The Effectiveness of Two-Stage Exams in Promoting Learning in Introductory Physics Courses

Simmer Mand

University of British Columbia
Dr. Ives

Two-Stage Exams

- First Stage: Students complete an assessment individually.
- Second Stage: Students complete the same assessment in the groups of 3 to 4.

Group Stage : Students figure the solutions through the use of Immediate Feedback Assessment Scratch Cards

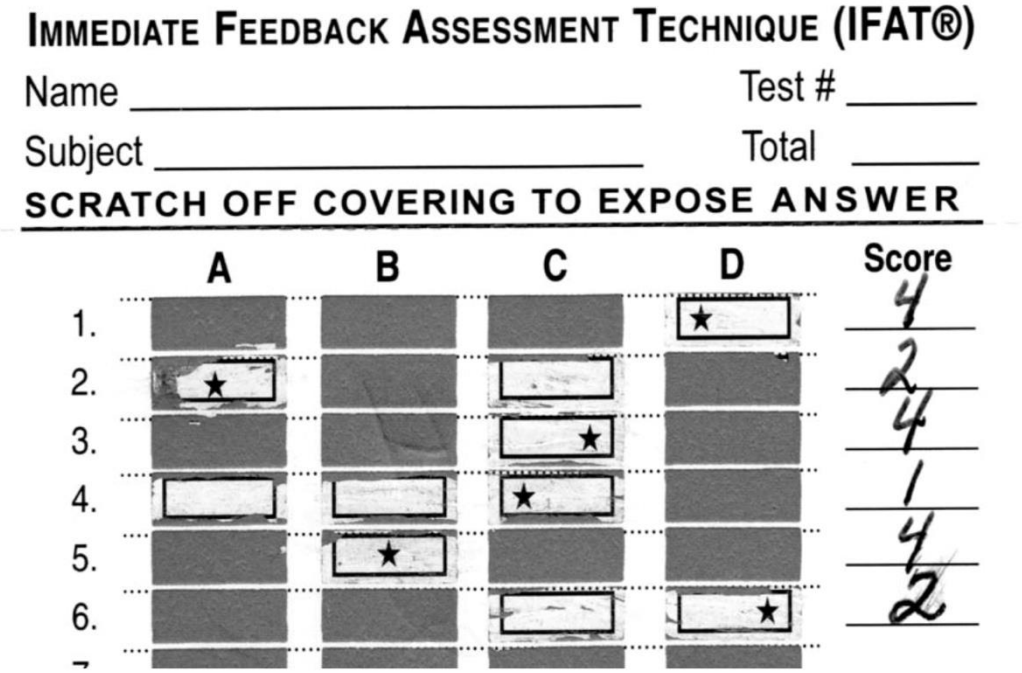

Figure 1: Students discuss their answers and agree upon an answer, which they scratch on the card. If their answer is correct, they see a star and get all the allotted points. If they don't see a star, they discuss farther to scratch the answer. They repeat the procedure until they find answer.
Gillespie, Judy (2007)

Benefits of Two-Stage Exams

- Immediate Feedback is provided through the use of Scratch cards.
- Group Stage: Allow student interaction and learning from peers unlike in a traditional exam.
- Better retention of material

Objective

•Look at the factors in the Group Discussion that result in maximum learning.

