Traditional lecturing compared with various forms of active learning. (Freeman et al. 2014)

”To test the hypothesis that lecturing maximizes learning and course performance, we metaanalyzed 225 studies that reported data on examination scores or failure rates when comparing student performance in undergraduate science, technology, engineering, and mathematics (STEM) courses under traditional lecturing versus active learning. ”

“The overall mean effect size for performance on identical or equivalent examinations, concept inventories, and other assessments was a weighted standardized mean difference of 0.47 (Z = 9.781, P << 0.001)—meaning that on average, student performance increased by just under half a SD with active learning compared with lecturing.”

“The overall mean effect size for failure rate was an odds ratio of 1.95 (Z = 10.4, P << 0.001). This odds ratio is equivalent to a risk ratio of 1.5, meaning that on average, students in traditional lecture courses are 1.5 times more likely to fail than students in courses with active learning.”

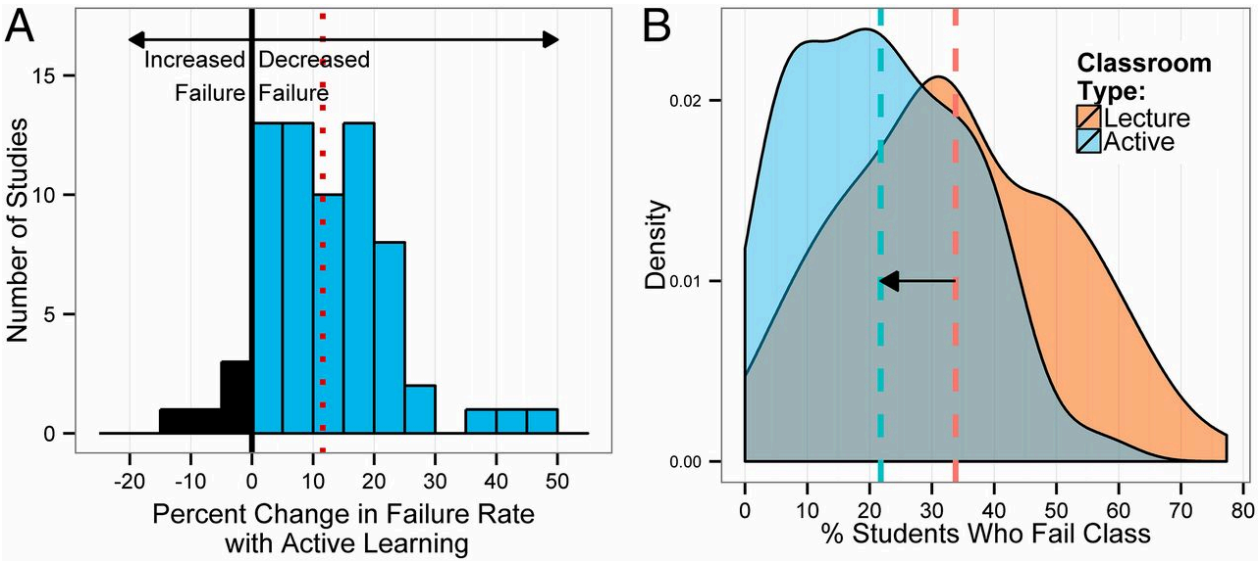

Fig. 1 Changes in failure rate. (*A*) Data plotted as percent change in failure rate in the same course, under active learning versus lecturing. The mean change (12%) is indicated by the dashed vertical line. (*B*) Kernel density plots of failure rates under active learning and under lecturing. The mean failure rates under each classroom type (21.8% and 33.8%) are shown by dashed vertical lines.

Heterogeneity analyses

Analyses indicate:

- no statistically significant variation among experiments based on the STEM discipline of the course in question, with respect to either examination scores.
- average effect sizes were lower when the outcome variable was an instructor-written course examination as opposed to performance on a concept inventory.
- significant variation in terms of course size, with active learning having the highest impact on courses with 50 or fewer students
- no statistically significant difference in active learning’s effect size when comparing

- no statistically significant difference in introductory versus upper-division courses
- no heterogeneity based on methodological quality
- Analyzing variation with respect to controls over instructor identity also produced no evidence of heterogeneity.

