MATHEMATICS COMPETITIONS, GENDER, AND GRADE LEVEL: DOES TIME MAKE A DIFFERENCE?

Gilah C. Leder, David G. Pederson and Graham H. Pollard

La Trobe University – Bundoora, University of Canberra and University of Canberra

In this paper we draw on data from a large mathematics competition, for the years 1987 to 2000 and use two different but closely related measures to investigate possible gender differences in performance. Our analyses revealed that small gender differences in favour of males persisted but had decreased over time. Consistent with reports from previous studies, gender differences in performance were more marked at senior high school than at junior high school grades.

INTRODUCTION

Gender differences in performance on mathematical tasks and participation in post compulsory mathematics courses have attracted much attention over the past three decades. A careful reading of the literature reveals that there is considerable overlap in the performance of males and females (see, e.g., Fennema, 1974; Leder, 2001). Friedman's (1995) appraisal: "while gender differences in mathematics are small and apparently decreasing over time, they still exist" (p. 22), offers an economical summary of the major research findings. However, when achievement is reported in terms of (usually low-stake) classroom grades, females are often rated slightly higher than males (Kimball, 1989). Gender differences in performance, most often in favour of males, continue to be reported when above average performance is considered, for students in advanced post compulsory mathematics courses, and on selected mathematical tasks assessed through standardised or large scale testings. For example, data from the large *Third International Mathematics and Science Study* [TIMSS], in which 41 countries and some 15,000 schools participated, revealed that there were few differences in average mathematics achievement by gender in grades 4 and 8 but that there were substantial gender differences in mathematics achievement in favour of males in grade 12 (Mullis, Martin, Fierros, Goldberg, & Stemler, 2000). These authors further argued:

> The trends in achievement by gender are so pervasive across countries and the sampling procedures employed so rigorous that a clear pattern can be discerned across primary, middle, and secondary school. The gender gap in achievement becomes larger as students progress through school in most countries (Mullis et al, 2000, p. 5).

Findings from a recent large scale testing program in the USA (National Assessment of Educational Progress [NAEP], 1999) point to a more pervasive performance difference on that instrument. Those data revealed a consistently higher performance by males at three age levels, 9, 13, and 17, with the difference largest for the oldest age group.

The TIMSS data, like many studies before it, indicated that performance can be affected by question content:

> Internationally, in mathematics, males tended to perform higher than females on items employing spatial reasoning, reading maps and diagrams, as well as problems involving

percentages or area. Females tended to perform higher on items requiring common algorithms. (Mullis et al, 2000, p. 98)

Males' higher performance on items involving geometry and topology are frequently reported (see, for example, Hyde, Fennema, & Lamon, 1990), although again there is some evidence that the magnitude of performance differences appears to be decreasing (Friedman, 1995).

In addition to question content, it has also been shown that the format of assessment may affect apparent gender differences in mathematics achievement. On average, males – as a group - seem to do better than females on multiple-choice items, but not on unstructured items or on those which require an essay-type response (see e.g., Halpern, 2002; Leder, Brew, & Rowley, 1999).

In the remainder of this paper we draw on a unique and large data base, the Australian Mathematics Competition [AMC]. In earlier explorations of the AMC data some gender differences in performance were found (Leder & Taylor, 1995; Taylor, Leder, Pollard, & Atkins, 1996). Here we examine, for data spanning the years 1987 to 2000, whether gender differences in performance continue to be found, whether they varied with grade level and whether any differences found were consistent over time. (We also examined whether gender differences in performance were affected by question topic area: arithmetic, algebra, geometry, and "other". However, space constraints do not allow a description of the coding used to define question category, nor of the metric devised to correct for possible differences in correct response rates for different topic areas and needed to enable a realistic comparison to be made of the performance means for different topic areas.)

The scope and format of the AMC are described in the next section.

THE AUSTRALIAN MATHEMATICS COMPETITION [AMC]

The AMC began in 1978. Each year three papers are set: one for students in grades 7 and 8, one for students in grades 9 and 10, and one for students in grades 11 and 12. These are known as the Junior, Intermediate, and Senior papers respectively. Females and males have been approximately equally represented in the entries for the Junior and Intermediate papers, but each year more boys than girls have elected to sit for the Senior paper.

Students are given 75 minutes to answer each paper, which contains 30 questions, and are asked to choose the correct response from a set of five alternative responses. Each of the first ten questions, the second ten questions and the third ten questions in each paper are awarded 3 marks, 4 marks and 5 marks, respectively, for a correct response. One quarter of the marks assigned to a question for a correct response is deducted for an incorrect response.

