Gender Differences in Children's Mathematics Achievement: Perspectives from the Longitudinal Study of Australian Children

Colin Carmichael
Charles Sturt University
<ccarmichael@csu.edu.au>

With reports of declining enrolments in mathematics related degrees and low female participation rates in these degrees, the issue of gender differences in mathematics remains relevant. Results of recent studies suggest gender differences in mathematics are nuanced and that small differences in the early years can manifest as larger differences in later years. This study explores differences in teachers' ratings of children's achievement across a number of mathematical content domains. It is based on observations from the K-cohort of the Longitudinal Study of Australian Children in 2006, when the children were aged between six and seven, and in 2008, when they were aged between eight and nine. Gender differences in achievement are analysed using the Mantel-Haenszel procedure associated with the implementation of the Rasch model. Results indicate that teachers rate girls higher on tasks related to data, whereas they rate boys higher on tasks related to place-value and computation. Implications of these findings are discussed.

Australia is producing fewer graduates in the Science, Technology, Engineering and Mathematics (STEM) disciplines than are needed and this may be exacerbated by continuing differences in gender-participation rates. A recent report from Australia's Office of the Chief Scientist (Chubb, Findlay, Du, Burmester, & Kusa, 2012) noted that the proportion of first-degrees awarded in the STEM disciplines in Australia (18.8%) was much lower than in China (52%) and Japan (64%). Moreover, recent statistics from the annual Graduate Destination Survey (Graduate Careers Australia, 2012) indicate that fewer females than males are entering these occupations, suggesting that relatively low female participation rates exacerbate national skill-shortages in STEM related occupations.

Gender differences in STEM participation rates are likely influenced by students' experiences with mathematics at school, even in their early years. Although mathematics is compulsory in Australia up to the end of Year 10, participation in post-compulsory mathematics courses is declining, with a greater decline for females (Forgasz, 2006). Factors influencing this declining participation reportedly include: student's previous achievement in mathematics; their mathematics self-concept; their interest; and, their perceptions regarding the usefulness and difficulty of mathematics (McPhan, Morony, Pegg, Cooksey, & Lynch, 2008). In relation to mathematics achievement, a recent meta-analysis suggests that there is little or no evidence that male and female mathematics achievement means differ (Lindberg, Hyde, Petersen, & Linn, 2010). Nevertheless, the evidence suggests that there is a difference in male and female mathematics achievement variances (Robinson & Lubienski, 2011; Strand, Deary, & Smith, 2006) in that there are more males than female in the upper (and lower) percentiles of the achievement distribution. In addition to this, growth trajectories in mathematics for males appear to be steeper than for females (Leahey & Guo, 2001; Vale et al., 2011). As a result, small gender differences in the early years may manifest themselves as larger differences in later school, especially in the upper quintiles of the achievement distribution that feed into STEM related careers. This study, therefore, explores the presence or otherwise, of gender differences in the mathematics achievement of children in the early primary years.

In V. Steinle, L. Ball & C. Bardini (Eds.), Mathematics education: Yesterday, today and tomorrow (Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia). Melbourne, VIC: MERGA.

Background

Factors that influence gendered differences in mathematics achievement have both genetic and environmental sources. Males tend to display an advantage in tasks requiring visual-spatial skills (Tzuriel & Egozi, 2010), whereas females display an advantage in tasks requiring verbal skills (Strand, et al., 2006). These findings point to neurological differences between males and females that have been supported by brain imaging showing distinct gender differences in brain activation during mathematical calculations (Keller & Menon, 2009). In addition to this, recent research has linked spatial abilities with prenatal and/or neonatal levels of sexual hormones (Berenbaum, Korman-Bryk, & Beltz, 2012). Eco-cultural factors, however, play a major and sustained role in gender differences. Gender stereotypes regarding mathematics, for example, are known to impact on achievement and career choice (Kiefer & Sekaquaptewa, 2007).

Gender differences in mathematics achievement appear to be quite nuanced, in that students' responses differ according to the mathematical content, age of child and the type of question and method used to assess their achievement. In their study of early primary aged children, Vale et al. (2011) reported that males outperformed females in place-value and computation, yet Strand et al. (2006) reported that older females outperformed male peers in algebra and computation. Features of the problems used to assess students and the required solution strategy for those problems also appear to create different levels of difficulty for males and females. Lowrie and Diezmann (2011), for example, reported that males perform better than females on pr oblems containing graphics that required the decoding of information along horizontal or vertical continuums. Further, Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, and Morely (2000) reported that females perform better than males on problems that require conventional, algorithmic based methods of solution.

