THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE
APPLIED ECONOMIC & SOCIAL RESEARCH

Research Insights

Are we testing students accurately? How multiple-choice exam questions increase the gender gap in test scores.

Standardised exams, often used to test students and screen job candidates, regularly include multiple-choice questions. These questions are considered objective and easy to mark. But, do multiple-choice tests lead to an unfair comparison between students? This Research Insight suggests men have an advantage in multiple-choice questions compared to other formats of assessments.

ISSN 2653-004X

No. 05/21

May 2021

Are standardised tests an adequate way to examine student performance?

Standardised tests are widely used to measure educational outcomes and provide economic opportunities. These tests are designed to measure the competencies of students and to track them over time. In Australia, students take several standardised tests, such as NAPLAN, HSC, and PISA (the section Standardised Tests in Australia under 'Further Information' describes these tests and their importance). Standardised tests, especially in math, regularly rely on multiple-choice questions. In 2016, for example, the Numeracy section of NAPLAN contained more than 70 per cent multiple-choice questions. In 2015 the mathematics section of PISA was made up of between 30 to 70 per cent multiple-choice questions. Every year, the mathematic HSC tests include a section with multiple-choice questions, which prompt students to identify the correct response from a set of possible answers.

Performance on multiple-choice question tests may reflect more characteristics of the students than their knowledge or skill. For example, confidence plays a role in ruling out incorrect responses and/or risk may play a role in the willingness to guess.

As males in mathematics are more confident and more likely to guess than females, observed differences in performance on multiple-choice tests could conceal gender differences in test-taking abilities rather than gender differences in knowledge or skills.

Using data from the PISA test, this Research Insight shows that males' greater performance in mathematics can be partially attributed to the format of the test (i.e. a large number of multiple-choice questions) rather than gender difference in mathematics knowledge. Indeed, boys perform much better than girls in multiple-choice compared to open-ended questions. Besides, the proportion of multiple-choice questions in the exam negatively affect females' performance on other questions. Indeed, when females are assessed using tests that have more multiple-choice and fewer open-ended questions they perform much worse than males in all questions.

