

Review of: "The Positive Impact of Dropping the Lowest Test Score on Academic Performance and Stress Levels in MathBased Graduate Courses"

Qihui Chen¹

¹ China Agricultural University

Potential competing interests: No potential competing interests to declare.

Comments on "The Positive Impact of Dropping the Lowest Test Score..." by Toma et al.

General comments

The study under review examines what form of exam structure (e.g., exam reviews versus dropping the lowest scores) can relieve student stress and (thus) school performance in college-level math-oriented courses based on a "quasi" experiment conducted in an engineering course (with three different sections) at the New York Institute of Technology. The chosen topic and the design of the "quasi" experiment are both interesting, but there are a great many other points (e.g., small sample size, lack of controls, and lack of random assignments, etc.) that merit further consideration. My more specific comments are provided below, largely by section of the paper.

Specific comments

1. The Abstract should be made more concise yet, at the same time, more informative. For example, the 2^d and 3rd sentences contain very similar information to that in the 4th and 5th sentences. These sentences can be combined. Meanwhile, more information on the method adopted in the study should be provided to better inform the reader.
2. The Introduction may benefit from a restructuring effort.
 - 1) The current version of the Introduction seems to be talking too much about the literature (those proposed methods). It's more important to quickly get to the gaps in the literature and your contribution[s]; you can always have a stand-alone section on the literature.
 - 2) 1st paragraph: I don't think motivating the paper using suicide attempts is reasonable—How many students made suicide attempts because of academic pressures? I think it suffices to mention stress and anxiety issues (that may not be related to suicides) to motivate the study.
 - 3) 2nd paragraph: What's the point of mentioning the Sokolowski & Ansai study but not mentioning their findings? So they did some research, yet do their findings support your arguments here?
 - 4) 4th paragraph, 2nd line: it's better to change "writers" to "term-paper writers" or "essay writers." Also, I think the

argument on students' feeling of unfairness is not well-grounded: in math-related courses, how likely can the final exam be given as an essay exam? That argument may be reasonable in, say, an economics course in which the exam questions may have open-ended answers, but I don't think it fits into a Fluid Mechanics course.

5) 5th paragraph, 1st line: The term "research studies" should just be "studies."

6) 2nd last paragraph: why "sought input from engineering students" rather than other types of students? Is there anything particular about engineering students (esp. in terms of helping you answer your research question[s])? And who are they? You described their opinions before even introducing them to the reader.

3. The design of the "experiment" should be improved.

1) The enrollment sizes of the three sections seem too small to yield meaningful statistical results (in particular, the law of large numbers may not work with only about 20 students in each section).

2) Did the students know about the differences in exam structure before enrolling in the course? If so, they might have self-selected into different sessions of the course based on their unobserved strength and preferences. This will likely introduce biases in your results.

3) Are the lectures of different sessions given at the same time? Can a student enrolled in section 1 attend sections 2 and 3, for example, to participate in the exam review? If so, your results might be biased.

4) Did any students drop out during the course? If so, you might need to address the potential issue of "attribution bias."

Any of these issues may invalidate your findings. So, please address them carefully in the paper.

4. Results should be tied to statistical significance.

1) Yes, you have provided evidence of "differences" across different sections, but are these differences statistically significant? If not, they might be simply driven by pure chances. Put differently, if you repeat the experiment in another college, you might see totally different results.

2) You might also need to gauge the practical significance of your results. For example, is a 7% difference really worth paying attention to? With a class size of about 20, a 7% difference can be driven by the behavior of one or two students.

3) You might want to run some regressions to control for differences in students' characteristics.

4) It would be interesting to explore potential heterogeneity in the impact of dropping the lowest score across student groups (e.g., females versus males).

5. Many of the points discussed in the "Discussion" section were not directly derived from your experiment. You should avoid extrapolating your findings too far. Also, the discussion sections should be revised to include the following points:

1) How are your results compared to those in the literature?

- 2) policy implications derived from your experiment.
- 3) Limitations of your study and some directions for future research.
6. I don't think it's a good idea to put many citations in the conclusion section. Most of them should be discussed already in the discussion section.
7. There are some grammatical errors and typos that need to be corrected.