Who Answers Complex Multiple-choice Questions in Physics Correctly?

Nicholas T. Young[1], Mark Mills[1], Rebecca L. Matz[1], Eric Bell[2], Caitlin Hayward[1]

[1]Center for Academic Innovation, University of Michigan
[2]Department of Astronomy, University of Michigan

We know there are grade inequities in physics and STEM in general.[1]

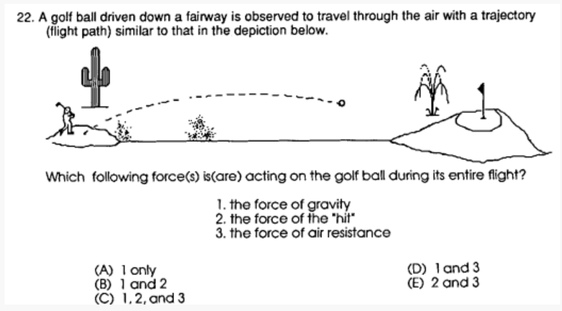
And inequities in grades have been tied to inequities on high stakes exams.[2,3]

Could the type of questions we ask contribute to that gap?

If so, we'd expect students with different preparation and societal advantages to have differing levels of performance on those questions.

Let's use student responses from a computerized a not-for-credit, test-prep program to find out!

A possible culprit: **complex multiple-choice questions**



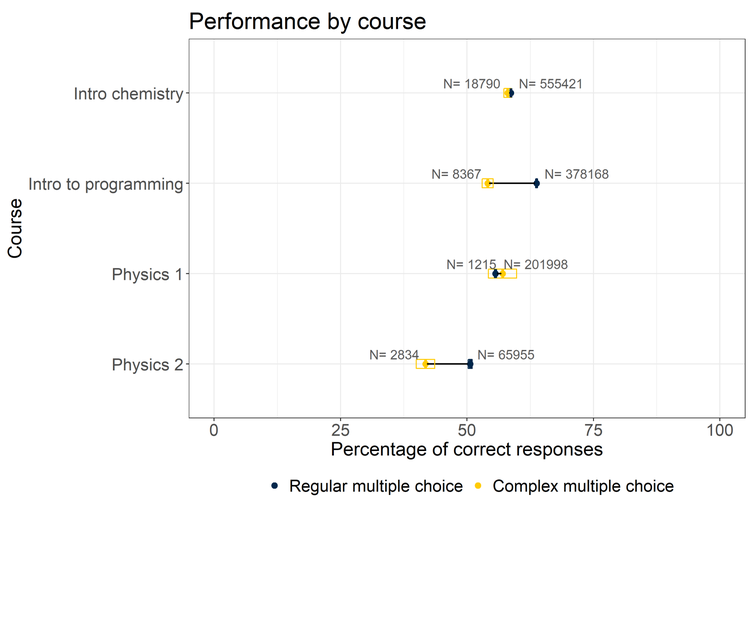
Sample complex multiple-choice question from the Force Concept Inventory[4]

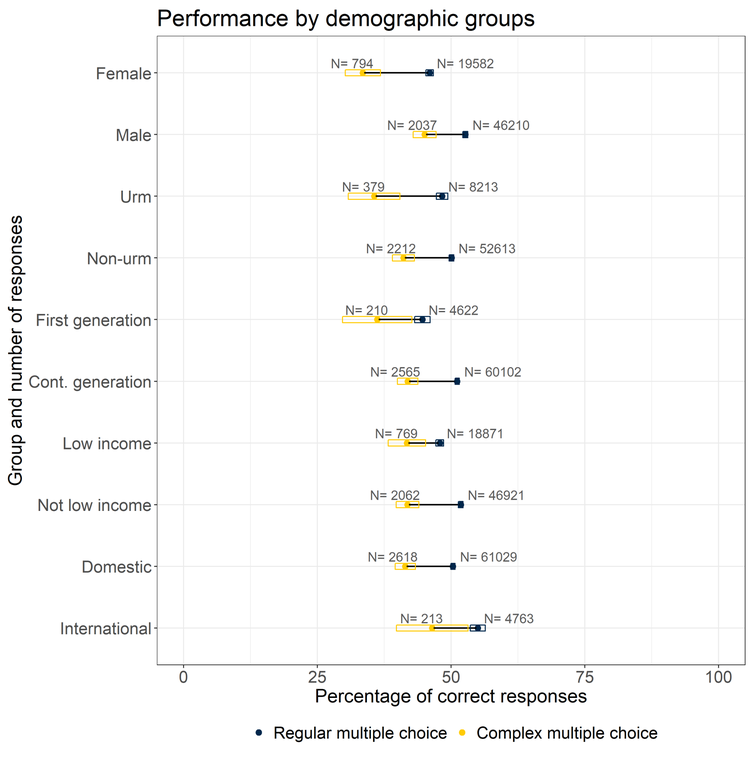
- These questions are claimed measure higher-order thinking[5] and have been shown to be more difficult.[6]
- But performance on these questions is also more variable than standard multiple-choice questions.[7]
- Format of question may also allow students to guess better, possibly helping "lower ability" students answer them correctly.[7]

