Two-Stage Exams Improve Student Learning in an Introductory Geology Course: Logistics, Attendance, and Grades

Katherine Knierim,[1,a] Henry Turner,[2] and Ralph K. Davis[2]

ABSTRACT

Two-stage exams—where students complete part one of an exam closed book and independently and part two is completed open book and independently (two-stage independent, or TS-I) or collaboratively (two-stage collaborative, or TS-C)—provide a means to include collaborative learning in summative assessments. Collaborative learning has been shown to have positive benefits, including increased student engagement and learning. To try to improve student learning, as measured by improvement in exam scores, two sections of introductory geology were taught using two-stage exams. It was hypothesized that class scores would be higher for semesters using two-stage exams—whether part two was TS-C or TS-I—than for semesters using traditional (T) exams. The median attendance rate was approximately 67% for all testing methods and was significantly greater when exams were TS-C (69%) rather than TS-I (53%). The class score was significantly greater during semesters when exams were TS-C (81%) but was not significantly different between T and TS-I semesters. To assess individual student learning over time, part one of the first exam and part one of the comprehensive final exam were compared. Across the F and D grade ranges, improvement on individual exam scores was significantly greater for the TS-C semester than for the TS-I and T semesters. Student learning, as measured by individual exam scores, improved due to the use of TS-C exams. The improvement in class scores due to the collaborative portion of two-stage exams was independent of increased attendance rates, greater for the lower-achieving students, and not observable if part two of the exam was completed as a take-home exam (TS-I). © *2015 National Association of Geoscience Teachers*. [DOI: 10.5408/14-051.1]

Key words: two-stage exams, collaborative learning, lecture

INTRODUCTION

The benefits of collaborative learning in the university classroom—as part of an active learning environment—include improved attendance, engagement, and student learning (Yuretich et al., 2001; McConnell et al., 2003, 2006; McKeachie and Svinicki, 2006a; Lasry et al., 2008). However, it can be difficult to translate student collaboration to the assessment portion of a course while still ensuring class grades reflect individual achievement. Two-stage, cooperative, or pyramid exams provide a means to include collaborative learning in summative assessment. During two-stage exams, students complete an exam in two stages or parts; part one typically consists of a closed-book exam that students complete independently, and part two consists of an open-book exam that students can complete independently (two-stage independent, or TS-I) or collaboratively (two-stage collaborative, or TS-C; Yuretich et al., 2001; Zipp, 2007; Bloom, 2009; Macpherson et al., 2011). On part two of the exam, students may complete identical test questions, new (and often more difficult) questions, or some combination of the two (Science Education Resource Center, 2011). Generally, the exam parts are weighted so that part one is worth 70%–80% of the grade and students' grades cannot be decreased by completing part two (Yuretich et al., 2001; Bloom, 2009; Science Education Resource Center, 2011).

Why does collaborative learning benefit students? When students collaborate, they may be more motivated to learn because of the community environment created through the interaction (Bransford et al., 2000). Peer learning also provides an opportunity for students to practice new vocabulary and put material into their own words (McKeachie and Svinicki, 2006a), which is essential as students adapt new knowledge to their existing ideas (Bransford et al., 2000). In addition, collaborative learning provides immediate feedback for students, which can be difficult to achieve in large classes, allowing them to fill in knowledge gaps (Cortright et al., 2003). The benefits of active, peer learning can also be extended to assessments if a portion of an exam (e.g., part two of two-stage exams) is completed collaboratively. There is a growing body of research showing the additional benefit of collaborative testing, separate from other in-class active learning techniques (Cortright et al., 2003; Zipp, 2007; Bloom, 2009; Macpherson et al., 2011).

Introductory geology at the University of Arkansas (General Geology, GEOL 1113) is taught as a traditional lecture course with class sizes of approximately 200 students. As is common with large-enrollment university courses (Yuretich et al., 2001), attendance is generally low, many students are unenthusiastic about course content, and classroom management proves to be a constant challenge. Summative assessments are completed in a traditional, closed-book format (T) and using machine scoring, so exams are multiple-choice, true/false, or matching formats that rely mostly on knowledge-based information recall. To try to improve student learning, as measured by improvement in exam scores, two sections of GEOL 1113 were

Received 23 August 2015; revised 6 March 2015; accepted 12 March 2015; published online 13 May 2015.

[1]Environmental Dynamics Program, University of Arkansas, Fayetteville, Arkansas 72701, USA
[2]Department of Geosciences, University of Arkansas, Fayetteville, Arkansas 72701, USA
[a]Author to whom correspondence should be addressed. Electronic mail: katherine.knierim@gmail.com. Tel.: 479-575-3355. Fax: 479-575-3469

TABLE I: Exam procedures for four sections of GEOL 1113 taught by instructor KK.

