The Assessment of Females in Mathematics

MEGAN CLARK, SHARLEEN FORBES AND THORA BLITHE

Abstract:

Many assessment procedures in mathematics have been shown to disadvantage females in some situations. The authors show that females prefer questions set in “environmental” or “people” contexts and indicate that they do better on familiar, rather than problem-solving, questions. Performance of male and female students (both Maori and non-Maori) gives clear evidence that in secondary school national mathematics examinations, and in some university mathematics papers, females do better in the internal assessment than in the examination while the reverse is true for males. The New Zealand practice of moderating internal assessment by examination results is therefore weighted against female students. All assessment of mathematics performance should incorporate a number of different procedures in order to provide a more authentic measure of various mathematical abilities.

The international debate about the appropriateness of some assessment procedures has, in part, been stimulated by the differential performance of male and female students in traditional examinations.

In New Zealand there have been a number of analyses of secondary school mathematics examinations (Stewart, 1981; Reilly et al., 1987; Forbes, 1988; Morton et al., 1988, 1989 and 1993; Forbes et al., 1990; Blithe et al., 1993; Forbes and Mako, 1993; Forbes, 1994) which all show consistent (but not always significant) gender differences in favour of males. There is a greater range of achievement within each gender than between the genders but typically the top grades are dominated by males. As Table 1 indicates, for Bursary Mathematics with Statistics, these differences are not consistent across school authorities – state, integrated, or private, or school type – single sex or co-educational.

Table 1 Bursary Mathematics With Statistics, 1992 Results

<table>
<thead>
<tr>
<th>School Authority</th>
<th>State</th>
<th>Integrated</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single</td>
<td>Co-ed</td>
<td>Single</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>Sex</td>
<td>Sex</td>
</tr>
</tbody>
</table>

- **Males No.**
  - Single: 1710
  - Co-ed: 3896
  - Mean Score %: 55.03

- **Females No.**
  - Single: 496
  - Co-ed: 178
  - Mean Score %: 50.87

- **Total**
  - Single: 2939
  - Co-ed: 7129
  - Mean Score %: 53.67

Source of all tables: authors’ research.

In addition, performance is related to the amount of mathematics studied. Not surprisingly, the students taking both Bursary mathematics papers perform better on average than those taking one paper only. This is the case for both genders. However, a smaller percentage of females (31 percent in 1992) than males (48 percent in 1992) take both papers (Burgess, 1994).

Assessment

Assessment in any area of education is intended to evaluate the students’ understanding, knowledge and/or skills gained through their course of study and can be an invaluable diagnostic tool for both teachers and students.

There are a number of forms of assessment in current use in mathematics. These include:

- **timed/speed tests** (problems need to be solved within a set period of time);
- **multi-choice tests** (students choose from yes/no, true/false, or a range of options);
- **essays**;
- **individual projects** (usually done over a period of days or weeks);
- **group projects**;

176 Megan Clark, Sharleen Forbes and Thora Blithe
mastery learning (one level has to be ‘mastered’ before progressing to the next);
- two or three hour written examinations; and
- internal assessment (generally done by teachers throughout the course and often involving a combination of some or all of the above methods).

A limited amount of research has been done comparing assessment methods to determine those which may best suit women, Maori or ethnic minorities. However, women themselves cannot be classified as just one group. Forbes (1992) showed that a reduction in gender differences in performance in mathematics of one group of the New Zealand population (Pakeha) does not necessarily lead to a similar reduction in another group (Maori). International studies have shown that gender differences vary across different countries (Hanna, 1988).

Curriculum-Assessment Interface

In many countries, the gender-specific nature of mathematics curricula has been a topic of debate for a number of years. Concerns that have been expressed include:
- that the subject areas selected either for inclusion or for emphasis reflect “male” priorities; (Nancy Shelley (1992) has pointed out that the emphasis and content of many calculus courses, particularly at tertiary level, reflect military needs);
- that the topics, or mathematical processes, given highest status are those that more males than females succeed in (Kimball, 1989; Clark, 1993); and
- that the language, exemplars and illustrations in curriculum documents, guide notes and textbooks reflect the male world (Northam, 1986).