Experimental Design

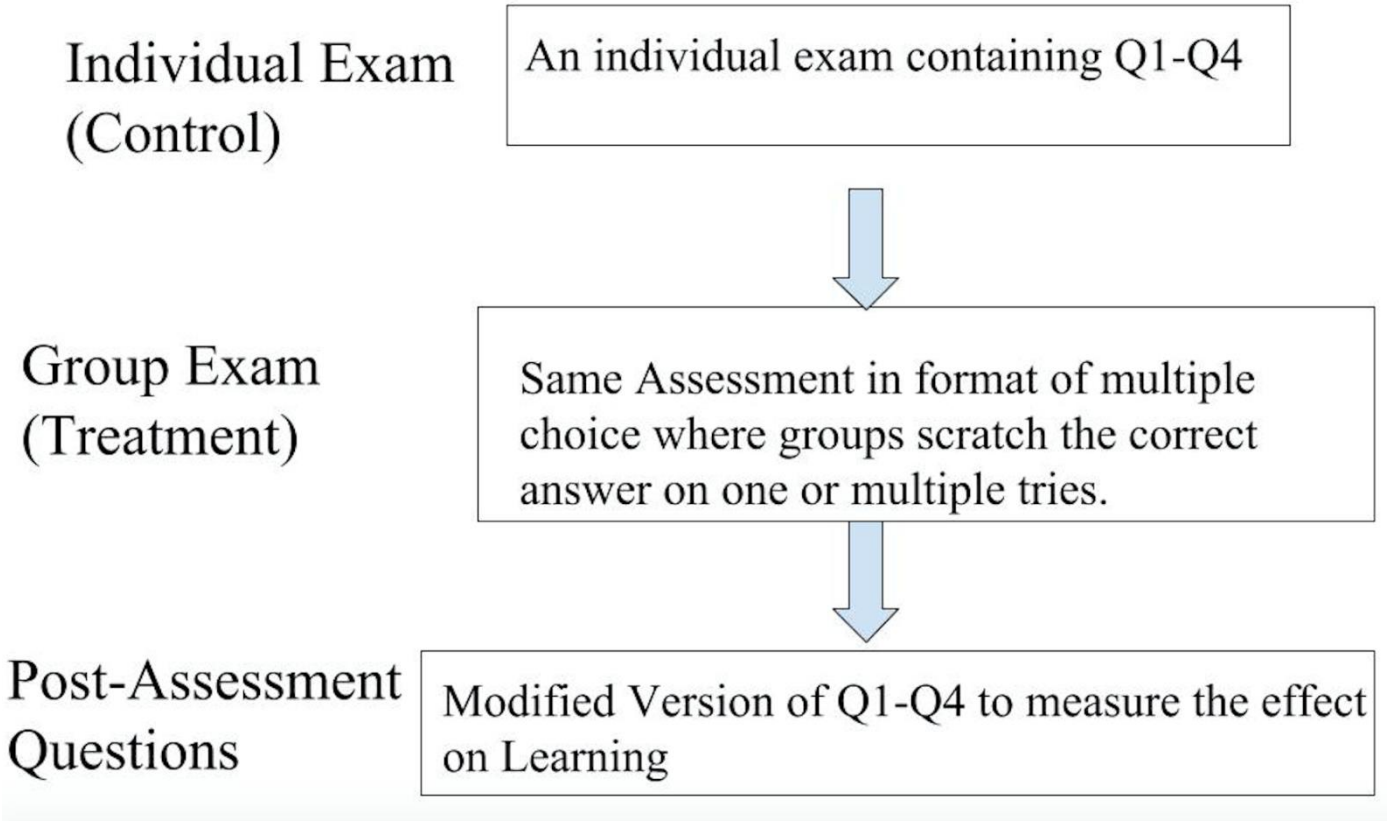

Figure 2: The students participate in the control group of the experiment by taking the individual section of two-stage exam. Students participate in the treatment group of the experiment by taking the group portion of the exam. The iclicker questions (given as post-assessment questions given 2-3 days later) determine how the students have performed as compared to individual section of their exams.

Experimental Design

- Individual Section of the Exam: Student's performance on each question.
  - Correct = 1
  - Incorrect = 0

- Group Section of the Exam: Group's performance on each question through use of scratch cards.

- Post Assessment Iclicker Question: Student's performance on each question.
  - Correct = 1
  - Incorrect = 0

Logistic Regression

- Using the data on how students perform on each question individually and after the exam, we tried to fit the logistic regression model, while controlling for factors affecting the group discussion.

- In logistic regression, the log odds of outcome is modelled using the combinations of different predictor variables.

- The outcome variables are dichotomous.

Simplest Model
Does Students' performance on each question individually effect their Post-Assessment Performance?