Semester	Exam Type[1]	Approx. No. Questions (per exam)	Weighting[2]
Fall 2011	T, closed book	50	NA
Spring 2012	T, closed book	50	NA
Fall 2012	TS-C	30	Part one = 75%, part two = 25%
Spring 2013	TS-I	50	Part one = 75%, part two = 25%

[1]Instructor HT proctored T, closed-book exams.
[2]NA = not applicable.

taught using two-stage exams. There is extensive research in geoscience education regarding improved student learning through field trips (Elkins and Elkins, 2007), in-class ConcepTests (McConnell et al., 2006), and lecture tutorials (Yuretich et al., 2001), but two-stage exams may provide another mechanism to engage students in collaborative learning (Zipp, 2007; Bloom, 2009; Macpherson et al., 2011). In addition, efforts were made to create a two-stage exam methodology that could be easily implemented in other large-enrollment courses.

METHODS

Study Setting

The University of Arkansas is a Research I university with 2012–2013 enrollment of 24,537 students, of which 83% were undergraduate students (University of Arkansas, 2014). Traditionally, nonscience majors from the Fulbright College of Arts and Sciences, Walton College of Business, College of Education and Health Professions, and College of Engineering enroll in GEOL 1113 to earn a required laboratory science credit. Sections of GEOL 1113 typically include four summative assessments: three section exams and a comprehensive final exam, although the final exam may or may not be required. Attendance is not mandatory, but an incentive may be offered with a small amount of extra credit. Attendance is taken by different instructors in a variety of ways, but most commonly use sign-in sheets or short "pop" quizzes.

Exam Logistics

Two-stage exams were proctored in two sections of GEOL 1113 during fall 2012 and spring 2013 (Table I). Part one was completed in class, closed book, and was worth 75% of the total exam grade. Part two was completed either in class or as a take-home exam, open book, and was worth 25% of the exam grade. Part two was optional, and a student's score could not be lowered by completing it. In addition, part two consisted of the identical set of exam questions completed for part one; this facilitated machine scoring of the exams, because students were not penalized if they answered a question correctly on part one but incorrectly on part two. Therefore, the score for part two depends on the answers to part one. An example Excel spreadsheet is available online (http://dx.doi.org/10.5408/14-051s1) to provide the equations that automatically score exams if an instructor is able to receive a digital version of student responses to every question on both parts of the exam. The Excel spreadsheet allows instructors to quickly score two-stage exams for large classes.

During fall 2012, part two of the two-stage exams was completed in class, open book, and collaboratively, with students working in groups of three to four (TS-C). The time requirements for completing both parts in class necessitated shortening the exams (Table I). During spring 2013, part two of the two-stage exams was completed as a take-home exam, which was open book, and students had the option of working with their peers (but were not provided a set time and location to work collaboratively). Therefore, spring 2013 is described as TS-I, but some students chose to collaborate with their peers (as evidenced by informal instructor inquiry and observing students immediately working in the hall after part one of the exam).

Statistical Analyses

Exam scores, class scores, and attendance from 20 sections of GEOL 1113 ($n = 3{,}162$) taught between fall 2008 and spring 2013 by two instructors (HT = 15 sections and KK = 5 sections) were aggregated to analyze trends in attendance and scores among semesters with different testing types (i.e., T versus TS-C or TS-I exams). Exam scores were the primary measure of student learning (final class scores were the average of exam scores), similar to other research designs (Zipp, 2007; Macpherson et al., 2011). Although pre- and postsurveys, such as the Geoscience Concept Inventory (GCI; Libarkin and Anderson, 2005), provide a validated means to compare student learning across institutions, using exam scores in this research allowed the plethora of archival data to be utilized. All HT sections were conducted using T exams, and Table I summarizes the exam procedure for four KK sections. It was hypothesized that class scores would be higher for semesters using two-stage exams—whether TS-C or TS-I—than for semesters using T exams. If two-stage exams result in higher class scores simply due to the open-book portion of the test (regardless of the degree of student collaboration), then class scores from both semesters using two-stage exams are predicted to be greater than T exam scores; therefore, it would be difficult to separate the effects of the open-book portion of the test from the effects of student collaboration. Attendance was analyzed in conjunction with class and exam scores to assess possible differences among semesters.

Differences for class scores and attendance rates among exam methods were tested for statistical significance using a nonparametric one-way analysis of variance (Kruskal-Wallis ANOVA on Ranks) at an α of 0.05 in Sigma Plot v. 12.5. Nonparametric statistics were used because most data were not normally distributed, as confirmed by the Shapiro-Wilk Normality Test; introductory geology scores and attendance rates tend to be skewed toward lower values. In the few cases in which data passed the Shapiro-Wilk Normality Test, nonparametric statistical procedures were used to maintain consistent treatment of the data. Post hoc tests to identify