The Competition has become both a national and international event. More than 90% of Australia's high schools and some 30% of eligible students (i.e., over half a million students) now participate. As well, over the years students from an increasing number of other countries have entered the Competition, with students from 38 different countries doing so in 2000.

Because of the large number of students attracted to the Competition, the organisers elect to use questions which are readily able to be computer marked, i.e., multiple choice questions. Much care is taken by the problems committee in designing the actual items to be used. Basic manipulation arithmetic, algebra, and geometry questions are included, as are routine and non-routine problems from the same domains. Some questions are closely linked to work likely to have been covered in class. Other items are intentionally expected to be unfamiliar to the students sitting for the Competition papers. The acknowledged limitations of the Competition papers - multiple choice questions to be answered in a limited period of time – must be balanced against the extensive penetration of the Competition into the Australian school population and thus the large and diverse group of students reached by the Competition papers.

A COMMENT ON CONTEXT

Reducing gender inequities has been a high priority, over the past three decades or so, in Australia as well as in many other countries. Means to achieve this have included grants to schools to initiate special intervention programs, media campaigns to encourage females to continue with mathematics and enter traditional male fields which rely on strong mathematical background, and putting in place legislation to address discriminatory practices in fields such as education, employment, and welfare. However, during the 1990s, increasing concerns began to be voiced about boys' educational performance (see, e.g., Forgasz & Leder, 2001). In Australia, these concerns led to the publication of the influential report *Boys: Getting it right. Report on the inquiry into the education of boys* (House of Representative Standing Committee on Education and Training, 2002). A list of recommendations to improve the quality and educational environment for students, and for boys in particular, is included in the report.

More boys than girls still elect to take the most demanding mathematics subject offered at the senior high school level. However, performance data presented in the report indicated that, as a group, girls now outperform boys in almost all subjects examined in state wide examinations held at the end of high school (grade 12). In mathematics, too, girls – on average – obtain a higher mark than do boys. These findings are at variance with the performance data for large scale testings reported at the beginning of this paper. It is noteworthy that items found on the grade 12 examination papers include short answer items as well as more open-ended items which require a description of the process used to reach a solution, as well as reaching the solution *per se*.

This brief sketch indicates the context in which the longitudinal data, described in the remainder of this paper, were gathered.

THE STUDY

Retrievable AMC data were available for the years 1987 to 2000 and so the analysis was on the performance results for Australian students for those 14 years. As indicated above, the aim was to determine whether gender differences were found, varied with grade level and changed over time.

The total number of questions posed on the three papers over the time period considered was 1260, the product of 14 (years) x 3 (papers) x 30 (questions)[1]. Since each question paper was attempted by students at two different grade levels, there were thus 2520 *items of information* available for analysis.

Measuring gender differences in performance – operational definitions

Two measures of gender difference were used.

- One measure was the difference in the percentage of males (MC) and females (FC) who chose the correct response for a given item, i.e., (MC – FC). This measure is denoted by (M-F), focuses solely on correct responses, and does not distinguish between omitted and incorrectly answered items since both are treated as incorrect answers.
- The second measure was the difference in the percentages of males (MC|R) and females (FC|R) who selected the correct response for a given item, given that they choose a response for that item, i.e., [MC|R - FC|R]. This measure is denoted by (M-F)|R and excludes omitted items.

THE RESULTS

Measuring gender differences in performance – differences over time

As described earlier, for each question there were five alternative responses. The probability of choosing the correct response for any item by chance was thus 0.2. We therefore considered FMC|R, the percentage of females and males combined who chose the correct response to a question, given that a response was chosen, and eliminated from our analyses all items for which FMC|R was less than 20%, since it was considered that those questions would not provide useful information on the difference in achievement between males and females. This reduced the initial data set from 2520 to 1964 *items of information*.

<u>Comparison of two seven-year periods</u>

To allow possible changes in performance over time to be explored, we clustered the 14 years of performance data into two: from 1987 to 1993 – designated as Time 1 or T1 - and from 1994 to 2000 – designated as Time 2 or T2, and calculated gender differences in performance in terms of the two measures described earlier. For example, for students in grade 7, the mean of (M–F) for Time 1 was 2.43 and for Time 2 the corresponding mean was 2.15. There was therefore a decrease in the mean gender difference of 0.28 percentage points from one seven year period to the next. Similar calculations for each grade and both measures of gender difference gave the means shown in Table 1. These data reveal that mean gender differences in performance (in favour of males) were consistently less for Time 2 (the years 1994 to 2000) than for Time 1 (1987 to 1993). The effect sizes (Cohen, 1988) corresponding to the differences in means for Time 1 and Time 2 are also shown in Table 1. They were consistently less than 0.2, i.e., consistently small according to Cohen's definition.