Given the earlier discussion, this study seeks to identify gender differences in the mathematics achievement of early primary school children across a range of mathematical content. More specifically it examines teacher ratings of achievement on overall mathematics achievement, and then examines specific mathematical content and skills. The study uses teacher ratings for pragmatic reasons but also because they are arguably a more valid source of data for younger children than standardised test scores (Soler & Miller, 2003).

Method

The study is based on a secondary analysis of data obtained from the Longitudinal Study of Australian Children (LSAC), details of which are reported in Sanson, Nicholson, Ungerer, Zubrick, and Wilson et al. (2002). LSAC utilises a cross-sequential design to follow two cohorts of children: a Birth (B) cohort of approximately 5000 children aged between 6 and 12 m onths of age; and a Kindergarten (K) cohort of approximately 5000 children aged between 4 years 6 months and 5 years. Moreover, it uses a stratified-cluster design that provides a large representative sample of the Australian population of children. This study is based on t he K-cohort from LSAC and the responses of their teachers to modified versions of the Academic Rating Scale (National Center for Educational Statistics, nd) in 2006 when they were aged between six and seven years and in 2008, when they were aged between eight and nine years. Further details of the student sample, instruments used, and analyses undertaken are discussed below.

Student sample

Of the 4983 children first recruited into the K-cohort, 4464 remained in Wave 2 during 2006, and 4331 remained in Wave 3 during 2008. During both of these waves, teachers of participating students were requested to provide a number of data on the study child including ratings of their mathematical achievement. Not all teachers provided these data, with only 3632 teacher responses in Wave 2 and 3643 responses in Wave 3. Of the students for whom teacher responses were available, 51.2% were male in Wave 2 and 51.6% in Wave 3.

Instruments

Teachers were asked to assess their student's proficiency to a number of items using a 5-point ordinal scale that ranged from 1 (*Not yet*) through to 5 (*Proficient*). There was also an additional category for not applicable, but responses in this category were assumed to be independent of the student and treated as missing values. Actual items used in both waves are shown, in abbreviated form, in the following table where those used in Wave 2 have codes prefixed by a six (the year of the wave) and those used in Wave 3 an eight. Full versions of significant items are provided in the discussion, with all items available from the Australian Institute of Family Studies (2006, 2008).

Table 1
Items used in study

Code	Item
6ars1	Continue a pattern using three items
6ars2	Demonstrates an understanding of place value
6ars3	Models, reads, writes and compares whole numbers
6ars4	Counts change with two different types of coins
6ars5	Surveys, collects and organises data into simple graphs
6ars6	Makes reasonable estimates of quantities
6ars7	Measures to the nearest whole number using common instruments
6ars8	Uses a variety of strategies for maths problems
8ars1	Creates and extends patterns.
8ars2	Uses a variety of strategies to solve math problems.
8ars3	Recognises properties of shapes and relationships among shapes.
8ars4	Uses measuring tools accurately.
8ars5	Shows understanding of place value with whole numbers.
8ars6	Makes reasonable estimates of quantities and checks answers.
8ars7	Surveys, collects and organises data into simple graphs.
8ars8	Models, reads, writes and compares fractions.
8ars9	Divides a 2 digit number by a 1 digit number.

Data analysis

A Rasch rating scale model (Andrich, 1978), was applied to the eight items used in Wave 2 and the nine used in Wave 3 to create two interval measures of children's mathematics achievement. Gender differences in mean achievement levels were assessed before individual items were examined for evidence of differential item functioning (DIF) by gender. The latter was achieved with the software package Winsteps (Linacre, 2006), using the Mantel-Haenszel procedure described in (Linacre & Wright, 1989). The procedure estimates the difficulty of each item were it presented to males and females separately and then calculates the magnitude of any difference in these difficulties, together with tests for the statistical significance of any difference. As recommended by Andrich and Hagquist (2012), items that displayed the greatest absolute DIF were temporarily removed from the analysis to ascertain whether reported DIF in other items was still evident.