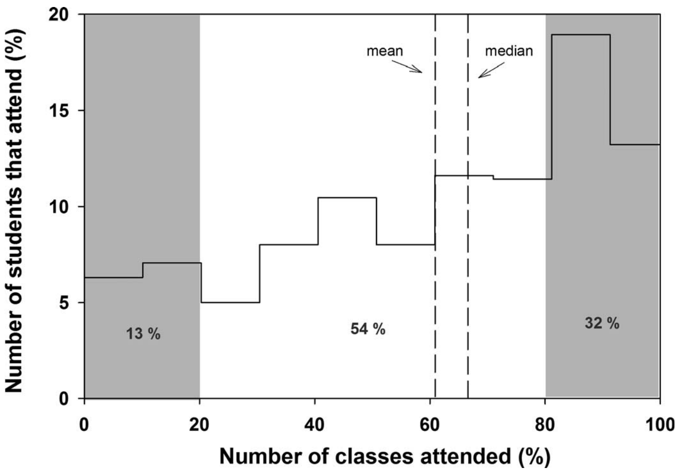

FIGURE 1: Attendance rates for all sections of GEOL 1113, where 1% equals approximately 32 students. Most students (~ 54%) attended 20%–80% of the classes, and the median attendance rate was 67%. Approximately 13% of the student body attended fewer than 20% of the classes. Approximately 32% of the student body attended more than 80% of the classes.

significant differences between groups were completed using Dunn's method.

The authors also hypothesized that two-stage exams—whether TS-C or TS-I—improved individual student learning compared to T exams, as measured by exam scores. To assess individual student learning over time, a more detailed analysis of semesters using two-stage exams was completed, with two semesters of T exams taught by KK serving as the control groups (Table I). Part one of the first section exam was compared to part one of the fourth and comprehensive final exam across grade categories (10-point divisions: A is ≥90%, B is 80%–89%, C is 70%–79%, D is 60%–69%, and F is ≤59%) using the following equation:

$$\Delta S = E1_{P1} - E4_{P1}, \quad (1)$$

where $E1_{P1}$ is the part one, exam one score and $E4_{P1}$ is the part one, final exam score. Part one exam scores measured student achievement when a student worked independently; therefore, only analyzing part one scores answers the question of whether individual student learning improves when two-stage exams are used. Retention of content over time was analyzed by comparing scores from exam one and the final exam, because the final exam included approximately 30% of the same material from the first section exam, with 10–12 questions identical to those in previous exams or study guides (~13%). Equation 1 was modified to use the total exam score for the two control semesters when T exams were used (Table I), because T exams have only one independent part. Statistical significance was tested using a Kruskal-Wallis ANOVA on Ranks to compare ΔS among semesters; statistics were completed in Sigma Plot v. 12.5 at an α of 0.05. In addition, Eq. 1 was only applied to students who completed both exam one and exam four; in spring 2013, the final exam was optional, so fewer students completed exam four than in fall 2012.

RESULTS

Attendance

Median class attendance for all students ($n = 3{,}162$) was 66.7% ± 29.1% and varied by instructor; HT attendance was significantly greater (66.7%) than KK attendance (60.0%), Mann-Whitney U Statistic(3,160) = 843,628, $p < 0.001$. Most students (~54%) attended 20%–80% of the classes throughout the semester, which encompasses the mean and median attendance rates (Fig. 1). The low-attendance students (attending fewer than 20% of the classes) made up approximately 13% of the student body, and the high-attendance students (attending more than 80% of the classes) made up approximately 32% of the student body (Fig. 1). Attendance was significantly and positively—although slightly—correlated with the final class score in GEOL 1113($r^2 = 0.2$; Fig. 2). When the final class score was binned by letter grade, the attendance rate was significantly different among all letter grades, Kruskal-Wallis H Statistic(4, 3,157) = 572, $p < 0.05$, with the lowest median attendance rate (33%) for the F students and the highest median attendance rate (87%) for the A students (Fig. 3).

Attendance rates were not significantly different among testing methods for instructors KK and HT (Table II). However, for only the semesters taught by KK, attendance varied significantly among testing methods (Table II). For example, attendance was significantly higher during the TS-C semester compared to TS-I, $H(2, 777) = 8.4$, $p < 0.05$, but was not significantly different between T and TS-I or TS-C semesters (Table II).

Two-Stage Exams

The median, final class score in GEOL 1113 was 77.6% and varied by test method. The class score was significantly higher during semesters when exams were TS-C than T (taught by either instructor) and TS-I, $H(3, 3{,}158) = 21.5$, $p < 0.05$, but were not significantly different among T and TS-I exams (Table II). Therefore, the class score was higher for the TS-C semester compared to the T and TS-I semesters (Table II). For semesters using two-stage exams and across all exams, the median score on part one was 69% and the median score on part two was 90%.

To better understand the effect of two-stage exams on student learning, four semesters taught by KK were compared, with T semesters serving as the control groups (Table I). The change in score on part one of the exam (or the total exam score for T semesters) between exam one and the final exam (ΔS) was significantly greater during semesters when exams were TS-C rather than T and TS-I, $H(3, 556) = 82.4$, $p < 0.05$, but were not significantly different among T and TS-I semesters (Table II). Therefore, more students showed improvement in their individual exam score between the first exam and the final exam for the TS-C semester compared to the T and TS-I semesters (Fig. 4). In addition, for the semesters when two-stage exams were used, students showed greater improvements between parts one and two of the final exam during the TS-C semester than during the TS-I semester (Fig. 5). Although the improvement was not significant, more students simply chose not to complete part two of the final exam during the TS-I semester (19%) compared to the TS-C semester (4%), as evidenced by students who showed no improvement between parts one and two of the final exam (Fig. 5).