---

[1] However, each year some questions were used in more than one paper so that the number of *different* questions attempted by students was only 906.

Although the results for the two measures of difference were similar to one another, the data in Table 1 further illustrate that the index chosen for measuring gender differences in performance can influence the apparent magnitude of that difference.

Table 1: Gender difference in performance over two periods: 1987 to 1993 (Time 1) and 1994 to 2000 (Time 2), separately for six grades.

Grade	Variable							
	(M–F)				(M–F)|R			
	Time1	Time2	T2 – T1	Effect size	Time1	Time2	T2 – T1	Effect size
7	2.43	2.15	-0.28	-0.073	1.77	1.48	-0.29	-0.078
8	2.75	2.41	-0.34	-0.088	2.12	1.82	-0.30	-0.081
9	3.38	3.06	-0.32	-0.089	2.61	2.39	-0.22	-0.060
10	4.57	3.99	-0.58	-0.148	4.10	3.37	-0.73	-0.189
11	4.64	4.37	-0.27	-0.071	4.42	3.98	-0.44	-0.109
12	6.11	6.06	-0.05	-0.010	6.37	6.13	-0.24	-0.055

Change over a 14-year period: from 1987 to 2000

On the assumption that the mean gender difference in performance was linearly related to time, the means in Table 1 were used to estimate the percentage changes in the mean gender difference from 1987 to 2000. For example, for students in grade 9, (M–F) for Time 1 (centred on 1990) was 3.38 and for Time 2 (centred on 1997) was 3.06. The estimated annual change in the mean of (M–F) was thus (3.06 – 3.38)/7 = – 0.046. The fitted value for (M–F) for 1987 was (3.38 + 3(0.046)) = 3.52 and the fitted value for (M–F) for 2000 was (3.06 – 3(0.046)) = 2.92. The estimated percentage change in (M–F) from 1987 to 2000 was therefore –17%. Similar calculations for each grade and for both measures of difference gave the percentages shown in Table 2.

Table 2: Estimated percentage change in the mean gender difference from 1987 to 2000

Grade	(M – F)	(M–F)|R
	Percentage change	Percentage change
7	-20	-29
8	-22	-25
9	-17	-15
10	-22	-30
11	-11	-18
12	-1	-7

For both measures, the difference in the performance of males and females was less in 2000 than in 1987, with the change being generally larger for (M–F)|R than for (M – F). For the latter, the difference was approximately 20% for students in grades 7 to 10, but smaller for students in grades 11 and 12.

Measuring gender differences in performance – differences by grade

Grade related gender differences in performance, for the 14 year period or for the *items of information* available for analysis, are summarised in Table 3. For both measures of gender difference the mean difference in favour of males increased markedly from grade 7 to grade 12. Except at grade 12, the gender difference in performance was larger for (M – F) than for (M-F)|R, i.e., larger when omitted answers were counted as incorrect responses.

Table 3: Mean gender difference from 1987 to 2000 for two measures, by grade

Grade	Measure	
	(M-F)	(M-F)|R
7	2.29	1.62
8	2.57	1.97
9	3.20	2.48
10	4.27	3.73
11	4.50	4.20
12	6.09	6.25

An example

As already indicated, some questions are used on more than one AMC paper. In 1993, the following question appeared on the Junior, Intermediate, and Senior AMC paper and was thus attempted by students in grades 7, 8, 9, 10, 11, and 12.

On my flight from Christchurch to Sydney, the following is shown on the information screen in the passenger cabin:

Current speed 864 km/h

Distance from Departure 1222km

Time to Destination 1 h 20 min

If the plane continues at the same speed, then the distance in kilometres from Christchurch is closest to

(A) 2300 (B) 2400 (C) 2500 (D) 2600 (E) 2700

Student performance on this question, at each grade level, is summarised in Table 4.

The data in Table 4 indicate that

- for both males and females, the percentage of students with a correct answer increased with grade level;
- more males than females obtained the correct answer at each grade level;

and

- the difference in the percentage of males and females with the correct answer increased with grade level.