- Model to examine how students' performance on each question on the exam effects their post-assessment performance.

$$POSTQ = SOLOQ + QNUM + (1|ID)$$

SOLOQ: A question being correct (1) or incorrect (0)
QNUM: Controlling for the difficulty of individual Question
(1|ID): Controlling for Individual Student

Simplest Model
Does Students' performance on each question individually effect their Post-Assessment Performance?

- SOLOQ is a very strong predictor of how a student performs on post-assessment question given 2-3 days later the exam.

- The odds of a student getting the post-assessment questions correct increases by a factor of 2.1 if they have gotten the questions correct on the individual section of the exam.

Simplest Model
Does Students' performance on each question individually effect their Post-Assessment Performance?

	Estimate	Std. Err	z-Value	Pr(>|z|)
Intercept	0.6072	0.1542	3.939	8.19e-0.5***
SOLOQ	0.7430	0.1285	5.782	7.39e-9 ***
QNUM1	-0.9892	0.1703	-5.810	6.26e-9***
QNUM2	-0.1857	0.1757	-1.057	0.29
QNUM3	-1.7613	0.1634	-10.777	< 2e-16***

Table 1 : Showing how performance on Individual Question predicts the students' performance on the post-assessment questions

What body of students tend to benefit the most from Group Discussions?

- Are these the students with higher grade overall that tend to show the most learning from the group discussion?

- Take the subset of students who don’t have the question right and control for their individual score to measure any learning on the post assessment question.

What body of students tend to benefit the most from Group Discussions?

- Subset : where student's individual question score is 0.

$$POSTQ = SOLOSCORE + GROUPQ + QNUM + (1|ID)$$

SOLOSCORE: Their Individual Score on the Exam
GROUPQ: Whether they get question correct in their groups on the first try
QNUM: Controlling for the difficulty of individual Question
(1|ID): Controlling for Individual Student

What body of students tend to benefit the most from Group Discussions?

	Estimate	Std. Err	z-Value	Pr(>|z|)
Intercept	-0.4064	0.3869	-1.051	0.2935
SOLOSCORE	0.1022	0.0512	1.993	0.0462 *
GROUPQ	-0.00291	1.684	-0.017	0.9862
QNUM2	-0.5282	0.2860	-1.847	0.0648
QNUM3	0.2567	0.2811	0.93	0.3611
QNUM4	-1.302	0.308	-4.224	2.4e-0.5***

Table 2: Showing how Individual Score predicts the students' performance on the post-assessment questions

We can compare this model to how Individual Score affect the post assessment performance if students were correct on the individual question

	Estimate	Std. Err	z-Value	Pr(>|z|)
Intercept	0.426	0.440	0.967	0.334
SOLOSCORE	0.117	0.0573	2.049	0.0405***
QNUM2	-1.22	0.281	-4.33	1.97e-5***
QNUM3	-0.548	0.323	-1.70	0.0891
QNUM4	-2.005	0.215	-9.312	<2e-16***

Table 3: Showing how Individual Score predicts the students' performance on the post-assessment questions

Simplest Model
What body of students tend to benefit the most from Group Discussions?

- Students with better understanding of the material tend to gain further from group discussions if they have minor conceptual flaw in their understanding.

- With 10 units increase in score, the odds of a student getting a question correct which they previously have it incorrect increases by the factor of 1.101.

Importance of Student's Individual Score

- Since we learned that Individual Score is a significant predictor in determining the students' success on the post assessment questions, we have to control for Individual Score in subsequent models.

Does Number of Tries in Group Discussion to Get the question right affect learning?

- In their groups, students have maximum four tries to get the question right.

- If they are getting question correct on the first try it could due to the fact that one person knowing answer is telling the correct answer ~ not so useful discussion.

- More than one try might lead to better discussion among the group ~ resulting in learning.

Does Number of Tries in Group Discussion to Get the question right affect learning?