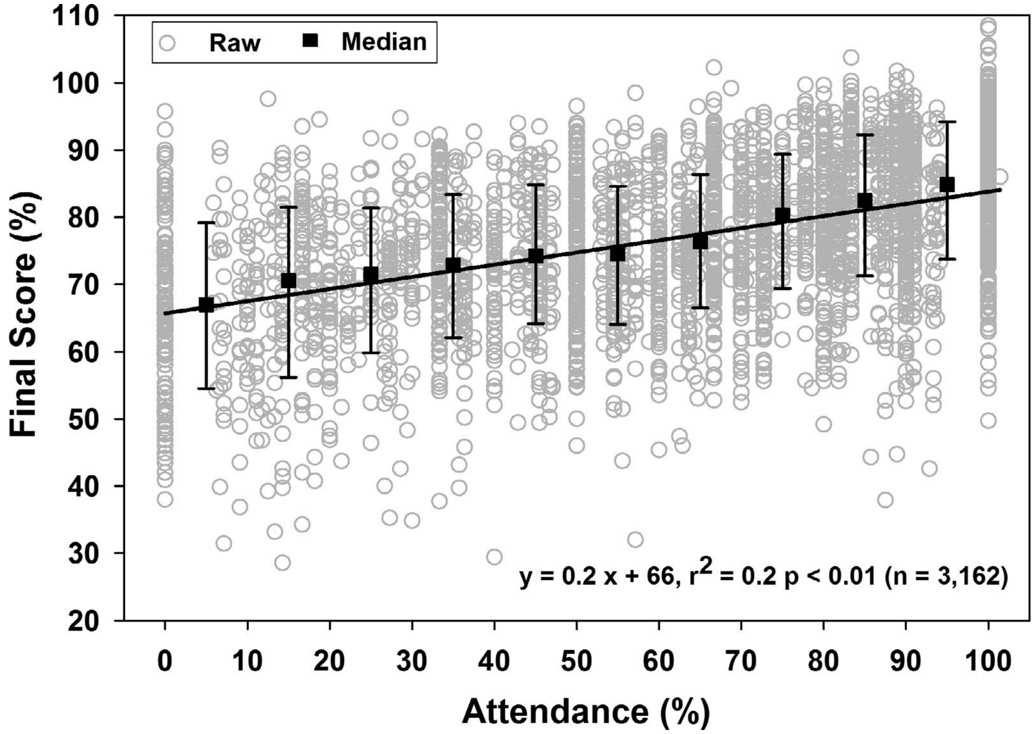

FIGURE 2: Attendance rate versus final class score for all sections of GEOL 1113. Gray circles show all data used to calculate the linear regression, black squares show the median class score for each attendance bin (0%–10%, 11%–20%, etc.), and error bars show one standard deviation.

The difference in the closed-book portion of the exam between exam one and the final (ΔS) can be further compared by binning students by letter grade, where the grade bin is based on the student's score on exam one (Table III). Across all letter grades, improvement in individual exam scores between exams one and four was greatest during the TS-C semester (Table III). ΔS was significantly greater for TS-C than for TS-I and T semesters for the F grade range, $H(3, 144) = 22.3$, $p < 0.05$, and D grade range, $H(3, 135) =$ 32,9, $p < 0.05$ (Table III). At the C grade range, significance in ΔS among semesters changed; ΔS was significantly greater for TS-C than for TS-I and one of the T semesters (but not both) for the C grade range, $H(3, 140) = 17.1$, $p <$ 0.05, and the B grade range, $H(3, 95) = 12.7$, $p < 0.05$ (Table III). In grade category A, the TS-C semester was the only semester in which students improved between exams one and four, although the difference was not significant (Table III). Attendance was not significantly different among testing methods for the F, D, B, and A grade ranges. Attendance did vary among semesters for the C grade range; although the significant difference was only between the two T semesters, $H(4, 140) = 8.8$, $p < 0.05$.

DISCUSSION

The perception of many GEOL 1113 instructors is that class attendance is generally low, because the cavernous auditoriums can appear quite empty on nonexam days. Analysis of attendance rates showed that median attendance was low over 5 y (~67%), but positively, almost one-third of the class attended more than 80% of the class sessions (Fig. 1). As attendance rate was significantly correlated with final class letter grade (Fig. 3), instructors are well served to encourage students to attend class. In large introductory science classes, inspiring and maintaining higher attendance rates can be a challenge (Yuretich et al., 2001), but breaking lectures up into shorter segments and involving active-learning activities in lectures can increase student attention (McKeachie and Svinicki, 2006b) and overall attendance rates (McConnell et al., 2006). Interestingly, for the semesters taught by KK, attendance was highest the semester that TS-C exams were used (69%). Therefore, two-stage exams did not inherently increase class attendance, but when part two of the exam was completed by "forced" collaboration (i.e., establishing a set time and place for students to complete the open-book portion of the exam), attendance was significantly higher. Although it has been shown that collaborative learning increases attendance (Yuretich et al., 2001), the authors are unaware of research that shows collaborative testing—independent of other in-class active learning techniques—improves attendance. Drawing from research on collaborative learning, collaborative testing may create positive attitudes about learning in general or may increase motivation through fear that students will appear ill-prepared in front of their peers, thus increasing class attendance (Bloom, 2009). Perhaps students feel that they are able to "get to know" their peers in class during the collaborative component of two-stage exams—which otherwise does not occur in many large lectures—thus increasing their likelihood of attending class on nonexam days. Without surveying students, the authors can only hypothesize about the cause of increased attendance for the semester taught using TS-C exams.