Although the two measures gave consistent results, the magnitude of the difference in mean performance in favour of boys was less for (M-F)/R than for (M-F), i.e., was less when comparison of performance was restricted to items actually attempted by students.

Table 4: Gender difference in performance for one question, at 6 grade levels

	Grade					
Measure of gender difference	7	8	9	10	11	12
MC [% of males who chose the correct response]	23.6	29.4	37.0	41.3	50.0	56.5
FC [% of females who chose the correct response]	18.2	23.2	28.8	31.7	40.1	44.6
(M-F) [defined as MC – FC]	5.4	6.2	8.2	9.6	9.9	11.9
(MC)|R [% of males who chose the correct response, given a response was chosen]	31.8	36.6	43.5	48.0	55.7	61.6
(FC)|R [% of females who chose the correct response, given a response was chosen]	28.0	32.3	37.8	42.0	49.6	54.2
(M-F)|R [defined as (MC)|R - (FC)|R]	3.8	4.3	5.7	6.0	6.1	7.4

A FINAL COMMENT

The AMC is a popular and carefully devised multiple choice mathematics problem paper, widely attempted by students in grades 7 to 12 in Australia as well as in a range of other countries. In this paper we examined Australian data gathered over a 14 year period. Our explorations confirmed that gender differences in mathematics performance in favour of boys persist, at least on multiple choice questions such as those found on the AMC papers, that these differences in performance appear to be decreasing over time, and that they are far more marked for students in the upper secondary grades than for those in the lower secondary grades. These findings are at variance with other (Australian) test data which indicate that males' performance in mathematics, as well as in various other subjects, is lower than that of females. Differences in the types of items found on the different test papers, and differences in the format of response required from students, may account for the different findings.

Use of two different measures for calculating gender differences gave consistent results which nevertheless varied in the strength of the differences observed. Thus reports of the magnitude of gender differences in performance on mathematics problems may well be affected by the choice of metric used for quantifying such differences.

References

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates

Fennema, E. (1974) Mathematics learning and the sexes: A review. *Journal for Research in Mathematics Education, 5*(3), 126-139

Forgasz, H. J., & Leder, G. C. (2001)The Victorian Certificate of Education – a Gendered Affair? *Australian Educational Researcher, 28*(2), 53-66

Friedman, L. (1995) The space factor in mathematics: Gender differences. *Review of Educational Research, 65*(1), 22-50

Halpern, D.F. (2002) Sex differences in achievement scores. Can we design assessments that are fair, meaningful, and valid for girls and boys*? Issues in Education, 8*(1), 1-19

House of Representative Standing Committee on Education and Training (2002, October) *Boys: Getting it right*. *Report on the inquiry into the education of boys*. Canberra: Commonwealth of Australia

Hyde, J.S., Fennema, E., & Lamon, S.J. (1990) Gender differences in mathematics performance: A meta analysis. *Psychological Bulletin*, *107*, 139-155

Kimball, M.M. (1989) A new perspective on women's math achievement. *Psychological Bulletin*, *105*, 198-214

Leder, G.C. (2001) Pathways in mathematics towards equity: a 25 year journey . In M. van den Heuvel-Panhuizen (Ed.) *Proceedings of the 25th conference of the International Group for the Psychology of Mathematics Education*, (Vol 1, pp. 41-54). Utrecht, the Netherlands: Freudenthal Institute, Faculty of Mathematics and Computer Science, Utrecht University

Leder, G. C., Brew, C., & Rowley, G. (1999) Gender differences in mathematics achievement: Here today and gone tomorrow? In G. Kaiser, E. Luna, I. Huntley (Eds.) *International comparisons in mathematics: The state of the art* (pp. 213 - 224). London: Falmer Press

Leder, G., & Taylor, P. (1995) Mathematics and gender: A question of testing? In B. Grevholm & G. Hanna (Eds.) *Gender and mathematics education. An ICMI Study* (pp. 271-280). Lund: Lund University Press

Mullis, I.V.S, Martin, M.O., Fierros, E.G., Goldberg, A.L., & Stemler, S.E. (2000) *Gender differences in achievement. IEA's third international mathematics and science study*. Chesnut Hill, MA: TIMSS International Study Centre

National Assessment of Education Progress [NAEP] (1999) *NAEP trends in academic progress*. Washington: US Department of Education

Taylor, P. J., Leder, G.C., Pollard, G..H., & Atkins, W..J. (1996) Gender differences in mathematics: Trends in performance. *Psychological Reports, 78*, 3-17