# DIFFERENCES BETWEEN THE SEXES IN MATHEMATICS ACHIEVEMENT IN AUSTRALIA

Tilahun Mengesha Afrassa John P. Keeves
The Flinders University Institute of International Education
The Flinders University of South Australia
<tilahun.afrassa@flinders.edu.au> <john.keeves@flinders.edu.au>

*In this paper an investigation is reported on whether the trend of sex differences in mathematics achievement of students at the lower secondary school level changed over time by bringing the mathematics achievement of the students to a common scale, which is independent of both the samples of students tested and the samples of items employed. The scale is used to examine the differences in mathematics achievement between the sexes in Australia over a 30 year period. Conclusions are drawn about such differences in mathematics achievement over time in Australia.*

## INTRODUCTION

During the past 30 years there is probably no other topic in Australian education that has attracted more discussion and more research, including widespread action research programs, than the issue of differences between the sexes in mathematics achievement in Australian schools. Reviews of research conducted in Australia by Leder and Forgasz (1992) and Barnes and Horne (1996) into gender and mathematics refer to "gender" rather than "sex" differences, and follow Megarry (1984) who argues for the use of "gender" to denote "the set of meanings, expectations and roles that a particular society ascribes to sex". Thus "gender" is seen to be a social construct, while "sex" refers to the biological category to which a person belongs. In this paper interest is primarily in recorded differences in mathematics achievement between males and females, the distinction being based on observable biological differences.

Leder (1992) has also reviewed changing perspectives in the area of gender as well as implicit sex differences in mathematics learning from an international viewpoint. However, this review does not examine factors that would account for gender differences both across countries and over time. Articles by Keeves (1973) and Baker and Jones (1993) address aspects of such differences which provide evidence that the measured differences between the sexes in mathematics achievement are at least in part socially based. It is only by the examination of such differences in naturally occurring circumstances that explanation can be advanced, since the conduct of randomized experimental trials is not possible.

Unfortunately much of the research into differences in mathematics achievement between the sexes in Australia has suffered from several serious shortcomings: (a) selection bias, since comparisons are made using groups that are not representative of a target population which is complete and has not suffered from self selection; (b) the sampling procedures employed are inadequate because, while large numbers of students are tested, they are drawn from too few schools; and (c) the estimation of error for significance testing fails to take into consideration the use of a cluster sample design. If sound comparisons are to be made for the detection of differences between the sexes in mathematics achievement, then large representative and random samples of students must be tested and significance tests with appropriate estimates of error must be employed.

The problems have been recognized for over 30 years of the inappropriate use in educational research of significance tests in situations where schools are selected as the primary sampling units and students are selected from within schools. Sometimes very crude estimates of error have been calculated which seek to make allowance for the use of a cluster sample design (for example, Keeves, 1968), although, in general, the problem has been ignored, because appropriate computer programs have not been available for the accurate estimation of sampling and measurement errors. In the past few years computer programs have become available, for example WesVarPC (Brick et al. 1997) and HLM (Bryk and Raundenbush, 1992), that can be readily used to estimate errors which would permit the testing of such data for statistical significance. Under these circumstances it would seem important for analyses to be undertaken of three data sets that involve large representative samples of Australian students assembled at different times over a period of 30 years in order to test whether there are significant differences between the sexes in mathematics achievement, and whether these differences have changed in magnitude and direction over time.

The aim of this paper is then to report whether differences can be detected between the sexes in mathematics achievement in Australia and whether such differences have changed in magnitude and direction over time as might be expected from the widespread programs that have been introduced in Australia to remove all aspects of discrimination associated with gender in schools particularly in the area of mathematics learning.

It is necessary to emphasize that several questions must be addressed before these analyses can be undertaken.

1 The samples selected must not be confounded by the effects of selection bias. This requires that the samples are drawn from schools across Australia, at a stage of schooling where the study of mathematics is compulsory and prior to the period where some students have dropped out from school or where the study of mathematics is optional.

2. It is necessary to establish that the instruments employed to measure mathematics achievement are assessing a single dimension and that measurement is not confounded by the presence of separate components of arithmetic, algebra and geometry. This requires verification that a single dimension of mathematics achievement is being assessed by the instruments employed.