New Zealand’s current national mathematics curriculum, introduced in 1993 (Ministry of Education, 1992a), spans from new primary school entrants to the end of secondary schooling. The curriculum contains five strands: number, measurement, geometry, algebra and statistics all nested in an over-arching mathematical processes framework. Three components of mathematical processes are identified: problem solving, developing logic and reasoning, and communicating mathematical ideas. Problem solving has become the current trend in mathematics education internationally. However, many studies have shown males are advantaged by items which require problem solving skills or an application of mathematics (Hyde, Fennema and Lamon, 1990) whereas females do well in areas such as computation or algebra (Willis, 1990; Wily, 1986; Wilder and Powell, 1989). Thus the current emphasis on processes is likely to disadvantage females.

Although the curriculum cannot be seen independently from the assessment processes associated with it, curriculum change is not always accompanied by assessment change. In her study of Asian-Pacific (APEC) countries Megan Clark (1994a) notes that the widespread use of multiple-choice test items tends to present mathematics in an absolutist way which is likely to deter females and conflicts with the desire of most APEC members for a curriculum that is inclusive of females.

Assessment procedures not only affects the mathematics content in the curriculum but also the emphasis given to different sections of the curriculum. In addition, they can influence the context in which the curriculum is taught. Context can affect both performance and question preference. Female students demonstrate a preference for test items set in a real-world context (Clark, 1994b). There is a body of evidence showing that females perform better on test items with a “feminine” or gender neutral context (Rosser, 1989; Wilder and Powell, 1989; Clark, 1993).

Performance in mathematics is also related to other subjects taken. In 1989, the New Zealand Universities Entrance Board, with respect to Scholarships examinations, stated that:

the large majority of scholarships are gained by students who enter two of their three scholarship papers in Chemistry, Mathematics with Calculus, Mathematics with Statistics, or Physics.

There is some content overlap between these subjects. Both the subjects and subject combinations are dominated by males. In his analysis of 1992 Bursary mathematics performance, Burgess (1994) found that:

combination of subjects is the most influential factor on performance in mathematics.

An understanding of the interface between assessment and curriculum is critical in any analysis of students’ abilities. We need to determine the mathematics skills measured by particular assessment methods, and to evaluate whether these are, in fact, the skills we desire future users of mathematics to have.
Overview of Assessment Procedures

Gender differences in mathematics performance are not constant across different forms of assessment. Many of the differences in favour of males noted in the international literature have been in timed, written tests. Meredith Kimball (1993) suggests that if classroom grades were used instead of standardised examinations to measure mathematics performance then gender differences would be in favour of females. Jan de Lange (1987) noted that male and female students perform much the same on both oral and "take-home" tasks.

The Traditional Mathematics Examination

Timed written tests have dominated mathematics assessment particularly at the higher levels but also in our primary schools (Carr, 1994). Within this structure there are a number of different areas which have been shown to impact differentially on the performance of females. These include:

Method of test administration

The method of test administration can affect levels of performance and gender differences. Bourke and Stacey (1988 cited in Willis, 1990) showed that achievement was higher overall when test questions were presented individually than when given as one single test, and that the improvement in marks was greater for females.

Question choice

Mathematics examinations often offer students a choice of questions. The intention is to improve the "fairness" of the examination. But male and female students make different choices.

In the Bursary Mathematics with Statistics examination females opt more for statistics than general mathematics questions. In Mathematics with Calculus they prefer complex numbers and differentiation questions both first introduced in the previous year. The preference shown by female students for statistics questions in the Mathematics with Statistics examination may be partially explained by the fact that fewer of them take both mathematics papers. The overlap between the papers is entirely in general mathematics (Forbes et al., 1990; Morton et al., 1993).

At the tertiary level in New Zealand gender differences in question choices within statistics courses have been found, with females preferring and having most success in questions set in either "people" or "the environment" context. In pure mathematics papers females demonstrated a significant preference over males for some standard, predictable, "safe" questions but their performance on these questions was similar to that of males (Blithe et al., 1993).

