“Two-stage” group exams can improve student learning

Brett Gilley and **Bridgette Clarkston**

Carl Wieman Science Education Initiative, University of British Columbia, Vancouver Canada
email: bclarkst@zoology.ubc.ca, bgilley@eos.ubc.ca; twitter: @funnyfishes, @modernhydra

Background

Learning through collaboration, even in a testing situation, has many benefits stemming from peer-to-peer interactions. A collaborative test, hereafter called a “two-stage exam”, typically has the following format:[1]

1st Stage:	Students write exam as individuals.
2nd Stage:	Groups of 3–5 students immediately complete a second identical (or very similar) exam. The 2nd Stage typically takes much less time.

Students self-report many benefits of two-stage exams, including: reduced test anxiety[2], greater motivation to study and greater motivation to think critically during a two stage exam[3]. There are studies[4] reporting improved retention when testing using two-stage group exams, however, these studies failed to control for the additional “time-on-task” of a two-stage exam format (in which students are exposed to the same questions twice).

Research Questions

1) Does collaboration during a two-stage exam improve students’ retention of concepts more than a test written individually?

2) What, if any, specific effects does collaboration during a test have on students’ retention of concepts?

Methods

The Course

- Earth and Ocean Sciences non-majors course about natural disasters.
- 2.5 hr classes, 5 days / week, 3 weeks in summer 2012.
- 98 students, 59 % first– and second-year, 41% third-year and above.
- Midterms each worth 30% total; within each: 85% for *individual test,* 15% for *group retest.*
- Study occurred over two midterms, each held on a Friday, with the *learning test* the following Monday.

Experimental Set-up: A Cross-Over Design (Figure 1)

Figure 1 outlines the experiment. We used two-stage exams as described above, with two extra parts:

- ***individual retest:*** Students repeated, as individuals, five 1st stage questions. Acted as the control treatment. Used to make sure students in the individual mode work on questions for the same amount of time.
- ***learning test:*** Individually-written quiz, 10 questions. Measure of students’ retention of concepts.

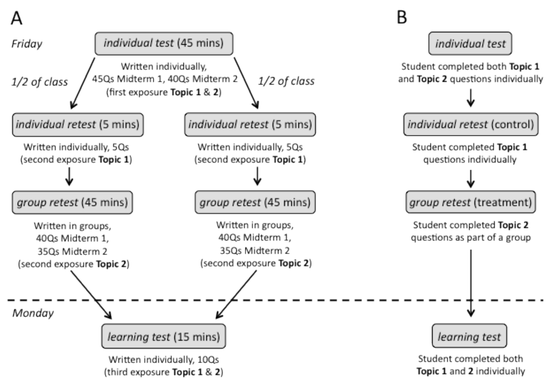

Figure 1: A. Flowchart of the experimental design used for both two-stage midterms. B. An example of how the experimental design was experienced by a given student during each midterm (topic order reversed for one-half of class).

Data Analysis

- Each midterm analyzed separately. We included only scores of students who wrote both *individual test* and *learning test* (midterm 1: n=79; midterm 2: n=71).
- Each student’s individual test score was paired with their learning test score for each of the Topic 1 and Topic 2 questions. Class data set was then divided according to whether students had answered questions from a given Topic during the *individual retest* (i.e., control) or *group retest* (i.e., treatment).
- Percentage learning gain and normalized change (midterm 1: n=67; midterm 2: n=53) were calculated for each student using their baseline *individual test* and the follow–up *learning test* scores.

Results and Discussion

1) *Does collaboration during a two-stage exam increase student’s retention of concepts more than a test written individually?*

Working in groups resulted in significantly greater retention of concepts by students, for both midterms (Table 1 and Figure 2).

Table 1: Class performance (mean % ± standard error) during each stage of the midterms. The differences in student performance between the baseline *individual test* and follow-up *learning test* assessments were compared within and between the *group* and *individual* conditions (p values). Midterm scores are shown for the relevant experimental questions only; scores for the full midterms were 70.5% ± 1.3 for the midterm 1 *individual test*, 88.4% ± 1.3 for the *group retest* and 73.1% ± 1.2 for the midterm 2 *individual test*, 77.5% ± 1.1 for the *group retest.* Effect size calculated using *cohen’s d.*

Condition	*individual test* (IT)	*retest* (ind. or grp.)	*learning test* (LT)	Difference LT – IT	p–value (paired t-test)
Midterm 1					
Group	64.6 ± 1.9	79.2 ± 3.1	77.7 ± 1.7	13.1 ± 2.1	t = 6.27, df = 78, p < 0.0001
Individual	64.8 ± 2.1	66.3 ± 2.1	68.6 ± 2.1	3.8 ± 1.5	t = 2.55, df = 78, p = 0.0127
		Difference between conditions		9.3 ± 2.6	t = 3.64, df = 78, p = 0.0005
			Effect size	0.54 ± 0.16 (SD)	—
			n	79	
Midterm 2					
Group	62.5 ± 2.7	77.4 ± 3.9	75.7 ± 2.8	13.2 ± 2.8	t = 4.67, df = 70, p < 0.0001
Individual	62.6 ± 2.7	64.4 ± 2.3	66.6 ± 2.7	4.0 ± 2.1	t = 1.94, df = 70, p = 0.0561
		Difference between conditions		9.2 ± 3.6	t = 2.54, df = 70, p =0.0132
			Effect size	0.39 ± 0.17 (SD)	—
			n	71	