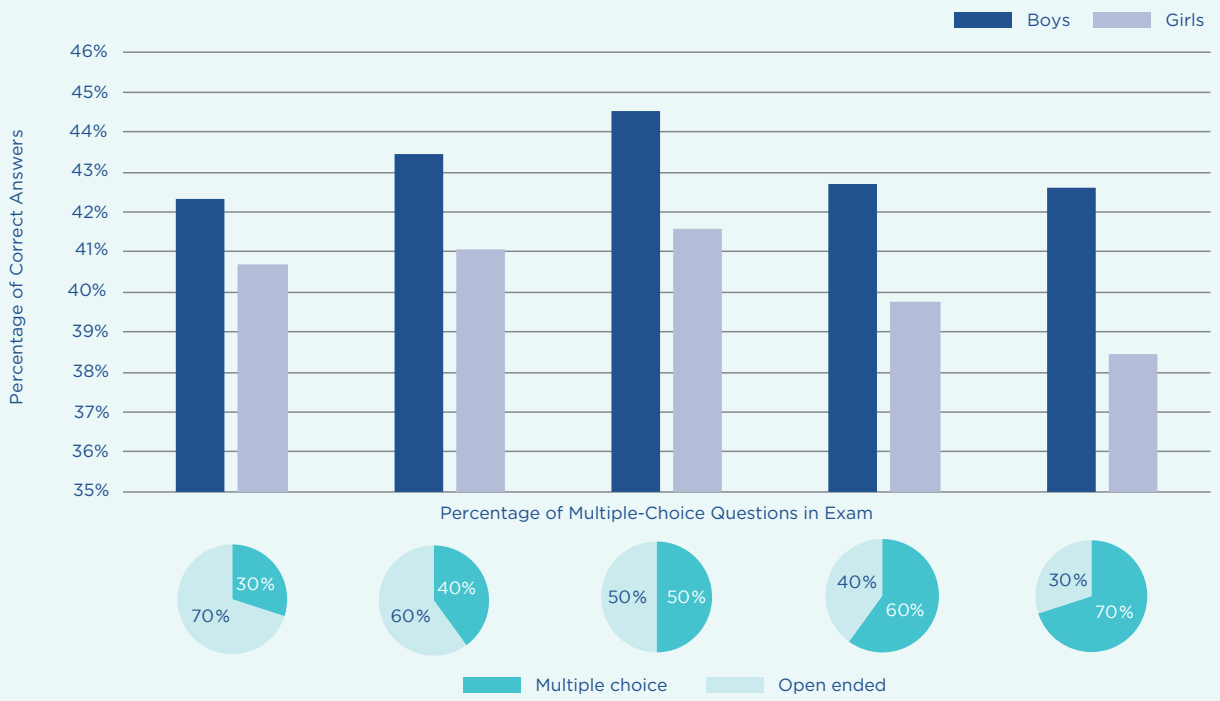
Key Insights

1 Males' advantage in math increases in exams that have a higher proportion of multiple-choice questions.

On average, boys receive higher scores relative to girls on NAPLAN and PISA math tests. Previous studies have shown that this difference can be partially explained by females' lower level of engagement in mathematics than males. Indeed, especially among top-performing students, females are less likely to think math is important and less likely to pursue math courses (Buckley 2016). However, the types of questions that students receive in their exams can themselves contribute to females' underperformance in math.

This can be demonstrated by using a random variation that exists in the PISA test. In the year 2015, the math tests were randomly distributed to students. Some students received a test that had mostly (i.e. 70 per cent) multiple-choice questions. Other students received a test with lower reliance on multiple-choice questions (i.e. 30 per cent). Figure 1 depicts the average performance by gender and test structure (proportion of multiple-choice and open-ended questions). Females performed worse, placing behind males, noticeably more when they receive an exam booklet with 60 per cent or more multiple-choice questions. This means that the gender gap in math scores widens by 50 per cent in favour of boys when the share of multiple-choice questions increases by 10 percentage points (e.g. from 50 per cent to 60 per cent).

Figure 1: Performance in Mathematics by Gender and by Proportion of Multiple-Choice Questions



Notes: This graph uses data from the PISA 2015 computer-based assessment

2 Multiple-choice tests cause girls to answer less carefully and to skip questions more often.

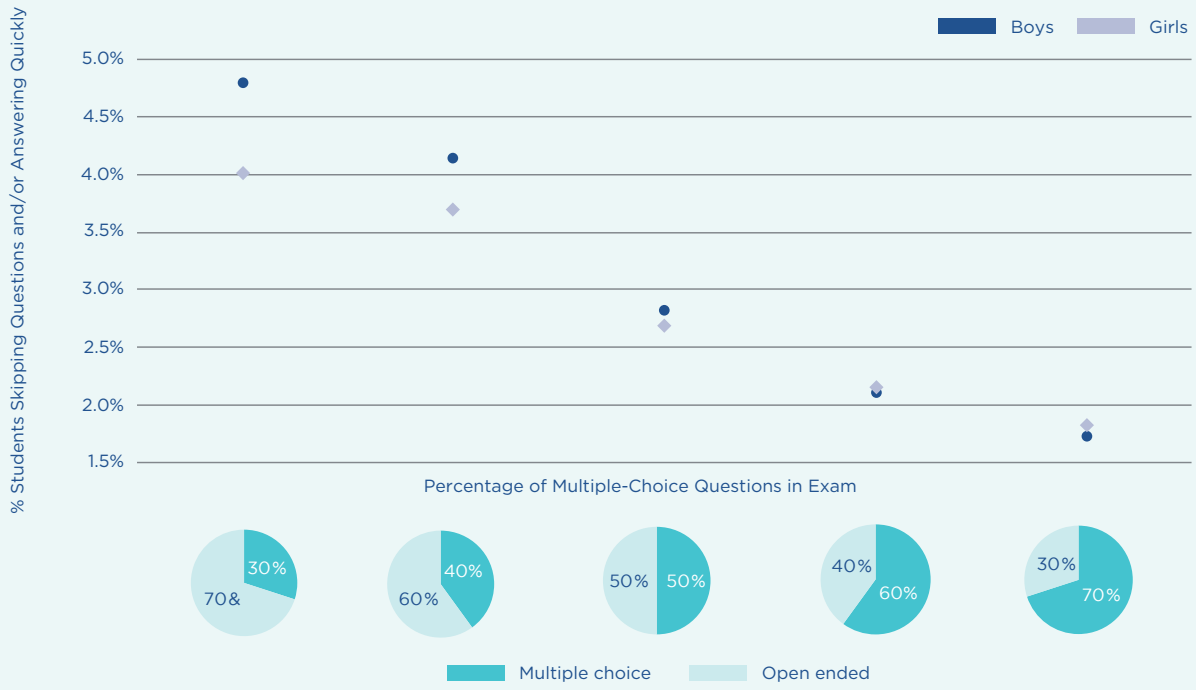
Is this difference in performance in multiple-choice questions tied to the approach students take to answer the questions? The PISA test allows us to explore this issue through its tracking of the time taken to respond to a question as well as its recording of omitted questions. Using the time taken to answer, PISA allows us to identify students who answer questions too fast (e.g. in under three seconds, which does not even allow for careful reading of the question). Indeed, answering questions too fast or skipping them entirely can be seen as a sign of low effort and inattentiveness.[1]