3. In order to make meaningful comparisons it is desirable to bring the different measures of mathematics achievement on the different occasions to a common scale. This can be done by the use of Rasch scaling of the measures and the use of concurrent equating procedures, provided that items and persons satisfy the requirement of unidimensionality, and there are common items in the tests across the different occasions.

4. The issue of whether the differences in mathematics achievement that might be observed should be classified as "sex" or "gender" differences can only be considered after such differences have been shown to be present in the sets of scores under consideration. If no differences in mathematics achievement are detected after replication then no hypotheses can be meaningfully advanced as to the origin of such differences between the sexes. It is important to recognize that a failure to detect differences which can be considered as other than occurring by chance, should not be taken to indicate that no differences between the sexes occur in mathematics education, or that there are not differences between the sexes with respect to those aspects of mathematics achievement that have not been examined in this study.

## THE DATA EMPLOYED IN THIS STUDY

The data employed in this study were obtained from the First, Second and Third International Mathematics Studies conducted in Australia and reported by Keeves (1968), Rosier (1980) and Lokan, Ford and Greenwood (1996) respectively. These are all studies that have used large random and representative samples of students from schools across Australia, prior to the stage of schooling where the study of mathematics is not obligatory, and where attendance at school ceases to be compulsory.

In the First International Mathematics Study (FIMS), conducted in 1964, two groups of students participated, namely 13-year-old students in Years 7, 8 and 9 and students in Year 8 of schooling. In total, 4320 students took part in this study. In FIMS only government schools in New South Wales (NSW), Victoria (VIC), Queensland (QLD), Western Australia (WA) and Tasmania (TAS) participated. In the Second International Mathematics Study (SIMS), which was administered in 1978, nongovernment schools and the Australian Capital Territory (ACT) and South Australia (SA) were involved as well as those states that participated in FIMS. In this study 5120 students were tested.

Meanwhile, in the Third International Mathematics Study (TIMS), which was conducted in 1994, government and nongovernment school students in all states and territories including the Northern Territory were involved. In the TIMS study 12,850 students were tested at the lower secondary school level.

In 1964 and 1978, the samples were age samples and included students from Years 7, 8 and 9 in all participating states and territories, although in FIMS a Year 8 sample was also tested. In TIMS the samples were grade samples drawn from Years 7 and 8 or Years 8 and 9. In ACT, NSW, VIC and TAS students in Years 7 and 8 were selected while in QLD, SA, WA and NT samples were drawn from Years 8 and 9, enabling a Year 8 Australia wide sample to be derived for purposes of comparison with the 1964 data.

In order to make meaningful comparisons of mathematics achievement of boys and girls over time by using the 1964, 1978 and 1994 data sets, the following steps were taken. The 1964 students were divided into two groups 13-year-old students in one group (FIMSA) and all Year 8 students including 13-year-old students at that year level as the second group (FIMSB) since in addition to an age sample, a grade sample had also been drawn. It is important to observe that 13-year-old students in Year 8 were considered as members of both groups. In the first group, students were chosen for their age and in the second group for their year level. The 1978 students were chosen as an age sample and included students from both government and nongovernment schools. In order to make valid comparisons between the 1978 sample and the 1964 sample, the 1978 government school students were divided into two groups. The first group included all government students who participated in the study (SIMSG), and the second group included all government school students in the five states excluding students from SA and ACT (SIMSR). Meanwhile, in TIMS the students were chosen as a grade sample. The common sample group for all states and territories was the Year 8 students. In order to make the TIMS samples comparable with the other samples, only Year 8 government school students in the five states that participated in FIMS and SIMS are considered as the TIMSR data set in this study. While the samples employed in these three studies were large and the response rates were, in general, adequate, being in excess of 80 per cent for the FIMS and SIMS studies, there was some evidence of selection bias or response bias with respect to sex, in so far as male students were over represented in the FIMS and SIMS studies and under represented in the TIMS study.