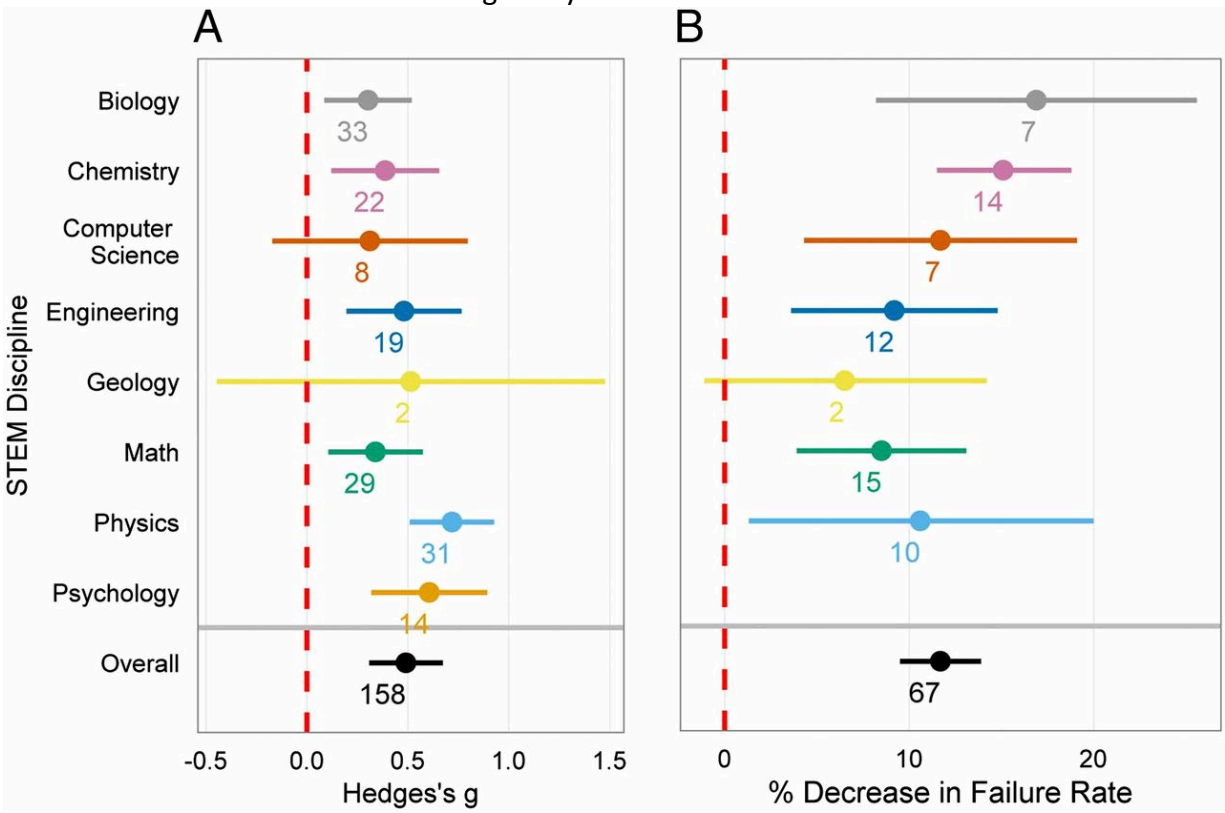

Fig. 2 Effect sizes by discipline. (*A*) Data on examination scores, concept inventories, or other assessments. (*B*) Data on failure rates. Numbers below data points indicate the number of independent studies; horizontal lines are 95% confidence intervals.

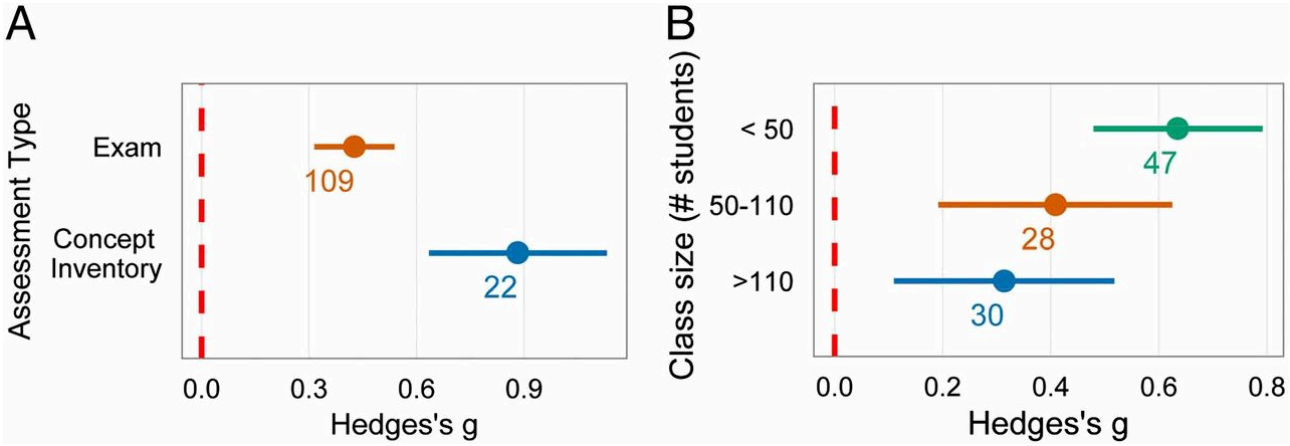

Fig. 3 Heterogeneity analyses for data on examination scores, concept inventories, or other assessments. (A) By assessment type—concept inventories versus examinations. (B) By class size. Numbers below data points indicate the number of independent studies; horizontal lines are 95% confidence intervals.