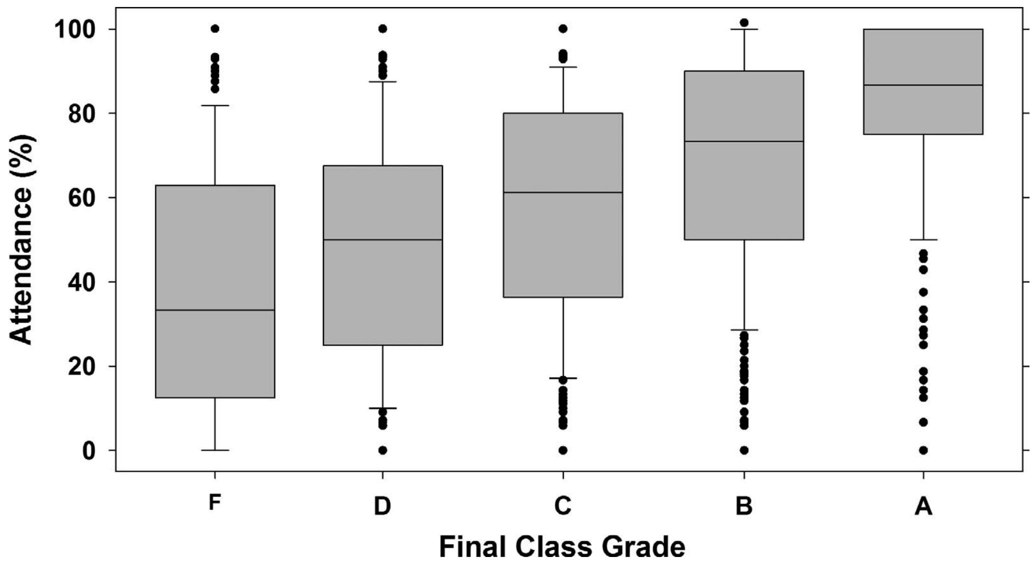

FIGURE 3: Box plots of attendance rate for final class scores, binned by letter grade, for all sections of GEOL 1113. Box boundaries represent the interquartile range, error bars represent the 10th and 90th percentiles, and black circles show data outside of the 10th or 90th percentiles (or statistical outliers).

Final class scores were significantly higher during the semester when TS-C exams were used (81%) but were not significantly different among TS-I and T semesters taught by either instructor. Although attendance was also higher during the TS-C semester (compared to other semesters taught by KK), which could contribute to the higher class scores, the TS-C class scores were significantly higher even compared to semesters taught by HT; as discussed previously, attendance rates were significantly higher for HT (67%) than for KK (60%). Therefore, the TS-C exams likely provided an additional benefit in student learning outside of the observed benefit achieved from higher attendance rates (Fig. 3). Observed increases in student learning using two-stage exams were not simply due to the "bump" in the score from the open-book portion of the exam, because TS-I exam scores were predicted to be greater than T exam scores, which was not the case. Any benefit that the students received in their class score likely came from the truly collaborative component of the two-stage exams, when part two of the exam was completed in class (TS-C). Qualitatively, it was observed during fall 2012 that few students chose not to complete part two of the exam compared to spring 2013, when students completed part two as take-home exam and many students opted not to complete the take-home portion. As students are enrolled in other classes, have work, have families, and attempt to participate in outside social engagements, there may simply be a benefit from providing students with a structured time to complete part two of the two-stage exam. For example, during the final exam (Fig. 5), approximately 19% of students chose to not complete part two during the TS-I semester, compared to 4% during the TS-C semester. Also, perhaps some of the lower-achieving students felt overwhelmed about the exam and did not have the resources to correct their mistakes from part one. No matter the cause, the improvement in student learning, as measured by class scores, was significantly greater during the TS-C semester compared to the TS-I and T semesters. Although results are from only one semester using TS-C exams, based on statistical significance, the collaborative learning component during part two of the exams seems to improve student learning, as measured by exam scores.

To better understand which students benefited from the two-stage exams, exam scores were broken down by letter

TABLE II: Summary results for final class score, attendance rates, and ΔS by testing method.[1] Data are arranged according to statistical groupings, and median values within a column followed by the same superscript letter are not significantly different at the $p < 0.05$ level.

