 week later than did participants in the other three conditions. Saville et al. (2006) subsequently compared interteaching to lecture in a graduate-level special education course and in an undergraduate research methods course. They found that interteaching produced better exam scores and that students generally preferred interteaching. Although these studies suggest that interteaching might be an effective method of instruction, we are aware of no published studies that have examined which of the several components of interteaching contribute to its effectiveness.

As mentioned above, a primary component of interteaching is the pair discussion, in which

We thank Jennifer Burnett and Brad Kincheloe for their help with data collection.

Address correspondence concerning this article to Bryan K. Saville or Tracy E. Zinn, Department of Psychology, MSC 7704, James Madison University, Harrisonburg, Virginia 22807 (e-mail: savillbk@jmu.edu or zinnnte@jmu.edu).

doi: 10.1901/jaba.2009.42-369

students spend class time discussing their answers to items contained in a prep guide. To improve the quality of these discussions—or more specifically, to ensure that students take the time to teach one another as effectively as possible—Boyce and Hineline (2002) introduced the concept of quality points. Quality points refer to a cooperative contingency in which part of a student's exam grade depends on how well his or her partner performed on certain exam questions. Specifically, if a student and his or her partner both do well on an essay question that they discussed together in class, each receives a small number of points toward his or her course grade. For example, if an essay question is worth five points and both students earn four or five points (i.e., an A or B) on that question, each earns an additional number of points toward his or her course grade. But if one or both students earn fewer than four points on that question, neither receives quality points. Boyce and Hineline suggested that quality points across all exams should account for approximately 10% of a student's course grade.

Although many studies have suggested that the addition of an explicit cooperative contingency improves various measures of performance (e.g., Johnson, Maruyama, Johnson, Nelson, & Skon, 1981), it is not known whether quality points would have the same positive effect. Therefore, the purpose of the present study was to examine the extent to which quality points affected exam scores in a group of college students.

METHOD

Participants

Participants were 44 undergraduate students (16 men, 28 women) in two sections of an introductory psychology course. The students' median age was 18 years, and all were classified as either freshmen ($n = 33$) or sophomores ($n = 11$). There were 22 students (6 men, 16 women) in the first section (SEC 1), which met from 9:30 a.m. to 10:45 a.m. on Tuesdays and

Thursdays, and 22 students (10 men, 12 women) in the second section (SEC 2), which met from 10:00 a.m. to 10:50 a.m. on Mondays, Wednesdays, and Fridays. The first author was the instructor for SEC 1, and the second author was the instructor for SEC 2.

Materials and Procedure

Because we could not randomly assign participants to the different sections, during the first week of class we collected the following self-reported demographic data: (a) cumulative grade point average, (b) number of credit hours taken during the semester, (c) whether students were currently employed, (d) whether students were currently involved with significant others, and (e) whether students were currently members of a fraternity or sorority. These data helped us to determine the extent to which students in the two sections were similar prior to our experimental manipulation.

The general method for this study followed that of Saville et al. (2006, Study 2). During each class, we divided students into pairs by asking them to "find someone you have not worked with yet." (On rare occasions when there was an odd number of students in class, we allowed one group to have 3 students.) Students were free to choose their own partners, with only one constraint: They could not work with the same partner more than three times during the semester. After finding a partner, students spent approximately two thirds of the class time (i.e., 50 min for SEC 1, 30 min for SEC 2) discussing items on their prep guides. During this time, the instructor moved among the pairs, answering questions and facilitating discussion. After the discussions, students took approximately 5 min to complete record sheets that provided the instructor with feedback regarding students' understanding of the material. The instructor then used this information to construct a clarifying lecture that began the next class session, lasted approximately one third of the class time (i.e., 20 min for SEC 1, 15 min for SEC 2), and preceded the pair

discussion for that day. For participating in the pair discussions, students received participation points that across the semester totaled 10% of their overall course grades (Boyce & Hinline, 2002).

After each unit of information, students from both sections took the same 25-point exam. Each exam consisted of two five-point essay questions and several other objective questions (e.g., fill in the blank, short answer) that required students to define concepts, apply information, and show higher level comprehension of the information covered in the prep guides and presented during the clarifying lectures. Students took a total of six exams during the semester.