Exam layout

The order that questions, or sections of questions, occur in examination papers may influence the order in which students answer questions. Many New Zealand teachers suggest to students that when they have a choice they should do an "easy" question first, in order to settle themselves down. Students do show a slight tendency to do better on questions answered early in an examination (with no apparent gender differences) but this is confounded by a tendency to answer questions in exactly the order presented. In Bursary Mathematics with Statistics where females prefer and perform better on the statistics questions, the paper has for a number of years been laid out with all the general mathematics questions preceding all the statistics questions. Students who attempt questions in the order in which they occur may be disadvantaging themselves.

Even within a single question, layout may be an important consideration. Clark (1994b) reported that when first-year university statistics students were given a table containing male data in the left column and female data in the right, female students were significantly more likely to "reach across" the table and compute the female measures first, leading in some cases to error.

Multi-choice tests

A number of studies in different countries indicate that males are favoured by multi-choice questions in mathematics (Wildier and Powell, 1989; Bolger and Kellaghan, 1990).

Essay questions

Essay questions are little used in mathematics tests; however, females perform better in these questions (Eccles and Jacobs (1986), Slack and Porter (1980) both cited in Willis, 1990; Ministry of Education, 1992b). In the 1987 and 1988 New Zealand Bursary Mathematics papers one of only two questions where the gender difference in means was significantly in favour of females was an essay question (Morton et al., 1993).
Context
From 1990-1993 in a first-year Victoria University mainstream statistics course Probability and Data Analysis male performance was better overall. This was most marked in abstract questions (problems presented entirely in formal or symbolic notation). Of the non-abstract questions a large proportion involved material likely to be more familiar to males than to females (e.g., computing, machinery) and the rest were judged to be equally accessible. No parts of the questions came from what we might term the “world of women”.

In 1991, in another first-year statistics course, students were given a problem with two options available: analysis of psychological test scores or strengths of concrete. The mathematics content of the two choices was identical. 76 percent of the 217 students took the first option presented, consistent with previous results that students tend to answer questions in the order presented. Women were more likely (50 percent) to take the second option than men (32 percent) if the first was about “concrete”. These results support the idea that particular contexts are effective with different subsets of the students and are consistent with Purser and Wily’s (1992) work with fifth-formers in New Zealand. They illustrated:

how much the students, particularly girls, were influenced by the gender orientation of the context of the question.

Time restrictions
There is evidence that eliminating time restrictions in tests can improve the performance of female students (Rosser, 1989).

Comparison of Assessment Methods
Internal assessment versus written examination
Two investigations were undertaken to compare mathematics performance in written, national secondary school examinations with that attained in internal assessment.

(a) The first secondary school investigation compared the performance of individual students in both School Certificate and Sixth Form Certificate mathematics. The former is usually a three-hour end-of-year written examination. The latter is internally assessed (using a variety of means including tests and projects). In School Certificate, students receive one of seven grades (D, C2, C1, B2, B1, A2, A1 in ascending order). In Sixth Form Certificate students are assigned a grade from 9 (lowest) to 1 (highest) for their year’s work. Sixth Form Certificate grades available for schools to allocate to students are determined by the school’s overall performance in the previous year’s School Certificate (that is, Sixth Form Certificate is moderated by School Certificate). The higher the School Certificate grade in mathematics the more likely students are to continue with Sixth Form Certificate mathematics; that is, approximately 90 percent of those awarded B1 or higher, 50-65 percent of those getting B2 and <50 percent of those getting a C or lower. However, in both 1991 and 1992, regardless of the mark they obtained in School Certificate, females were less likely to continue in mathematics than males.

Table 2(i): 1992 Sixth Form Certificate Mathematics Grades by 1991 School Certificate Grade: Males

