Further investigations into the effectiveness of collaborative group exams

Joss Ives
Dept. of Physics and Astronomy
University of British Columbia, Vancouver, Canada
joss@phas.ubc.ca
learnfacilitation.wordpress.com
@jossives

Scan for contact info

Summary

To quantify the learning impact of collaborative group exams, a randomized crossover design was used in 2014 and 2015 in an introductory calculus-based physics course where each student participated in both the treatment and control groups. Questions from each of the two midterms were designed to form near-transfer pairs with the end-of-course diagnostic, which was used as a retest to measure learning.

In both years, improved learning was shown in the treatment group for retest questions associated with the second midterm (retest within 2 weeks of the midterm). The 2014 data show no improved learning in the treatment group for retest questions associated with the first midterm (6-7 weeks prior to retest) and the 2015 data show a decrease in learning for retest questions associated with the first midterm (4-6 weeks prior to retest).

A likely explanation for this difference is that there is a time-based decay of the learning impact from the groups exams. However, additional studies are needed to investigate the difference in results between 2014 (null) and 2015 (decreased learning) for the longer times between midterm and retest.

The model:

In the following mixed-effects logistic regression model, a positive $\beta_1$ indicates the group exams had a positive effect on retest success. The analysis was run separately for each year and for the retest questions associated with midterm one (Q1.1-1.x) and for those associated with midterm two (Q2.2-2.x):

$$\log_{odds}text{Retest success}_{ijk} = \beta_0 + \beta_1 \times text{Pre} + \beta_2 \times text{Question} + \beta_3 \times text{Treatment} + \epsilon_i,$$

where,

- Retest success is the (binary) success on the learning test of Student $i$ on Question $j$;
- Pre is a binary variable that indicates if Student $i$ answered correctly the question isomorphic to Retest Question $j$;
- Question is a categorical variable representing question number and account for differences in question difficulty; and
- $\epsilon_i$ is a random intercept for Student $i$, which accounts for differences in student ability.

Results:

Shorter timeframes (retest given within 2 weeks of the 2nd midterm):

- Treatment: (collaborative group exam) predicted success for retest questions Q2.1-Q2.x, (2014: $\beta_1 = .203$, SE $= .079$, $p = .011$ & 2015: $\beta_1 = .363$, SE $= .083$, $p < .001$)

Expressed as odds ratios, the odds of answering a question correctly on the learning test versus not answering it correctly increased by a factor of 1.22 (95% CI [1.05, 1.43]) in 2014 and a factor of 1.44 (95% CI [1.22, 1.69]) in 2015 for those in the treatment as compared to the control.

Longer timeframes (retest given with 4-6 weeks): For 2014, no statistically significant predictive power for retest questions Q1.1-1.x.

For 2015, treatment predicted success for retest questions Q1.1-Q1.x ($\beta_1 = .199$, SE $= .077$, $p = .011$)

The fits between the model and data were good (2014: $\chi^2(8) = 279.1$, p < .001 & 2015: $\chi^2(10) = 446.5$, p < .001)

A mixed-effects logistic regression showed improved learning for short timeframes (up to 2 weeks) but null (2014) or decreased learning (2015) for longer timeframes (4-7 weeks)

Question by question comparison of retest performance

Note: Question numbering schemes from 2014 and 2015 are not the same.

Matched question pairs

The midterm exam questions were designed to form matched near-transfer pairs with questions on the locally developed end-of-term course diagnostic.

Question validation:

Diagnostic question validation via:
- Expert feedback and student interviews
- Classical Test theory analysis ongoing

2014 Exam question validation via:
- Four course instructors
- Graduate student TA feedback

Results from previous studies

References


Joss Ives
Dept. of Physics and Astronomy
University of British Columbia, Vancouver, Canada
joss@phas.ubc.ca
learnfacilitation.wordpress.com
@jossives

Scan for contact info

Two-stage collaborative group exams and study design

All students first completed the midterm exam individually:
- 2014: Midterm 1: n = 679, Midterm 2: n = 673
- 2015: Midterm 1: n = 701, Midterm 2: n = 703

Treatment: Immediately after the individual exams are collected, students self-organized into collaborative groups of 3 or 4 and retook a subset of the original exam questions (different subsets for conditions A-C).

Retest: The end-of-term diagnostic contained near-transfer questions that partnered with those from the original exam.

The time between the first midterm (questions 1.1-1.x) and the diagnostic was 6-7 weeks in 2014 and 4-6 weeks and the time between the second midterm (questions 2.1-2.x) and the diagnostic was 2-2 weeks in 2014 and 3-2 weeks in 2015.

A mixed-effects logistic regression model showed improved learning for short timeframes (up to 2 weeks) but null (2014) or decreased learning (2015) for longer timeframes (4-7 weeks)

Question by question comparison of retest performance

Note: Question numbering schemes from 2014 and 2015 are not the same.

Matched question pairs

The midterm exam questions were designed to form matched near-transfer pairs with questions on the locally developed end-of-term course diagnostic.

Question validation:

Diagnostic question validation via:
- Expert feedback and student interviews
- Classical Test theory analysis ongoing

2014 Exam question validation via:
- Four course instructors
- Graduate student TA feedback

Results from previous studies

References


Joss Ives
Dept. of Physics and Astronomy
University of British Columbia, Vancouver, Canada
joss@phas.ubc.ca
learnfacilitation.wordpress.com
@jossives

Scan for contact info

Two-stage collaborative group exams and study design

All students first completed the midterm exam individually:
- 2014: Midterm 1: n = 679, Midterm 2: n = 673
- 2015: Midterm 1: n = 701, Midterm 2: n = 703

Treatment: Immediately after the individual exams are collected, students self-organized into collaborative groups of 3 or 4 and retook a subset of the original exam questions (different subsets for conditions A-C).

Retest: The end-of-term diagnostic contained near-transfer questions that partnered with those from the original exam.

The time between the first midterm (questions 1.1-1.x) and the diagnostic was 6-7 weeks in 2014 and 4-6 weeks and the time between the second midterm (questions 2.1-2.x) and the diagnostic was 2-2 weeks in 2014 and 3-2 weeks in 2015.

A mixed-effects logistic regression model showed improved learning for short timeframes (up to 2 weeks) but null (2014) or decreased learning (2015) for longer timeframes (4-7 weeks)

Question by question comparison of retest performance

Note: Question numbering schemes from 2014 and 2015 are not the same.

Matched question pairs

The midterm exam questions were designed to form matched near-transfer pairs with questions on the locally developed end-of-term course diagnostic.

Question validation:

Diagnostic question validation via:
- Expert feedback and student interviews
- Classical Test theory analysis ongoing

2014 Exam question validation via:
- Four course instructors
- Graduate student TA feedback

Results from previous studies

References