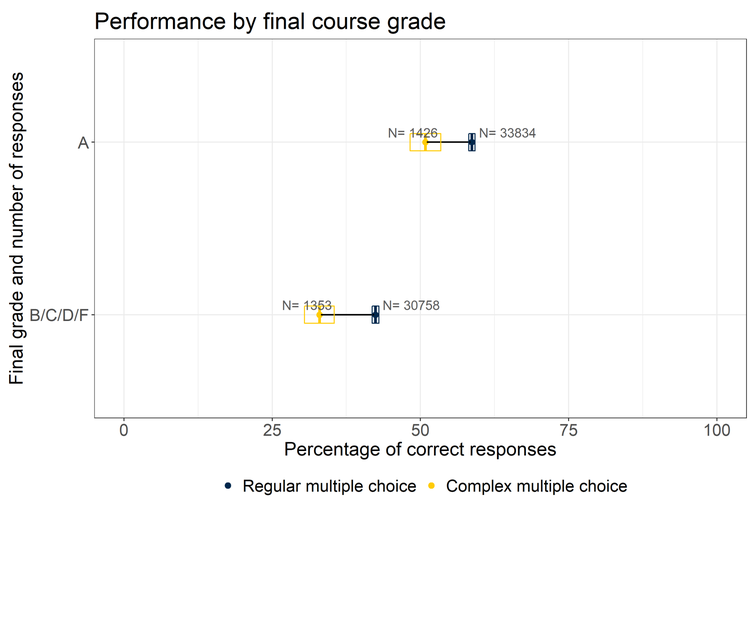
What does our data* show?

In general, students do not perform better on complex multiple questions compared to regular multiple-choice questions. In some courses, performance is comparable and in others, students do better on regular multiple-choice questions

However, there doesn't appear to be any disparate impacts of using complex multiple-choice questions (in second semester introductory physics).







In some courses, students do appear to do worse on complex multiple-choice questions.

In Physics 2, most groups of students persist in doing worse on complex multiple-choice questions compared to regular multiple-choice questions.

However, the difference in performance on the two types of questions is approximately the same for each demographic pair.

All students do better on regular multiple-choice questions, regardless of their final course grade.

But the difference in performance on the two types of questions is about the same, regardless of the final grade earned.

What's next?

For a deeper understanding of complex multiple-choice questions, we can compare student performance on similar questions asked in both formats. We can also investigate why Physics 2 shows a performance difference while Physics 1 does not.

PROBLEM ROULETTE
Learn more about our data source:
https://problemroulette.ai.umich.edu/

***Data:**

- Responses to previous exam questions in a not-for-credit, test-prep setting by students currently enrolled in course (Winter 2019- Winter 2022 term)
- Demographics from student data warehouse. Low-income status computed from median family income in zip code of student's high school.
- Complex multiple-choice questions identified by script to identity answer patterns like "A & B" or "I & III"
- Identified questions then checked by hand to eliminate false positives
- Number of complex multiple-choice questions identified:
  - intro chemistry: 22
  - Intro to programming: 24
  - Physics 1: 5
  - Physics 2: 44

We would like to thank Erin Murray for her assistance in accessing the data.

[1]Whitcomb, K. M., Cwik, S., & Singh, C. (2021). Not all disadvantages are equal: Racial/ethnic minority students have largest disadvantage among demographic groups in both STEM and non-STEM GPA. AERA Open, 7, 23328584211059823.

[2]Simmons, A. B., & Heckler, A. F. (2020). Grades, grade component weighting, and demographic disparities in introductory physics. Physical Review Physics Education Research, 16(2), 020125.

[3]Salehi, S., Cotner, S., Azarin, S. M., Carlson, E. E., Driessen, M., Ferry, V. E., ... & Ballen, C. J. (2019, September). Gender performance gaps across different assessment methods and the underlying mechanisms: The case of incoming preparation and test anxiety. In Frontiers in Education (Vol. 4, p. 107). Frontiers Media SA.

[4]Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. The physics teacher, 30(3), 141-158.

[5]Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning?. Journal of Applied Research in Memory and Cognition, 7(3), 323-331.

[6]Haladyna, T. M., & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. Applied measurement in education, 2(1), 51-78.

[7]Albanese, M. A. (1993). Type K and other complex multiple-choice items: An analysis of research and item properties. Educational Measurement: Issues and Practice.