## MEASUREMENT PROCEDURES EMPLOYED IN THE STUDY

In this study the procedures employed to compare the achievement differences between male and female students on the three occasions involved the use of the Rasch model to scale students' responses to the mathematics test items. The Rasch model has been shown to be the most robust of the item response models (Sontag, 1984), and was used in this study primarily to equate students' performance in mathematics on a common scale for the Australian investigations conducted in FIMS, SIMS and TIMS.

### Unidimensionality

In order to employ the Rasch model for calibrating the items in the mathematics tests it was necessary to examine whether or not the items satisfied the requirement of unidimensionality (Hambleton and Cook, 1977). If the items were found not to satisfy the condition of unidimensionality, it would not be possible to employ the Rasch procedures for the calibration of the tests. Furthermore, it would not be meaningful to compare differences between the sexes with respect to mathematics achievement since a total score would be meaningless.

Consequently, confirmatory factor analysis procedures were employed using LISREL 8 (Jöreskog and Sörbom, 1993) to examine the unidimensionality of the mathematics test items. Confirmatory factor analysis is a statistical procedure employed for investigating relations between a set of observed variables and the underlying latent variables (Byrne, 1989; Spearritt, 1997). The results of the confirmatory factor analyses of FIMS and SIMS data sets revealed that a nested model in which the mathematics items were assigned to a general higher order factor, which was labelled as Mathematics as well as to three specific correlated first-order factors of Arithmetic, Algebra and Geometry provided the best fitting model. In addition, in the confirmatory factor analyses undertaken, no evidence was found to reject the assumption of the existence of one general factor involved in the mathematics tests, in so far as in the nested model the Mathematics factor extracted more of the total variance than did the specific first-order factors taken together. Therefore, the mathematics test items in the FIMS and SIMS studies are considered to satisfy strongly the requirement of unidimensionality. The item cluster-based design procedure employed in the construction of the TIMS data sets would seem to preclude the use of confirmatory factor analysis to test the unidimensionality of the TIMS data set and confirmation of unidimensionality must be provided by the introductory steps in the Rasch analysis.

### Developing a common mathematics scale

The calibration of the mathematics test items with concurrent equating permitted a scale to be constructed that extended across the three tests, namely FIMS, SIMS and TIMS test items on the mathematics scale. The fixed point of the scale was set at 500 with one logit, the natural metric of the scale, being set at 100 units. The fixed point of the scale, namely 500 was taken as the mean of the difficulty level of the calibrated items in the FIMS test administered in 1964.

### Rasch Analysis

Three groups of students namely FIMS (4320), SIMS (5120) and TIMS (12,850) were employed in the calibration and scoring analyses. The necessary requirement for calibration in Rasch scaling is that the items and persons must fit the Rasch scale. In order to examine whether or not the items and persons fitted a unidimensional scale, it was important to evaluate both the item fit statistics and the person fit statistics. In FIMS two items (Items 13 and 29), in SIMS two items (Items 21 and 29) and in TIMS one item [(Item T1b No 148) with one item (Item 94), which was excluded from the international TIMSS analysis] were removed from the calibration analyses due to the misfitting of these items to the Rasch model with infit mean square values outside the range of 0.77 to 1.30 (Adams and Khoo, 1993). Moreover in this analysis no person was deleted, because the for all cases was less than 5 (Wright and Stone, 1979). However, students with a zero score or with a perfect score were automatically excluded from the calibration procedure, since they did not provide useful information for the purposes of scale calibration, although such students were included in the scoring of the data.

## COMPARISON OF SEX DIFFERENCE IN FIMS, SIMS AND TIMS

In this section the differences in mathematics achievement between the two sexes are examined for the three occasions. These comparisons for Australia in 1964 and 1978, differ from those carried out in previous studies (Keeves, 1968; Moss, 1982) in that proper account can now be taken of the complex design of the samples employed in testing for statistical significance. However, in testing for significant differences, while multiple comparisons are involved no use is made of the Bonferroni adjustment (Finn, 1997), because the thrust of the comparisons is more toward the detection of no differences, than towards the detection of highly significant ones.