The findings demonstrate that girls who receive an exam with more multiple-choice questions are more likely to show a lack of effort during the test. In general, boys are more disengaged in the test than girls. Yet, this difference in engagement tends to revert for students that receive an exam with a higher proportion of multiple-choice questions.

Indeed, on average girls are 1.3 per cent less likely than boys to take the exam carelessly. But, a 10-percentage-point increase in the share of multiple-choice exam questions decreases this gap to 1 per cent. As girls are more likely to present analytical procedures and solutions while boys are more likely to just state the answers (Liu and Wilson 2009), these results suggest that more multiple-choice questions could have a disengaging effect on girls.

1 The PISA does not penalise students who answer multiple-choice questions incorrectly. Moreover, the test allows more than enough time to complete the test. This means that students are always better off attempting all questions.

Figure 2: Percentage of Students Skipping Questions or Answering Questions Quickly by Gender and by Proportion of Multiple-Choice Questions



3 The increase in gender gaps associated with more multiple-choice questions is greater for students whose mothers do not work in STEM occupations.

Can we better explain the gender gap in performance on multiple-choice questions-based tests? Performance on multiple-choice questions is likely to depend on students' confidence and beliefs about their mathematics knowledge. For example, a higher level of confidence affects how fast students can rule out incorrect responses.

PISA did not provide us measures of students' level of confidence in 2015. To overcome this issue, this study uses maternal occupation as a measure of girls' level of confidence and beliefs in mathematics. Indeed, previous literature has shown that girls with mothers working in STEM occupations are more confident in mathematics and less likely to believe the stereotypes that boys are better than girls in mathematics (van der Vleuten 2018).

The evidence shows that the negative effect of multiple-choice questions on females' performance disappears among girls whose mothers work in STEM-related occupations. This suggests that, especially among females with a low level of confidence in mathematics, multiple-choice exams may not be the most appropriate tools to measure their level of knowledge. Indeed, these types of exams capture students' characteristics, such as their confidence, rather than just their mathematics knowledge.

Can standardised tests be more gender equal?

Multiple-choice questions are considered impartial, and widely adopted as they save time and effort for graders. The findings, however, point out that these questions may not be a gender-neutral way to compare students' competencies. There are at least three steps test designers and policy makers could take to reduce gender gaps in standardised tests.