Results

Seven of the eight items in Wave 2 (items 6ars2 through to 6ars8) were used to form a measure of children's mathematics in 2006 that explained 85% of the variance in responses and reported a reliability of $\alpha = 0.97$. Similarly, eight of the nine items in Wave 3 (Items 8ars1 through to 8ars8) were used to form a subsequent measure of these children's mathematics achievement that explained 84% of the variance and reported a reliability of $\alpha = 0.97$. The two discarded items (6ars1 and 8ars8) reported infit statistics outside the acceptance range of 0.8 through to 1.3 (Keeves & Alagumalai, 1999). Both items were more specific than the others in the scales and tended to elicit erratic responses that did not conform to the Rasch model's requirements.

For both waves, male mean achievement scores were higher than female means and there were larger than expected proportions of males in the top deciles. In 2006 the mean achievement score for males was 0.08 logits higher than that of females, and in 2008 it was 0.23 logits higher for males. These differences, however, were not statistically significant at the 5% level. In addition to this comparison of means, the proportion of males in the top decile of the distributions was compared with the proportion of males in the entire group. In 2006, males accounted for 54.8% of the children in the top decile, whereas they accounted for 51.2% overall. In 2008, they accounted for 57.5% of children in the top decile whereas they accounted for 51.6% overall. The latter result is statistically significant at the 5% level.

DIF analysis was then conducted on the seven remaining items from the Wave 2 questionnaire and the eight from the Wave 3 questionnaire in two separate analyses. Results for the analysis are reported in Table 2, which shows the number of male and female respondents for each item, the estimated item difficulties based on the male and female samples, the difference in these estimates and the significance of this difference. Negative differences indicate the item was easier for males than for females. A Bonferroni adjustment was applied, meaning that only differences with a reported *p*-value of 0.00 were regarded as statistically significant at the 5% level.

As is seen from the table, Items 6ars2 (place value) and 6ars5 (data) from the 2006 test displayed evidence of DIF, although the difference in difficulty in both cases is quite small. When the item with the greatest absolute DIF (Item 6ars2) was removed from the analysis, Item 6ars5 still demonstrated evidence of DIF, suggesting both items possessed real as opposed to artificial DIF (Andrich & Hagquist, 2012). Similarly in 2008, Items 8ars2 (problem-solving), 8ars5 (place value) and 8ars7 (data) displayed evidence of DIF. When the item with the greatest absolute DIF (8ars7) was removed from the analysis, however, Item 8ars5 failed to show DIF, suggesting its initial inclusion may be incorrect and caused by the procedure itself (Andrich & Hagquist, 2012). Item difficulty differences for Items 8ars2 and 8ars7 can be regarded as slight to moderate (Linacre & Wright, 1989).

It should also be noted that there were a number of missing responses and responses that were assigned as not applicable. The former ranged from 74 (2%) in Item 8ars7 down to 31 (0.8%) in Item 6ars3, whereas the latter ranged from 337 (9.3%) in Item 6ars7 down to 11 (0.3%) in Item 8ars1. An analysis of the characteristics of these missing values, including those assigned not applicable, revealed no noticeable gender influence.

Table 2
Results of DIF analysis

	Males		Females		Differences		
Item	Number of responses	Item difficulty	Number of responses	Item difficulty	Difference in difficulty (DIF)	*t*	*p*
6ars2	1748	-0.98	1740	-0.75	-0.23	-3.68	0.00
6ars3	1774	-1.23	1768	-1.15	-0.08	-1.12	0.26
6ars4	1647	1.08	1642	1.17	-0.09	-1.40	0.16
6ars5	1755	-0.21	1748	-0.40	0.19	3.05	0.00
6ars6	1770	-0.35	1768	-0.37	0.02	0.36	0.72
6ars7	1609	1.05	1636	0.89	0.16	2.47	0.01
6ars8	1774	0.58	1761	0.58	0.00	-0.03	0.98
8ars1	1847	-1.25	1722	-1.05	-0.20	-2.57	0.02
8ars2	1847	0.22	1723	0.64	-0.42	-5.93	0.00
8ars3	1821	-0.11	1687	-0.07	-0.04	0.43	0.66
8ars4	1838	-0.28	1707	-0.46	0.18	2.56	0.01
8ars5	1842	-1.24	1716	-1.03	-0.21	-2.78	0.00
8ars6	1812	0.81	1697	0.78	0.03	0.44	0.66
8ars7	1798	0.17	1692	-0.31	0.48	6.68	0.00
8ars8	1760	1.52	1645	1.42	0.10	1.50	0.06

Discussion

In general, the results based on children's overall achievement tended to conform to cited research. In line with findings from Lindberg et al. (2010), mean levels of mathematics achievement for these children did not differ significantly by gender. There was, however, evidence that a greater than expected proportion of males occupied the upper reaches of the achievement distribution in the second wave. In addition to this, the magnitude of the male/female achievement difference, albeit small, appeared to increase during the two year period encapsulated by these two waves, supporting the notion that male achievement trajectories may exceed those of females.