### The t -test

In order to determine the level of statistical significance between the mean scores on FIMS, SIMS and TIMS in mathematics achievements a $t$ statistic was calculated, which took into account errors from three scores: (a) sampling error, (b) errors of calibration, and (c) equating error. Since the samples all involved a cluster sample design with schools sampled with a probability proportional to size at the first stage and students sampled from within schools at the second stage, it was necessary to use the WesVarPC (Brick et al. 1997) computer program to test the data for statistical significance, taking into account both the stratification and cluster sample design employed in all three studies.

### Effect size

In this paper both the standardized effect size and the magnitude of effect on the calibrated scales are used to examine the level of practical significance of the differences between FIMS, SIMS and TIMS in mathematics achievements over time.

In this study effect size values less than 0.20 are considered as trivial, while values between 0.20 and 0.50 are considered as small. Furthermore, effect size values between 0.50 and 0.80 are taken as moderate and values above 0.80 are treated as large (Cohen, 1992).

### Effect of mathematics learning in one year

Tilahun (1998) has undertaken analyses to estimate the extent of learning achieved by Australian students in the TIMS study in one year. It is estimated that the growth in mathematics achievement between Year 8 and Year 9 is 0.32 logits. From this information it is possible to examine the practical significance of differences between the male and female students in terms of a year of learning of mathematics.

## COMPARISONS OF SEX DIFFERENCES BETWEEN 1964 AUSTRALIAN STUDENTS

Table 1 and Figures 1 and 2 show the three comparisons which are considered in FIMS. The estimated mean score differences on the three comparisons are seemingly in favour of boys. This suggests that boys achieved at a higher level than girls. However, the effect size is trivial and t-values are non-significant. Thus, in all the comparisons the differences are neither practically nor statistically significant. It should be noted that Keeves (1968) using a very crude approximation for the design effect of the complex sample reported a significant difference. However, the procedures used currently in significance testing, make proper provision for the complex sample design using the jackknife routine in WesVarPC (Brick et al. 1997).

### Comparisons of Sex Differences between 1978 Australian Students

In the SIMS data set four comparisons are undertaken and presented in Table 1 and Figures 1 and 2. The comparisons are between boys and girls in government schools, nongovernment schools, in all schools, and the restricted sample of schools.

The mean score differences between the two sexes in government schools (both Government and Restricted) indicated that girls achieved at a higher level than boys. However, the

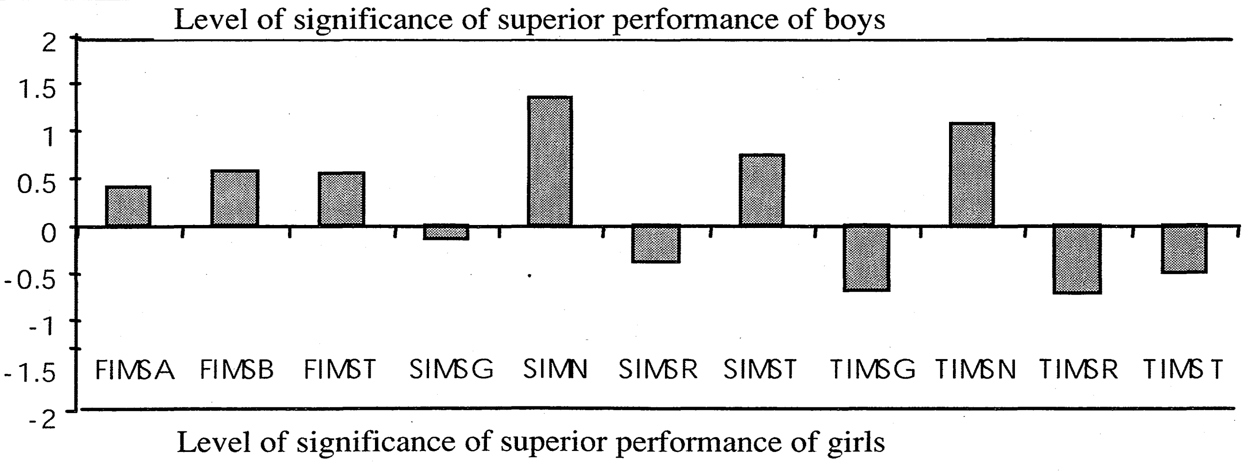
*Table 1*
*Descriptive statistics for mathematics achievement of all students for the three occasions by sex*