To examine the effects of quality points on exam scores, we used an alternating treatments design (Barlow & Hayes, 1979), switching between quality points and no quality points several times during the semester. In addition, we counterbalanced across sections, such that while quality points were in effect for one section, they were not in effect for the other section. More specifically, the quality points contingency was in place for students in SEC 1 on Exams 2, 4, and 6 and for students in SEC 2 on Exams 1, 3, and 5. The addition of quality points to each exam worked as follows: If, on a given essay question, both students who discussed that question in class earned either four or five points, each received three additional quality points toward his or her course grade. But if one or both students received fewer than four points, neither received quality points for that question. Thus, because each exam contained two essay questions, students could earn zero, three, or six quality points toward their course grades. Overall, quality points accounted for approximately 8% of each student's final course grade. We described the quality points contingency in our course syllabus. Therefore, students knew when the contingency was in effect; we did not, however, inform them of the overall purpose of

the study until the end of the semester. On the last day of class, we informed students of the purpose of the study, at which time each signed a consent form that allowed us to use their data.

Interobserver Agreement

Two graduate teaching assistants (GTA) who were naive to the purpose of the study independently graded 7 of the 22 exams from each section (i.e., 32% of each of the six exams). To ensure that the grading of one GTA did not influence the other, the GTAs placed the scoring for the exams on separate sheets of paper. We used a fairly stringent criterion when determining agreements and disagreements in grading: An agreement occurred only when the GTAs computed exactly the same overall score on an exam. We then calculated the level of interobserver agreement by taking the number of agreements divided by the number of agreements and disagreements and converting this ratio to a percentage. Agreement scores across the six exams ranged from 70% to 96%, with a mean score of 88%. Most often, the exam scores from each GTA were within one point of each other, and disagreements typically occurred on essay questions when the GTAs were a half point apart in their scoring. When there were disagreements, the GTAs subsequently discussed their grading and came to an agreement regarding the final exam score.

RESULTS AND DISCUSSION

One student in SEC 2 did not provide us with any demographic data. Of the remaining 43 students, 1 from SEC 2 did not list a cumulative grade point average, and 1 from SEC 2 did not provide information on his involvement with a significant other or with a fraternity. Thus, the following demographic comparisons are based on the remaining data. We found no differences between SEC 1 and SEC 2 on any of the self-reported demographic measures: (a) cumulative grade point average, $t(40) = 0.81$, $p = .43$; (b) number of credit

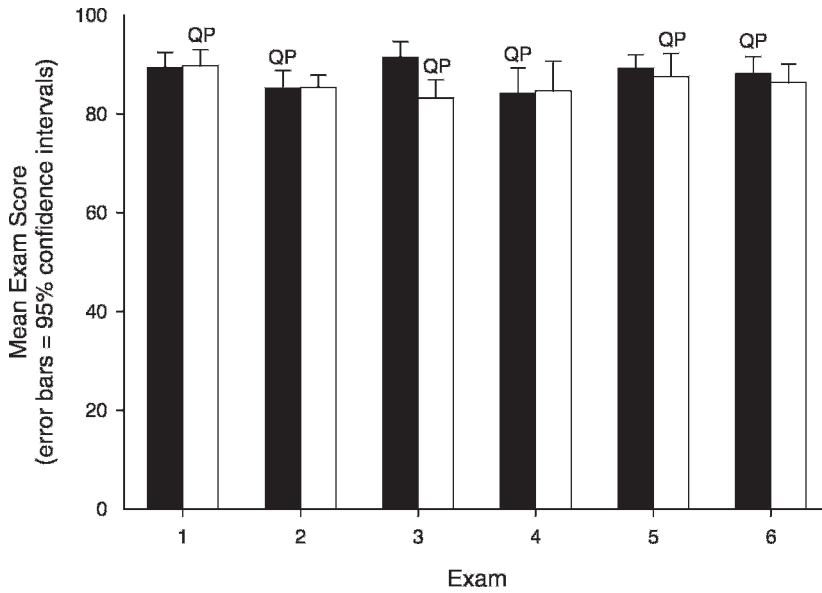


Figure 1. The mean scores for SEC 1 and SEC 2 on each of the six exams. Error bars represent 95% confidence intervals. Filled bars indicate the scores for SEC 1, and open bars indicate the scores for SEC 2. QP indicates which section had the quality points contingency in effect on each exam.

hours taken during the semester, $t(41) = 0.04$, $p = .97$; (c) employment status, $\chi^2(1, N = 43) = 1.10$, $p = .30$; (d) involvement with a significant other, $\chi^2(1, N = 42) = 0.06$, $p = .81$; and (e) involvement with a fraternity or sorority, $\chi^2(1, N = 42) = 2.69$, $p = .11$. Thus, it is unlikely that preexisting demographic differences between sections contributed greatly to our results.