Exam Type	Attendance (%)			Class Score (%)			ΔS (%)		
	Median	Std. Dev.	*n*	Median	Std. Dev.	n	Median	Std. Dev.	*n*
T, HT	67[2]	25	2226	78[a]	11	2226	NA	NA	NA
T, KK	60[a,b]	29	447	78[a]	13	447	3[a]	12	274
TS-C	69[a]	27	165	81	10	165	15	14	165
TS-I	53[b]	29	168	75[a]	11	168	5[a]	12	121

[1]ΔS = Difference in exam score between part one (closed book) of the first exam and part one (closed book) of the final exam; Std. Dev. = standard deviation; NA = not applicable.
[2]Attendance rates were not significantly different among semesters including HT. Therefore, attendance was tested for statistical significance for KK semesters only.

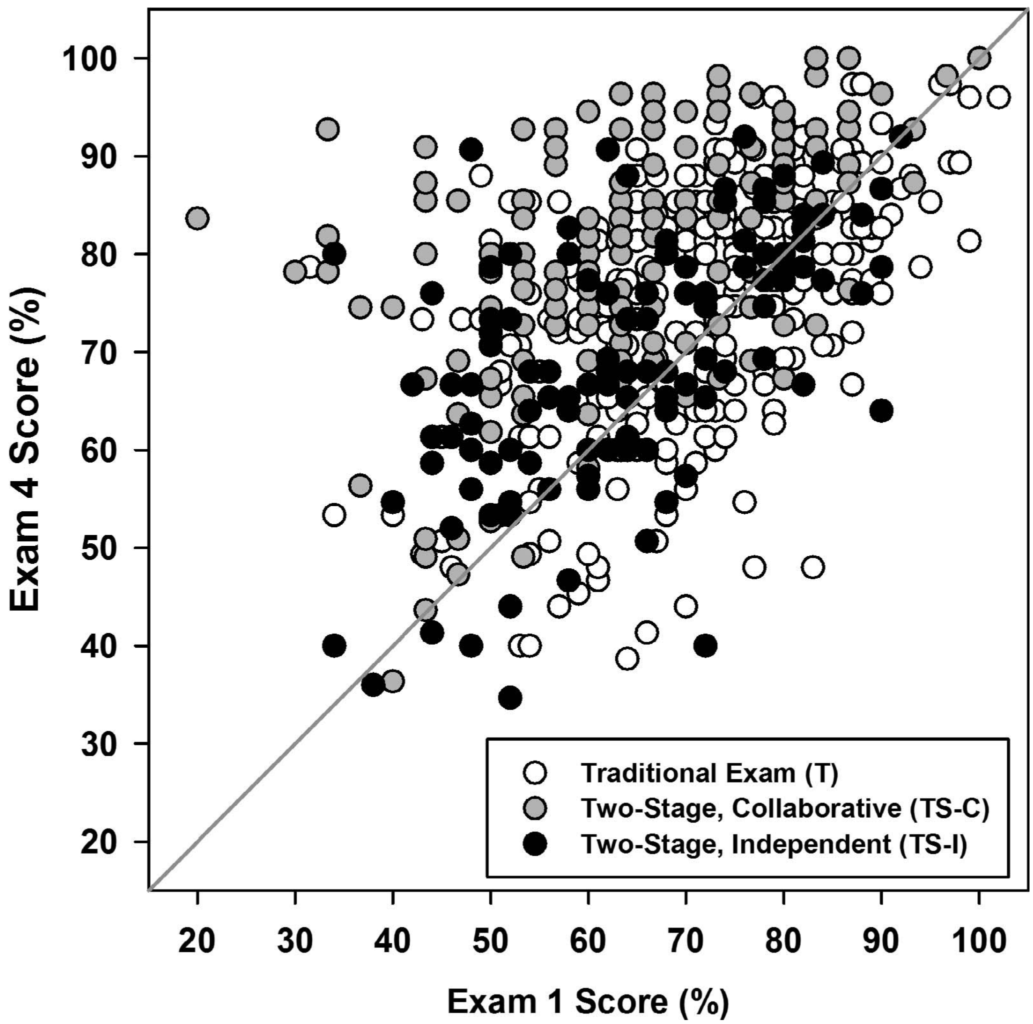

FIGURE 4: Individual student exam scores for exam one versus the final exam (exam four) by test type. A 1:1 line is shown for reference, where students who plot above the line showed improvement from exam one to exam four.

grade category. During the TS-C semester and across the F and D ranges, students significantly improved a median of 23.5% and 16.4%, respectively (compared to TS-I and T semesters), on their individual exam scores between exam one and the final exam (Table III), where approximately one-third of the test material was repeated. As letter grades on part one of exam one increased, the median improvement decreased for the C (12.1%), B (7.2%), and A (0.8%) ranges (Table III). Therefore, the lower-achieving students benefited the most from the collaborative component of TS-C exams, as was found in another study (Macpherson et al., 2011). Finding ways to help lower-achieving students improve is critical in large-format lecture classes—especially considering that some students in the F and D grade categories are statistical outliers in terms of attendance (Fig. 3), which means that some students can have high attendance rates but still fail or barely pass introductory geology. The cause of improvement on part one scores between exams one and four during the TS-C semester is not known. As discussed in Macpherson et al. (2011), perhaps low-scoring students benefit directly from the peer teaching or indirectly from greater motivation to learn the material. Whether the TS-C exams increase scores either directly (through the collaborative learning process) or indirectly (through either positive student motivation because of better attitudes about learning or negative motivation due to fear of appearing ill-prepared), the benefits are clear. Student learning, as measured by individual exam scores on two-stage exams, was significantly greater due to the use of TS-C exams compared to TS-I and T exams, especially for the lower-achieving students.