<table>
<thead>
<tr>
<th>SC Grade</th>
<th>Ethnicity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Maori</td>
<td>26</td>
<td>39</td>
<td>21</td>
<td>11</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>43</td>
<td>28</td>
<td>20</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1051</td>
</tr>
<tr>
<td>A2</td>
<td>Maori</td>
<td>6</td>
<td>15</td>
<td>22</td>
<td>29</td>
<td>21</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>5</td>
<td>14</td>
<td>27</td>
<td>31</td>
<td>18</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>2185</td>
</tr>
<tr>
<td>B1</td>
<td>Maori</td>
<td>5</td>
<td>18</td>
<td>39</td>
<td>27</td>
<td>9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>1</td>
<td>6</td>
<td>23</td>
<td>35</td>
<td>26</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td>3183</td>
</tr>
<tr>
<td>B2</td>
<td>Maori</td>
<td>1</td>
<td>4</td>
<td>22</td>
<td>35</td>
<td>26</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>1</td>
<td>5</td>
<td>20</td>
<td>37</td>
<td>26</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>1764</td>
</tr>
<tr>
<td>C1</td>
<td>Maori</td>
<td>11</td>
<td>36</td>
<td>44</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>2</td>
<td>5</td>
<td>33</td>
<td>34</td>
<td>25</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>244</td>
</tr>
<tr>
<td>C2</td>
<td>Maori</td>
<td>17</td>
<td>33</td>
<td>50</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>11</td>
<td>17</td>
<td>39</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>D</td>
<td>Maori</td>
<td>50</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Non-Maori</td>
<td>33</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Total No. Maori</td>
<td>17</td>
<td>34</td>
<td>48</td>
<td>92</td>
<td>186</td>
<td>183</td>
<td>111</td>
<td>42</td>
<td>1</td>
<td>714</td>
<td></td>
</tr>
<tr>
<td>Total No. Non-Maori</td>
<td>581</td>
<td>635</td>
<td>101</td>
<td>155</td>
<td>187</td>
<td>166</td>
<td>777</td>
<td>331</td>
<td>8</td>
<td>8447</td>
<td></td>
</tr>
</tbody>
</table>
The value-added component of the sixth form year (i.e., the change in a student’s knowledge or skills over the period of study) was analysed by comparing the Sixth Form Certificate male and female grade distributions for 1991 and 1992 (for those students who had sat School Certificate in the previous year) within each School Certificate grade. As expected, there is a strong correlation between the two grades. That is, students with an A1 in School Certificate are the most likely to receive a grade 1 in Sixth Form Certificate in the following year. Tables 2 (ii), (iii) show the grade distributions for these students in 1992. A similar pattern occurred in 1991. For all School Certificate grades above C1 (about 45 percent) females were found to do significantly better in Sixth Form Certificate mathematics than males. The numbers of Maori students are much smaller but when there is a significant gender difference this is also in favour of females. There are no obvious consistent differences between Maori and non-Maori students.

Although overall a higher percentage of males attain the highest grade (A1), this is because fewer females than males achieve high School Certificate grades, and within each School Certificate grade fewer females proceed to sixth form mathematics. Within each School Certificate grade, however, both females and Maori are taking a bigger share of this internally assessed cake than would be expected from their previous written examination performance (Blite et al., 1993; Forbes, 1994).

(b) The second secondary school investigation compared the two methods of assessment in the same paper. The Bursary Mathematics with Statistics paper consists of:

- 80 percent end-of-year three hour written examination;
- 20 percent internal assessment (format left to individual schools).

Although the internal assessment does vary between schools, in general it is based on one or more practical projects.

After the internal assessment marks are submitted by each school to the central agency responsible for the awarding of qualifications (New Zealand Qualifications Authority 1993), they are moderated (scaled to have the same mean and standard deviation that the school achieved in the written examination). This moderation of internal assessment by examination indicates a lack of trust in either the “fairness” of internal assessment as a procedure, or in the “fairness” of individual schools.

Forbes & Mako (1993) found that the gender difference in the mean overall (final) mark was significant and so was each contributing component. However, while the gender difference in mean examination marks is in favour of males, that for the mean internal assessment marks was in favour of females. It is because the examination mark dominates the final mark by 4:1 that the overall gender difference is in favour of males. Male and female students in state co-educational schools are in similar teaching environments. For these students, the difference in their ranking within their school on each assessment method was analysed. On average, female students were ranked higher in their school on the internal assessment marks than on the examination marks. The reverse was the case for male students.
At Victoria University of Wellington, three first-year mathematics/statistics courses were analysed as these also have a mix of examination and internally assessed components. First-year mathematics students can opt to take two Keller Plan courses, for Calculus and Algebra and Geometry, instead of the more traditional lectures followed by a three-hour examination. The assessment in the Keller Plan courses is a mixture of mastery based assessment and examination. In each course there are 13 modules of work. The students attempt mastery tests on these modules which require a pass mark of 80 percent and can be re-sat as many times as required to pass. Once each test is passed it contributes five marks towards the final grade. Nine modules are compulsory in each course and a pass in ten modules will ensure passing the course. The final two-hour test is optional and worth 35 percent.