| | FIMS | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | FIMSA | | FIMSB | | Total | | | |
| | Male | Female | Male | Female | Male | Female | | |
| Mean | 462.0 | 457.0 | 455.0 | 448.0 | 459.0 | 453.0 | | |
| Standard Deviation | 101.0 | 90.0 | 87.0 | 76.0 | 98.0 | 87.0 | | |
| Jackknife Standard Error | 9.7 | 7.3 | 10.1 | 7.5 | 9.3 | 6.5 | | |
| Design Effect | 14.9 | 9.6 | 17.3 | 11.2 | 20.5 | 11.4 | | |
| Sample Size | 1530 | 1386 | 1619 | 1462 | 2275 | 2044 | | |
| | SIMS | | | | | | | |
| | Government | | Non-government | | Total | | Restricted | |
| | Male | Female | Male | Female | Male | Female | Male | Female |
| Mean | 442.0 | 443.0 | 488.0 | 464.0 | 453.0 | 448.0 | 440.0 | 443.0 |
| Standard Deviation | 105.0 | 100.0 | 117.0 | 97.0 | 110.0 | 99.0 | 105.0 | 99.0 |
| Jackknife Standard Error | 4.2 | 5.1 | 14.9 | 9.6 | 5.2 | 4.6 | 4.7 | 5.8 |
| Design Effect | 3.4 | 5.0 | 9.5 | 5.4 | 6.2 | 5.2 | 3.2 | 4.8 |
| Sample Size | 2095 | 1894 | 580 | 551 | 2675 | 2445 | 1614 | 1424 |
| | TIMS | | | | | | | |
| | Government | | Non-government | | Total | | Restricted | |
| | Male | Female | Male | Female | Male | Female | Male | Female |
| Mean | 423.0 | 431.0 | 488.0 | 474.0 | 443.0 | 448.0 | 422.0 | 431.0 |
| Standard Deviation | 125.0 | 124.0 | 117.0 | 110.0 | 126.0 | 121.0 | 126.0 | 126.0 |
| Jackknife Standard Error | 7.7 | 8.4 | 10.2 | 8.4 | 7.0 | 6.4 | 8.4 | 9.2 |
| Design Effect | 9.4 | 9.9 | 8.5 | 9.6 | 11.1 | 10.7 | 8.9 | 9.4 |
| Sample Size | 2479 | 2168 | 1111 | 1633 | 3590 | 3801 | 2030 | 1755 |

| | Mean difference | Effect size | t-value | Sign. level |
|---|---|---|---|---|
| FIMSA Male vs Female | 5 | 0.05 | 0.41 | Not Significant |
| FIMSB Male vs Female | 7 | 0.09 | 0.56 | Not Significant |
| FIMS Male vs Female | 6 | 0.06 | 0.53 | Not Significant |
| SIMS Government Male vs Female | -1 | -0.01 | -0.15 | Not Significant |
| SIMS Nongovernment Male vs Female | 24 | 0.22 | 1.35 | Not Significant |
| SIMS Male vs Female | 5 | 0.05 | 0.72 | Not Significant |
| SIMS Restricted Male vs Female | -3 | -0.03 | -0.40 | Not Significant |
| TIMS Government Male vs Female | -8 | -0.06 | -0.70 | Not Significant |
| TIMS Nongovernment Male vs Female | 14 | 0.12 | 1.06 | Not Significant |
| TIMS Male vs Female | -5 | -0.04 | -0.52 | Not Significant |
| TIMS Restricted Male vs Female | -9 | -0.07 | -0.72 | Not Significant |

Restricted = Those group of students in SIMS which are comparable with FIMSA, and students in TIMS which are comparable with FIMSB

*Figure 1*
*t-values for differences between boys and girls in Mathematics test score for FIMS, SIMS and TIMS*