The results from the DIF analysis confirm the view that discernible gender differences in mathematics are influenced by content. In both waves, girls tended to achieve higher than boys in aspects related to data (items 6ars5 and 8ars7). In these items teachers were asked to rate their student's ability to survey, collect and organise data into simple graphs. In 2006 (Item 6ars5) the statement was followed by the example "make tally marks to represent the number of boys and girls in the classroom, or making a bar, line, or circle graph to show the different kinds of fruit children bring to school for lunch and the quantity of each type". In 2008 (Item 8ars7), where the DIF was greater, the statement was followed by the example "charts temperature changes over time, or makes a bar graph comparing the population in different cities in Australia, or interprets a pictograph in which each symbol represents 5 people". Given the findings from Lowrie and Diezmann (2011) that boys tend to perform better than girls on problems that require encoding along vertical and horizontal continuums, such as bar and column graphs, it is difficult to suggest why the teachers tended to rate girls higher on these items. One hypothesis is that the predominantly female population of teachers in primary schools select contexts for their statistics lessons that have more interest for girls than boys, reflecting reported gender differences in interest for statistics (Carmichael & Hay, 2009). This may lead to gendered differences in engagement and ultimately achievement.

In line with earlier reported findings from Vale et al. (2011), the results suggest that teachers rate the achievement of boys higher than girls on tasks related to place value. This was evident in 2006 (Item 6ars2) where teachers were asked to rate their student's ability to demonstrate an understanding of place value (e.g. by explaining that fourteen is ten plus four, or using two stacks of ten and five single cubes to represent the number 25). One hypothesis for this bias in favour of boys is that they find it easier than girls to interpret the relationships in the place-value chart, which is commonly used when introducing and developing place-value concepts. This is supported by the findings from Lowrie and Diezmann (2011), discussed above.

Teachers in both waves were asked to assess their student's ability to use a variety of strategies to solve math problems (items 6ars8 and 8ars2). In 2006, the statement was followed by the example "using manipulative materials, using trial and error, making an organised list or table, drawing a diagram, looking for a pattern, acting out a problem, or talking with others", whereas in 2008, it was followed by the example "adds 100 and then subtracts 2 when doing the mental math problem 467+98". A moderate DIF was detected in the 2008 item, where the example focuses on computational strategies rather than problem-solving in general. This result tends to agrees with findings from Vale et al. (2011), who reported boys achieved higher in computation. Teachers did not, however, perceive any gendered difference in general problem-solving skills during 2006. This suggests that the gender differences reported by Gallagher et al. (2000) may emerge later in schooling, if indeed they still exist.

Study limitations

The study involved a secondary analysis of data, which prevents any experimental control of the variables. In addition, the observations were analysed on the basis that they were obtained from a simple random sample. The sample, however, was based on a stratified-cluster sample and the Rasch method of analysis used in the study, does not cater for this design. Further, the measures of achievement are based on teacher assessments rather than results in standardised tests and this could contribute to the reported differences.

Conclusion

The study has sought to examine teacher ratings of children's achievement in early primary school for evidence of gender differences. Unlike many of the studies reported in the paper, this study is based on a large representative sample of the population of Australian primary school children. Consequently evidence of items favouring one gender over another is unlikely to be attributed to differences in the sample, although teacher bias needs to be considered. Some of the results of the study are supported by other studies, adding weight to the suggestion that teacher bias was minimal. Surprisingly, the findings indicated that teachers rated girls' achievement in data higher than boys in both waves, that is when the children were aged between six and nine years old. That this result was found from the ratings of two different groups of teachers is noteworthy and suggests that further research into the teaching of data in the early years is needed. In addition, the results of the study, together with results from Vale et al. (2011), suggest a gender difference in the way children learn about place-value. Given the fundamental nature of this concept, further research on the learning of place-value may be warranted.