- Treated number of tries as a categorical variable:
  - Get the question correct on the first try: 1
  - Get the question correct on the second try: 2
  - Get the question correct with more than 2 tries: 3

- After fitting the model, no significant results were obtained meaning that number of tries in the group does not affect how students perform on the post-assessment questions.

Does Group Size affect learning?

- There were groups of either 3 and 4

- Through model fitting group size also did not turn to be significant predictor in determining how students perform on the post-assessment questions.

Does Number of People Correct in the Group affect the learning?

- First, students complete each question on their own.
- Look at the subset of students who have incorrect answers individually.
- We see what happens when they become a part of a group in which nobody was correct, one person was correct, and more than one person was correct.
- The result was not a significant predictor.

Does Number of People Correct in the Group affect the learning?

- Look at the subset of students who have correct answers individually.

- We see what happens when they become a part of a group in which they were only student correct vs. there are more people correct.

Does Number of People Correct in the Group affect the learning?

- If they are not the only student correct in the group, the odd of them getting the answer correct decreases by the odds of 1.85.

- One hypothesis: When there is only one student correct in the group, they have the most responsibility for explaining the correct answer to their group mates yielding in greater learning for that particular student.

Does Number of People Correct in the Group affect the learning?

POSTQ = NGCORR+ SOLOSCORE+ QNUM+ (1|ID)

NGCORR: 1 if only one person is correct or 0 if more than one are correct
SOLOSCORE: Their Individual Score on the Exam
GROUPQ: Whether they get question correct in their groups on the first try
QNUM: Controlling for the difficulty of individual Question
(1|ID): Controlling for Individual Student

Does Number of People Correct in the Group affect the learning?

	Estimate	Std. Err	z-Value	Pr(>|z|)
Intercept	0.27673	0.44807	0.618	0.5368
NGCORR	-0.62313	0.30174	2.065	0.0389*
SOLOSCORE	0.12071	0.05776	2.090	0.0366*
GROUPQ	0.5457	0.2513	2.171	0.0299*
QNUM2	-1.402	0.297	-4.716	2.4e-6***
QNUM3	-0.734	0.337	-2.182	0.0291*
QNUM4	-2.041	0.218	- 9.36	< 2.4e-0.16**

Table 3: Showing how only student correct in the group vs. more than one correct predicts the students' performance on the post-assessment questions

Can gender be a control variable when trying to fit different models?

- For Physics 101 midterm, the midterm average for males was 64.5% which was 4.7% higher than female midterm average (59.9%).

- Therefore, when the model only uses female as a predictor, the odds of a student getting the post assessment question correct decreases.

Can gender be a control variable when trying to fit different models?

POSTQ = FEMALE + GROUPQ + (1|ID)

NGCORR: 1 if a student is female and 0 if not.
GROUPQ: Whether they get question correct in their groups on the first try
QNUM: Controlling for the difficulty of individual Question
(1|ID): Controlling for Individual Student

	Estimate	Std. Err	z-Value	Pr(>|z|)
Intercept	1.2158	0.1999	6.081	1.2e-9***
FEMALE	-0.3022	0.1202	-2.513	0.0120*
GROUPQ	0.2314	0.1361	1.700	0.0891
QNUM2	-1.251	0.1833	-6.525	8.79e-12***
QNUM3	-0.570	0.1821	-3.124	0.00178**
QNUM4	-1.865	0.1682	-11.84	< 2.4e-0.16***

Table 3: Showing how student being a female predicts the students' performance on the post-assessment questions

Can gender be a control variable when trying to fit different models?

- However, when this model is controlled for students' responses for each question or for individual score, the significance of female predictor goes away.

- Therefore, both genders seem to benefit equally from the group exams.

Conclusion

- From the results, if only one student is correct in the group, that one student's learning tends to benefit the most from the group discussion.
- Compose a group where only one student is correct in the group, optimizing the learning of at least 0.25 students.
- Results are only valid for one data set (PHYS 101 Midterm I).
- To avoid statistical flukes, another data set (PHYS 101 Midterm II) has to be examined.