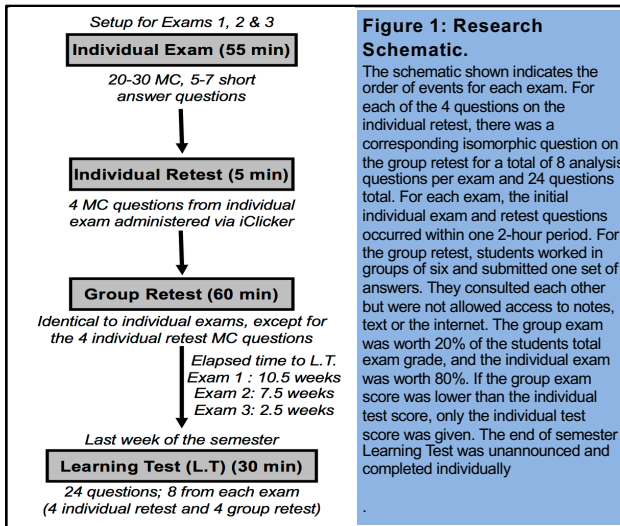




Examining the Impact of Collaborative Two-Stage Exams on Long-Term Retention

Dr. Stefanie R. DeVito Martin, Kathy Denning and Dr. Mark Mort
Undergraduate Biology, University of Kansas, Lawrence, Kansas

Abstract In a collaborative two-stage exam, students complete an exam individually and then immediately complete it again in groups, allowing students to immediately discuss and receive feedback on exam material while also giving them the opportunity to raise their exam grade. Currently, the impact of a collaborative exam on retention of material throughout the semester is unknown. We wanted to know whether the collaborative retest promotes retention of material that is sustained throughout the semester. We exposed students to exam questions three times; an initial individual exam, an immediate retest either in a group or individual setting and an unannounced individual learning test administered at the end of the semester. Isomorphic questions were used for the group and individual settings. This was done for three exams which were given 10.5, 7.5 and 2.5 weeks prior to the learning test, allowing us to examine learning gains from group and individual retest settings as a function of time. We find that the two-stage retest format improves learning gains when compared to an individual-only exam setting within 2-3 weeks of the initial exam, however retention decreases over time.



Biol 152 Class Performance During Each Exam (mean % +/- standard error)				
	Individual Test (IT)	Retest (RT)	Learning Test (LT)	Learning Gains (LT - IT)
Exam 1				
	2/15/16	2/15/16	4/28/16	
Group Retest	66.86 +/- 1.47	81.37 +/- 0.95	68.28 +/- 1.53	1.42
Individual Retest	66.82 +/- 1.39	69.08 +/- 1.38	68.72 +/- 1.45	1.9
Difference between conditions				-0.48
Delay between IT and RT: 10 weeks, 3 days				
Effect Size, 95% C.I.				-0.02, [-0.12, 0.08]
Exam 2				
	3/7/16	3/7/16	4/28/16	
Group Retest	82.67 +/- 1.45	96.93 +/- 0.7	74.65 +/- 1.61	-8.02
Individual Retest	77.83 +/- 1.48	79.72 +/- 1.37	69.58 +/- 1.39	-8.25
Difference between conditions				0.23
Delay between IT and RT: 7 weeks, 3 days				
Effect Size, 95% C.I.				0.01, [-0.09, 0.11]
Exam 3				
	4/11/16	4/11/16	4/28/16	
Group Retest	58.96 +/- 1.71	83.02 +/- 1.07	76.18 +/- 1.53	17.22
Individual Retest	56.72 +/- 1.73	56.84 +/- 1.71	60.61 +/- 1.82	3.89
Difference between conditions				13.33
Delay between IT and RT: 2 weeks, 3 days				
Effect Size, 95% C.I.				0.62, [0.52, 0.72]

Table 1: Average Class Performance by Exam. The average class score is shown for the first exposure (individual test), second exposure (individual retest or group retest, respectively) and third exposure (learning test). Scores represent the average class score (mean \pm SEM) on the four analysis questions for each condition. Only students who completed all three exams and the learning test are included in the analysis ($n = 212$). Learning gains are represented by the change in score from the individual test (first exposure) to the learning test (third exposure). Delay indicates the amount of time between the given exam and the learning test. Effect size was calculated from the difference in learning gains between the group vs. individual retest treatments for each exam using Cohen's d and a pooled standard deviation.

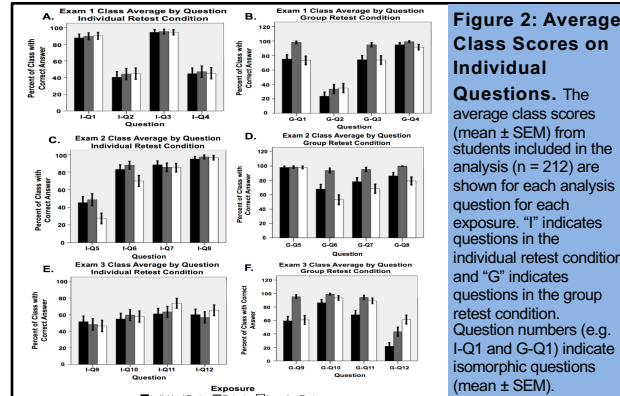


Figure 2: Average Class Scores on Individual Questions. The average class scores (mean \pm SEM) from students included in the analysis ($n = 212$) are shown for each analysis question for each exposure. "I" indicates questions in the individual retest condition and "G" indicates questions in the group retest condition. Question numbers (e.g. I-Q1 and G-Q1) indicate isomorphic questions (mean \pm SEM).