CONCLUSIONS

TS-C exams significantly increased attendance rates by approximately 16% compared to semesters taught by the same instructor (KK) using TS-I exams and by 9% compared to using T exams (although the difference was not significant). Therefore, two-stage exams did not inherently increase class attendance, but when part two of the exam was completed in class and dominantly through small-group collaboration (i.e., TS-C), then attendance was higher for the entire semester. Final class scores were also significantly greater during the semester when TS-C exams were used. Although attendance was higher during the TS-C semester compared to other semesters taught by KK, which could contribute to the higher class scores, the TS-C class scores

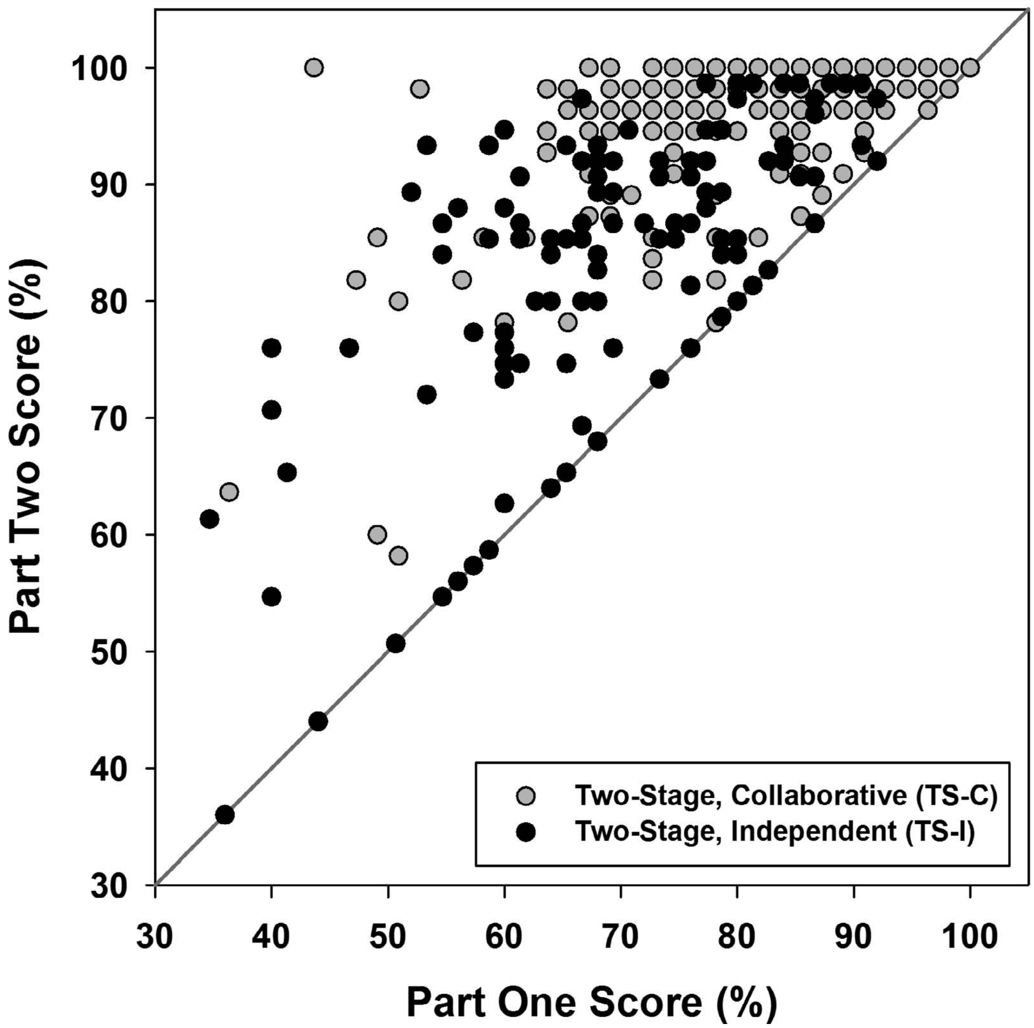

FIGURE 5: Part one versus part two scores for the two-stage final exams. A 1:1 line is shown for reference, and because of the exam scoring, it is not possible for students to decrease their exam scores (i.e., plot below the line). More students during the TS-I semester plotted on the 1:1 line, meaning that students chose not to complete part two of the exam (i.e., the take-home portion).

were significantly greater even compared to semesters taught by HT using T exams; attendance rates were significantly greater for HT (67%) than for KK (60%). Therefore, the TS-C exams likely provide an additional benefit in student learning outside of the observed benefit achieved from higher attendance rates. Lower-achieving students benefited the most from the collaborative component of TS-C exams. For example, F-range students significantly improved by a median of 23.5% on their individual exam scores between exam one and the final exam during the TS-C semester compared to the TS-I (12.0%) and T (14.4%) semesters. If improved exam scores are a measure of student learning, then the exact cause or causes of improvement during the TS-C semester compared to T and TS-I semesters are unknown. Students may benefit either directly from the collaborative learning process during part two of in-class exams or indirectly from changes in motivation; for example, students may be more motivated to

TABLE III. Change in individual exam score from exam one to exam four (ΔS), binned by letter grade on exam one. Median values within a column followed by the same superscript letter are not significantly different at the $p < 0.05$ level.