To determine whether there were significant differences between SEC 1 and SEC 2 on each of the six exams, we conducted a series of independent-samples t tests with a Bonferroni correction ($\alpha = .008$). Figure 1 shows the mean exam scores and 95% confidence intervals for SEC 1 and SEC 2 on each of the six exams. On five of the exams, there was no significant difference between the two sections (all p s $> .45$). There was, however, a significant difference between the sections on Exam 3, $t(42) = 3.49$, $p = .001$. SEC 1 ($M = 91.45$, $SD = 7.28$), for which the quality points contingency was not in effect, had a higher mean exam score than SEC 2 ($M = 83.18$, $SD = 8.39$), which

did receive quality points. Given this overall pattern of results, it is unlikely that the difference we observed on Exam 3 was a function of our manipulation.

There are several possible reasons for our observations. First, the lack of differences between sections may have been due to a ceiling effect (Volkert, Lerman, Troclair, Addison, & Kodak, 2008). Overall, the average exam scores for both sections were relatively high, typically falling somewhere between 85% and 90%, regardless of whether the quality points contingency was in effect. With little room for improvement, it is possible that our quality points manipulation may have been unable to affect exam scores in such a way that a significant difference between sections emerged. Although interteaching seems to produce higher exam scores than more traditional methods of instruction (e.g., Saville et al., 2005, 2006), increasing the difficulty of exam questions may allow future researchers to determine the extent to which quality points contribute to interteaching's effectiveness.

Second, in their description of interteaching, Boyce and Hinline (2002) suggested that quality points should account for approximately 10% of students' overall course grades. In our study, quality points were worth approximately 8% of students' grades. (Our decision to make quality points worth approximately 8% of students' overall course grades was a practical one. If we had made quality points worth 10% of the course grades, the number of quality points on any particular exam would have included fractions.) Although such a difference seems minor, it is possible that this slight reduction in the percentage of points earned via quality points may have affected our results. Specifically, the number of quality points available during the semester may not have been substantial enough to motivate students to engage in high-quality discussions. Thus, future researchers may wish to examine whether manipulating the percentage of points earned through quality points has an effect on measures of student learning.

Third, the inclusion of quality points, as implemented in the present study, may simply not contribute to interteaching's efficacy, or more specifically, to improvements in learning as measured by exam scores. Previous research supports this contention. Saville et al. (2005) compared interteaching to lecture, reading, and control, but did not include quality points in their interteaching condition. Nevertheless, they still observed that students in the interteaching condition performed significantly better on a short multiple-choice quiz taken 1 week later than did students in the other conditions.

Although the inclusion of a cooperative contingency often has positive effects on various measures of performance (e.g., Johnson et al., 1981), numerous studies have shown that delayed consequences have less effect on performance than immediate consequences do (e.g., Chung, 1965; Green & Myerson, 2004). In the present study, students often did not know how many quality points they received until the exams had been graded, which

typically occurred about 1 week later. This delay may have weakened any additional effect that quality points had on their exam scores.

Furthermore, because of the nature of interteaching, it is likely that other components may have exerted a stronger effect in our study. For example, whereas the inclusion of quality points in interteaching creates an explicit cooperative contingency, pair discussion creates an implicit cooperative contingency in which students help one another learn the course material. Thus, the immediate social consequences that students in our study received from their partners and their instructor during pair discussions may have had a greater impact on their exam performances than did delayed quality points. It is also possible, though, that implementation of quality points in another way could potentially have a greater effect on exam scores than the way in which we implemented them. For example, instructors could distribute quality points if students' discussions are "on target" during class (Saville et al., 2006, Study 1). Instructors might also choose to award quality points based on students' reports of how well their pair discussions went (Boyce & Hinline, 2002). Implementing quality points in these ways may have a more powerful effect on learning than did the delayed quality points we tested in the present study.

REFERENCES

- Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*, 199-210.
- Boyce, T. E., & Hinline, P. N. (2002). Interteaching: A strategy for enhancing the user-friendliness of behavioral arrangements in the college classroom. *The Behavior Analyst, 25*, 215-226.
- Chung, S. H. (1965). Effects of delayed reinforcement in a concurrent situation. *Journal of the Experimental Analysis of Behavior, 8*, 439-444.
- Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin, 130*, 769-792.
- Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goals structures on achievement. *Psychological Bulletin, 89*, 47-62.

- Keller, F. S. (1968). Good-bye teacher ... *Journal of Applied Behavior Analysis*, *1*, 79–89.
- 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., 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.
- Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.
- Volkert, V. M., Lerman, D. C., Trosclair, N., Addison, L., & Kodak, T. (2008). An exploratory analysis of task-interspersal procedures while teaching object labels to children with autism. *Journal of Applied Behavior Analysis*, *41*, 335–350.

Received August 30, 2007

Final acceptance March 30, 2008

Action Editor, Mark Dixon