Table 3: Comparison of Isomorphic Questions			
Question	Topic	Bloom's Level	Treatment
Exam 1 - Evolution			
I-Q1	Environmental Effects on Evolution	4 - Analysis	Individual Retest
G-Q1	Environmental Effects on Evolution	4 - Analysis	Group Retest
I-Q2	Phylogenetic Analysis	1 - Knowledge	Individual Retest
G-Q2	Phylogenetic Analysis	3 - Application	Group Retest
I-Q3	Population Genetics	2 - Comprehension	Individual Retest
G-Q3	Population Genetics	2 - Comprehension	Group Retest
I-Q4	Common Ancestors	1 - Knowledge	Individual Retest
G-Q4	Common Ancestors	1 - Knowledge	Group Retest
Exam 2 - Plant Structure, Function and Diversity			
I-Q5	Plant Evolution	2 - Comprehension	Individual Retest
G-Q5	Plant Hormones	2 - Comprehension	Group Retest
I-Q6	Plant Structure	2 - Comprehension	Individual Retest
G-Q6	Plant Structure	2 - Comprehension	Group Retest
I-Q7	Plant Function	3 - Application	Individual Retest
G-Q7	Plant Function	3 - Application	Group Retest
I-Q8	Plant Structure	1 - Knowledge	Individual Retest
G-Q8	Plant Hormones	2 - Comprehension	Group Retest
Exam 3 - Animal Structure, Function and Diversity			
I-Q9	Nervous System	3 - Application	Individual Retest
G-Q9	Nervous System	3 - Application	Group Retest
I-Q10	Hormones	2 - Comprehension	Individual Retest
G-Q10	Hormones	2 - Comprehension	Group Retest
I-Q11	Kidneys	1 - Knowledge	Individual Retest
G-Q11	Kidneys	1 - Knowledge	Group Retest
I-Q12	Development	2 - Comprehension	Individual Retest
G-Q12	Development	2 - Comprehension	Group Retest

Table 2: Comparison of Isomorphic Questions. The topic and Bloom's cognitive levels are listed for each question. Questions are grouped into isomorphic pairs. The average Bloom's cognitive level is indicated for each set of individual and group retest questions for each exam.

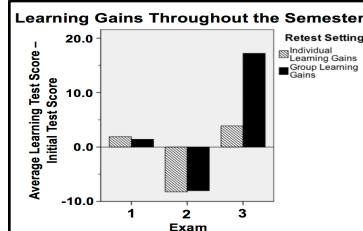


Figure 3: Average Learning Gains. The learning gains for the four analysis questions for each treatment on each exam are shown. Learning gains are calculated from the difference between the average class learning test score and the average class initial test score.

RM ANOVA Results					
	d.f.	S.S.	M.S.	F	P
Treatment	1	0.669	0.669	10.56	0.00135 **
Error	212	13.445	0.0634		
Delay	1	2.56	2.56	37.63	<0.0001 ***
Error	212	14.42	0.068		
Treatment x Delay	1	1.219	1.219	19.22	<0.0001 ***
Error	212	13.446	0.0634		

Table 3: ANOVA Analysis. A repeated-measures ANOVA was conducted to assess the main effects of treatment (group versus individual retest), delay between the initial- and learning tests (2.5, 7.5 or 10.5 weeks) and their interaction.

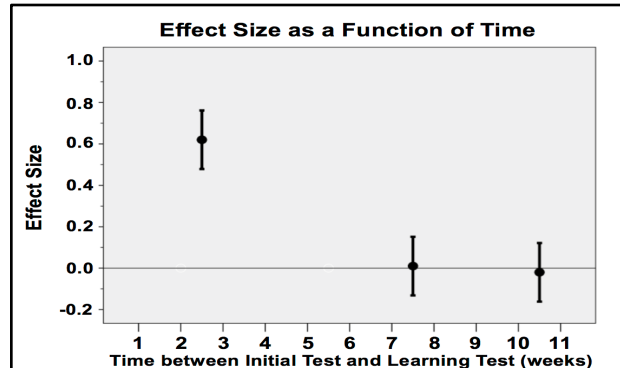


Figure 4: Effect Size Decreases as a Function of Time. The difference in average class score between the learning test and initial test (learning gains) were used to calculate the effect size between the group and individual retest treatments for each exam using Cohen's d and a pooled standard deviation. The effect sizes are graphed for each exam as a function of time that passed between the first and third exposure. Error bars indicate 95% confidence intervals.

Conclusions:

- ▶ A collaborative, group retest increases student exam scores
- ▶ A group retest increases retention of exam material compared to an individual retest up to ~2-3 weeks after the initial exam
- ▶ Retention of exam material decreases significantly between ~3-7 weeks after the initial exam

Future Directions:

- ▶ Where in the 3-7 week window does retention decrease?
- ▶ How quickly does retention decline?
- ▶ How does retention from multiple choice questions compare to other summative assessment types (e.g. short answer or group projects?)