Grade Bin	F		D		C		B		A	
Exam Type[1]	Median	*n*	Median	*n*	Median	*n*	Median	*n*	Median	*n*
T (F11)	14.7[a]	26	9.0[a]	39	6.3[a,b]	59	−3.0[a]	33	−6.0	8
T (S12)	14.0[a]	21	−0.7[a]	21	−0.3[a]	32	−0.7[a,b]	23	−5.7	12
TS-C	23.5	58	16.4	43	12.1[b]	29	7.2[b]	29	0.8	6
TS-I	12.0[a]	43	5.3[a]	36	2.7[a]	24	−1.0[a]	14	−7.3	4

[1]F11 = fall 2011, S12 = spring 2012.

prepare for TS-C exams so that they are able to share and discuss answers with their peers. No matter the causes, student learning, as measured by individual exam scores, improved due to the use of TS-C exams in introductory geology. The improvement in class scores due to the collaborative portion of two-stage exams was independent of increased attendance rates during the TS-C semester, greater for the lower-achieving students (F and D ranges), and not observable if part two of the exam was completed as a take-home exam. Future research could apply this exam methodology but use the GCI as the measure of student learning. Then, pre- and posttests could be compared to other institutions using T lectures versus in-class active learning activities, which may further elucidate the effects of TS-C exams on individual student learning.

Acknowledgments

Comments from the associate editor and two anonymous reviewers greatly improved this manuscript. The authors thank students at the University of Arkansas who have taken introductory geology.

REFERENCES

Bloom, D. 2009. Collaborative test taking: Benefits for learning and retention. *College Teaching,* 57(4):216–220, doi:10.1080/87567550903218646.

Bransford, J.D., Brown, A.L., and Cocking, R.R., eds. 2000. How people learn: Brain, mind, experience, and school. Washington, DC: Academy Press. p. 386. Available at http://www.csun.edu/~SB4310/How%20People%20Learn.pdf (accessed 15 August 2014).

Cortright, R.N., Collins, H.L., Rodenbaugh, D.W., and DiCarlo, S.E. 2003. Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education,* 27(3):102–108, doi:10.1152/advan.00041.2002.

Elkins, J.T., and Elkins, N.M. 2007. Teaching geology in the field: Significant geoscience concept gains in entirely field-based introductory geology courses. *Journal of Geoscience Education,* 55:126–132.

Lasry, N., Mazur, E., and Watkins, J. 2008. Peer instruction: From Harvard to the two-year college. *American Journal of Physics,* 76:1066–1069, doi:10.1119/1.2978182.

Libarkin, J.C., and Anderson, S.W. 2005. Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education,* 53:394–401.

Macpherson, G.L., Lee, Y.-J., and Steeples, D. 2011. Group examination improves learning for low-achieving students. *Journal of Geoscience Education,* 59:41–45.

McConnell, D.A., Steer, D.N., and Owens, K.D. 2003. Assessment and active learning strategies for introductory geology courses. *Journal of Geoscience Education,* 51:205–216.

McConnell, D.A., Steer, D.N., Owens, K.D., Knott, J.R., Van Horn, S., Borowski, W., Dick, J., Foos, A., Malone, M., and McGrew, H. 2006. Using ConcepTests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education,* 54:61–68.

McKeachie, W.J., and Svinicki, M. 2006a. Active learning. *In* Teaching tips: Strategies, research, and theory for college and university teachers. New York: Houghton Mifflin, p. 213–220.

McKeachie, W.J., and Svinicki, M. 2006b. How to make lectures more effective. *In* Teaching tips: Strategies, research, and theory for college and university teachers. New York: Houghton Mifflin, p. 57–73.

University of Arkansas. 2014. Admissions: Quick facts. Available at http://admissions.uark.edu/1454.php (accessed 15 August 2014).

Science Education Resource Center. 2011. Cooperative exams. Available at http://serc.carleton.edu/NAGTWorkshops/earlycareer2011/cooperative_exams.html (accessed 13 June 2014).

Yuretich, R.F., Khan, S.A., Leckie, R.M., and Clement, J.J. 2001. Active-learning methods to improve student performance and scientific interest in a large introductory oceanography course. *Journal of Geoscience Education,* 49:111–119.

Zipp, J.F. 2007. Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology,* 35:62–76, doi:10.1177/0092055X0703500105.