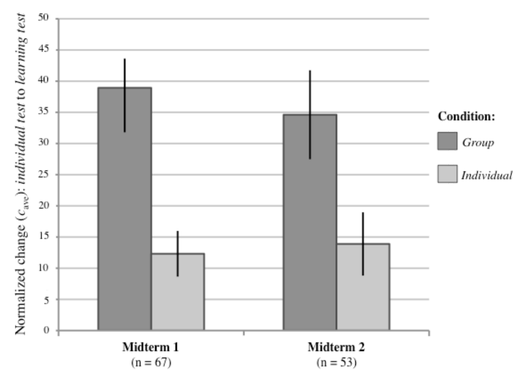

Figure 2: The improvement in student performance was greater for the group condition compared to the individual, measured as average normalized change, cave, for the class between the baseline (individual test) and follow-up (learning test) assessments. Normalized change was calculated for each individual student before determining the mean of the class. Each midterm was analyzed separately. Bars represent standard error.

2) *What, if any, specific effects does collaboration during a test have on students’ retention of concepts?*

The potential gain in retention for each student may be limited by their group’s score (Figure 3).

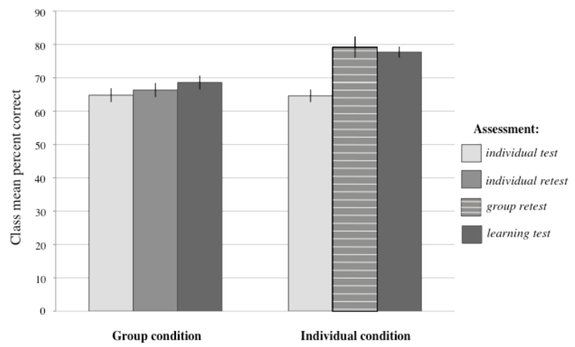

Figure 3: The improvement in student performance on the follow-up learning test was similar to (i.e., not significantly different from) what the groups achieved on the same questions during the group retest. (independent t-test, midterm 1: t = 0.41, df = 38 p = 0.68; midterm 2: t = 0.35, df = 50, p = 0.73) Bars represent standard error.

Results and Discussion *continued*

2) *What, if any, specific effects does collaboration during a test have on students’ retention of concepts?*

When comparing normalized gain by quantiles of the class (based on midterm mark) collaborative testing benefits all students equally, regardless of pre-intervention test performance (Figure 4).

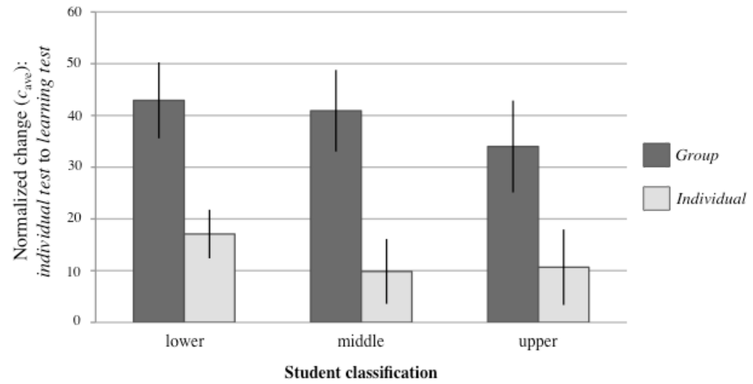

Figure 4: Normalized change (c_{ave}) for three classifications of students (lower, middle and upper) based on their baseline individual test scores. Students in the normalized change data set were separated into three quantiles of roughly equal size based on students’ scores on all questions during the individual test: lower (<= 50%; n = 20), middle (60%; n = 22) and upper (>= 70%; n = 25), respectively. A two-way ANOVA revealed no interaction between condition (group and independent) and quantile (F(5,128)=0.15, p = 0.86) on average normalized change (cave) in the group condition. Bars represent standard error.

Conclusions

Students showed a significantly higher gain in retention when tested in a collaborative setting over a traditional, individual-written test setting.

Students’ retention appears to be influenced by the performance of their group in the 2nd stage of the exam.

Regardless of their performance prior to the two stage exam, all students appear to benefit equally when tested collaboratively.

References

[1] Stearns, S. A. (1996). Collaborative exams as learning tools. *College Teaching*, *44*(3), 111–112.
[2] Russo, A. and Warren, S.H. (1999) Collaborative Test Taking College Teaching, Vol. 47, No. 1 (Winter, 1999), pp. 18-20
[3] Shindler, J.V. (2004) “Greater Than the Sum of the Parts?” Examining the Soundness of Collaborative Exams in Teacher Education Courses Innovative Higher Education Volume 28, Number 4
[4] 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 Physiol. Edu. 27: 102-108 pp, 2003

Acknowledgments

The authors would like to thank the Carl Wieman Science Education Initiative for funding and support for this project. We would also like to thank Carl Wieman, Francis Jones, Ido Roll, Louis Deslaurier, Mandy Banet, Laura Wier, James Day, Georg Reiger, Cynthia Heiner, Sarah Gilbert, Warren Code, all the current and former Science Teaching and Learning Fellows in the Carl Wieman Science Education Initiative and most importantly Students in the 2012 Summer Section of EOSC 114.

Want to try two-stage exams in your classroom?

Scan to see video of two-stage exams in action at UBC.
url: http://www.cwsei.ubc.ca/resources/SEI_video.html

Pick up a two-stage exam “best practices” guideline below.

Talk to us! Contact by email or twitter with your questions or ideas. Bridgette: bclarkst@zoology.ubc.ca; Brett: gilley@eos.ubc.ca