Acknowledgements

The author would like to acknowledge all involved in the LSAC study, which was funded as part of the Australian Government's Stronger Families and Communities Strategy by the Australian Government Department of Housing, Families, Community Services and Indigenous Affairs (FaHCSIA).

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Andrich, D., & Hagquist, C. (2012). Real and artificial differential item functioning. *Journal of Educational and Behavioral Statistics, 37*(3), 387-416.

Australian Institute of Family Studies. (2006). Wave 2, 2006, Teacher questionnaire: LSAC. Available from: http://www.growingupinaustralia.gov.au/

Australian Institute of Family Studies. (2008). Wave 3, 2008, Teacher questionnaire: LSAC. Available from: http://www.growingupinaustralia.gov.au/

Berenbaum, S. A., Korman-Bryk, K. L., & Beltz, A. M. (2012). Early androgen effects on spatial and mechanical abilities: Evidence from congenital adrenal hyperplasia. *Behavioral Neuroscience, 126*(1), 86-96.

Carmichael, C. S., & Hay, I. (2009). Gender differences in middle school students' interests in a statistical literacy context. In R. Hunter, B. Bicknell & T. Burgess (Eds.), *Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 89-96). Wellington: MERGA Inc.

Chubb, I., Findlay, C., Du, L., Burmester, B., & Kusa, L. (2012). Mathematics, engineering and science in the national interest. Canberra, ACT: Office of the Chief Scientist.

Forgasz, H. (2006). Australian year 12 "Intermediate" level mathematics enrolments 2000-2004: Trends and patterns. In P. Grootenboer, R. Zevenbergen & M. Chinnappan (Eds.), *Proceedings of the 29th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 211-220). Canberra: MERGA.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., & Morely, M. (2000). Gender differences in advance mathematical problem solving. *Journal of Experimental Child Psychology, 75*, 165-190.

Graduate Careers Australia. (2012). GradStats: Employment and salary outcomes of recent higher education graduates. Available from: http://www.graduatecareers.com.au/

Keeves, J. P., & Alagumalai, S. (1999). New Approaches to Measurement. In G. N. Masters & J. P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 23-42). Oxford: Pergamon.

Keller, K., & Menon, V. (2009). Gender differences in the functional and structural neuroanatomy of mathematical cognition. *NeuroImage, 47*, 342-352.

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes. *Psychological Science, 18*(1), 13-18.

Leahey, E., & Guo, G. (2001). Gender differences in mathematical trajectories. *Social Forces, 80*(2), 713-732.

Linacre, J. M. (2006). WINSTEPS Rasch measurement computer program (Version 3.61.2) [Computer Software]. Chicago: Winsteps.com.

Linacre, J. M., & Wright, B. D. (1989). Mantel-Haenszel DIF and PROX are equivalent! *Rasch Measurement Transactions, 3*(2), 52-53.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123-1135.

Lowrie, T., & Diezmann, C. (2011). Solving graphics tasks: Gender differences in middle-school students. *Learning and Instruction, 21*, 109-125.

McPhan, G., Morony, W., Pegg, J., Cooksey, R., & Lynch, T. (2008). Maths? Why Not? (Report): Australian Department of Education, Employment and Workplace Relations (DEEWR).

National Center for Educational Statistics. (nd). Early Childhood Longitudinal Study - Kindergarten (ECLS-K). Washington, DC: Department of Education.

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal, 48*(2), 268-302.

Sanson, A., Nicholson, J., Ungerer, J., Zubrick, S., Wilson, K., Ainley, J., et al. (2002). Introducing the Longitudinal Study of Australian Children, (LSAC Discussion Paper No. 1). Melbourne: Australian Institute of Family Studies.

Soler, J. & Miller, L. (2003). The struggle for early childhood curricula: A comparison of the English foundation stage curriculum, Te Wha "riki and Reggio Emilia. *International Journal of Early Years Education, 11*(1), 57-68.

Strand, S., Deary, I., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK national picture. *British Journal of Educational Psychology, 76*(3), 463-480.

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81*(5), 1417-1430.

Vale, C., Davidson, K., Davies, A., Hooley, N., Loton, D., & Weaven, M. (2011). Using assessment data: Does gender make a difference. In J. Clark, B. Kissane, J. Mousley, T. Spencer & S. Thornton (Eds.), *Proceedings of the 34th annual conference of the Mathematics Education Research Group of Australasia* (pp. 736-743). Alice Springs: MERGA.