Increase girls' confidence in their math ability

The lack of confidence in girls when it relates to their math skills undermines their test performance and could end up making gender stereotypes in math performance a self-fulfilling prophecy. Previous research shows that exposure to successful women in math (Porter, 2020) and giving objective feedback about students' ability (Moore, 2008) improves girls' confidence in math. Together with the results on girls with mothers in a STEM-occupation, this suggests that providing strong role models in math for girls can be a very effective policy for decreasing gender test scores and improving outcomes for women.

Prepare students to undertake standardised tests

Standardised tests are different in many dimensions to other assessments that students are used to. Testing students using standardised tests is a way to familiarise them with the format of tests and allow them to perform at their best. Suggestive evidence from the paper behind this Research Insight shows that boys' advantage in multiple-choice questions disappears among students in schools that prepare their students to take standardised tests and used standardised tests often as a tool to test students' competencies.

Greater balance between multiple-choice and open-ended questions in tests

A greater balance among multiple-choice and open-ended questions in tests such as NAPLAN and ATAR would likely decrease gender gaps in test scores. In addition, including more open-ended questions would give more weight to the analytical process and thinking on tests – an arguably more accurate measure of skills. While multiple-choice questions are less costly and allow for impartial grading (and because of this they are often preferred to open-ended questions in tests), they might be painting a gender-skewed picture of student ability and achievement. This could have large social and economic costs, due to the loss of talented and skilled women.

Further Information

This work was part of Silvia's Doctoral Dissertation in Economics at the Department of Economics at the University of Melbourne and the Melbourne Institute: Applied Economic & Social Research. Silvia is currently a Job Market Candidate. Her research lies in the area of labour economics and the economics of education and gender. A link to her website, including the latest version of the research paper behind this Research Insight, can be found at https://www.silviagriselda.com/.

Datasets and methods

This research uses primary data from The Programme for International Student Assessment (PISA) in 2015. The PISA dataset contains measures that capture student performance on standardised tests, and socio-economic characteristics of the students, teachers and school information.

The PISA assessment tests mathematics, reading and science competencies of a representative sample of 15-year-old students in more than 60 countries, including Australia.

The tests last about three and half hours and are low-stakes; they do not directly affect students' school grades, graduation, or other outcomes. Yet, the results are used by education policymakers to compare students' performance across countries, and over time. In 2015, students took a computer-based assessment. The data include students' answers to each question and the time they took to answer them.

Authors

Silvia Griselda
Melbourne Institute: Applied Economic & Social Research

Research Insights produced by the Melbourne Institute provide a clear and practical understanding of contemporary economic and social issues in Australia.

Supported by high-quality academic analysis, each Research Insight aims to make sense of complex issues to enable evidence-based decision making for policy and practice.

melbourneinstitute.unimelb.edu.au

Standardised Tests in Australia

In Australia, students take several standardised tests, which are all of national importance. The National Assessment Program – Literacy and Numeracy (NAPLAN) test is administered nationwide to all Australian students and used by government authorities, schools, and parents to compare and evaluate students' reading, writing, language and numeracy skills from Year 3 to Year 9. Australian students in New South Wales sit the Higher School Certificate (HSC) exams in their last two years of high school and their scores are then used to compute each student's Australian Tertiary Admission Rank (ATAR), a measure that universities use to grant admissions. Besides, the PISA international test, which is used in the analyses above, assesses the performance of 15-year-old students in mathematics, reading and science. These results are used by policymakers to compare the performance of Australian students over time and to that of other countries.

References & Endnotes

1. Buckley, S. (2016). Gender and sex differences in student participation, achievement and engagement in mathematics. *Melbourne: Research Developments*, ACER.
2. Coffman, K. B. (2019). Stereotypes and belief updating. *Harvard Business School.*
3. Moore, D. A. (2008). The trouble with overconfidence. *Psychological Review,* 502.
4. Porter, C. a. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 226-54.
5. van der Vleuten, M. a. (2018). Intergenerational transmission of gender segregation: How parents' occupational field affects gender differences in field of study choices. *British Educational Research Journal*, 294-318.