Six years (1989-1994) of results in each course were analysed. Results from year to year were very similar. While significantly more females sit the full 13 modules and require fewer marks in the test to get an A grade, males are more inclined to complete fewer modules and therefore have to make up more on the final test. This they manage to do. The mean marks obtained in the final tests in either course are not significantly different for males and females. The overall final marks obtained by combining the test plus the module mark also show no significant difference.

By contrast, the other course, Statistics for Natural and Social Sciences, is a service course taken by a wide variety of students with many different major subjects, and with some having minimal mathematical preparation. Over the years studied, there was a higher participation in the course by women than men (1992, 64 percent female, 1993, 59 percent female, 1994, 59 percent female), which is unusual in both university mathematics and statistics courses. The course has a three-hour final examination and an internally assessed component consisting of five linked projects throughout the year all performed on a single data set (specific to an individual student). The material in these projects is within a social context, (data on mothers and newborns) and collaboration on the work is appropriate up to a certain point. Females do significantly better than males in both the traditional timed examination questions and in the projects but the difference is more marked, in all years, for the projects than for the final examination.

There is, therefore, evidence that the two methods, internal assessment and written examinations, each advantage different groups of students in mathematics. It may be that the assessment method itself is preferred by a particular group, or that each method evaluates different skills, or a combination of these. In particular, female students appear to do better in internal assessment than in written examination. There seems little justification for the moderation of one procedure by the other.

Attitudes to Success
Males and females view success in mathematics differently. Females habitually attribute their success to hard work or luck and their failures to lack of ability, whereas males see it the other way round (failure is attributed to lack of work or bad luck and success to ability). Female students have a decreased expectation of success in advanced work (Leder, 1984; Barnes et al., 1987). Little research has been done on the relationship between the testing method and attitudes to success. We still need to help female students to become aware of their abilities and to gain greater confidence.

Use of Student Assessment as a Monitoring Tool
A trend which is of concern is the use of results in national examinations as a measure not only of a student's performance, but also of a teacher's or of a school's performance. We must be extremely careful not to use inappropriate measures to evaluate teacher performance. Little is known of the interaction between "good" and "bad" teachers and students' performance on tests. When students' marks are used to assess teachers or schools this is usually done with scant regard for different intakes of students, or for the "value-added" component.

Summary
Gender differences in mathematics performance are closing but there are still fewer women than men continuing in mathematics. The following have been shown to disadvantage females in some mathematics assessment situations: timed tests, multi-choice tests, question content, question context, and order/layout of questions. Females prefer questions set in "environmental" or "people" contexts and they do better on familiar or standard, rather than problem-solving, questions. They perform well in essay-type questions, and are advantaged by internal assessment procedures rather than examinations. The moderation or "out-weighing" of internal assessment results by examination results is likely, then, to discriminate against females.

National examinations need continually to be assessed and monitored to ensure that females, and other groups, are not being disadvantaged by their format and context. However, while females
have been generally considered as a single group it is acknowledged that female subgroups (by ethnicity or socio-economic status) need further investigation. The assessment of mathematics performance should incorporate a number of different procedures in order not to disadvantage any single group of students and to provide a more authentic measure of various mathematical abilities.

References


Kimball, M. Contribution to panel discussion at ICMI Study 93 Gender and Mathematics, 1993.


The authors

Megan Clark is Associate-Professor in the Institute of Statistics and Operations Research and Director of the Mathematics and Science Education Centre at Victoria University of Wellington. She has been a secondary school teacher in Wellington and has three children. Her main interests are women and Pacific Island learners of mathematics.

Sharleen Forbes is Manager of Public Policy for Statistics New Zealand and previously worked for Te Puni Kōkiri (Ministry of Maori Development). She has been a secondary school teacher in Wellington and has four children. She has a particular interest in the Education of Maori.

Thora Blithe is a Senior Lecturer in Mathematics at Victoria University of Wellington. She has a particular interest in women in mathematics.

All three authors are founding members of the EIME (Equity in Mathematics Education) collective and collaborate on much of their research.