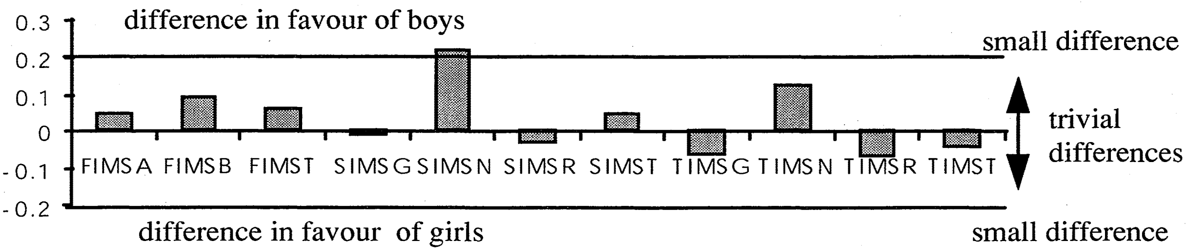


differences were neither practically nor statistically significant since the effect size and t-values were too small to be considered significant.

Furthermore, the estimated mean score difference between male and female nongovernment school students in SIMS is seemingly in favour of the boys, although the effect size (0.22) is small and the t-value (1.35) is non-significant. Thus, the difference is not statistically significant. Nevertheless, the achievement level of the boys is higher than that of the girls by approximately two-thirds of a year of mathematics learning in Australian schools as estimated in the mid 1990s, although this result may be influenced by the particular schools selected in the analysis and the result is not significant statistically.

The last comparison in SIMS is between boys and girls in all schools. The mean score difference is apparently in favour of boys. However, the effect size (0.05) and t-value (0.72) are very small. Hence, the difference between boys and girls in 1978 is not practically or statistically significant across Australian schools. Out of the four comparisons, only in government schools is the achievement of the girls slightly higher than that of the boys, in the remaining comparisons boys appear to achieve better than girls. These findings appear dissimilar to the findings in FIMS. In the 1964 data set all the differences are in favour of boys. This might indicate that over the 14-year period there was a small shift in achievement level differences between boys and girls. However, the differences are not statistically or practically significant.

*Figure 2*
*Effect size values for differences between boys and girls in Mathematics test score for FIMS, SIMS and TIMS*



## Comparisons of Sex Differences between 1994 Australian Students

The estimated mean score differences between the two sexes in government schools (both Government and Restricted) are in favour of girls (see Table 1 and Figures 1 and 2). However, the effect size and t-values are very small and the mean differences are neither practically nor statistically significant.

The other comparison is between the two sexes in nongovernment schools. Unlike the government school students the difference is seemingly in favour of boys. However, the difference is not statistically or practically significant since the effect size is trivial (0.12) and the t-value is very small, because the design effects are large.

The last comparison in TIMS considers all students (government and nongovernment together). The mean score difference is apparently in favour of girls. However, the effect size (-0.04) and t-value (-0.52) are too small to be considered significant. Consequently the mean difference is neither practically nor statistically significant.

Three of the four different comparisons in TIMS between boys and girls in mathematics achievement reveal that the mathematics achievement levels of the girls is slightly greater than that of the boys. When the findings in TIMS are compared with the findings in FIMS and SIMS, the direction of the difference between the sexes has apparently changed, but the changes are not significant. In Figures 1 and 2 these results for the differences in effect size and in the t-values employed for significance testing are presented graphically in order to summarize the results.

## DISCUSSION AND CONCLUSION

In 1964 the differences are in favour of the boys, while in 1978 the difference is in favour of the girls only in government schools. However, in 1994 the differences, except in nongovernment schools, are in favour of the girls. Nevertheless none of the differences are found to be statistically significant. These nonsignificant differences might seem to suggest that a difference between the sexes on mathematics achievement is starting to emerge in the 1990s in favour of girls and in contrast to the findings of Keeves (1972), Carss (1980), Moss (1982), Leder and Forgasz (1992). These researchers would appear to argue that a difference between the sexes in mathematics achievement in Australian schools starts to emerge at the junior secondary school stage in favour of boys. The apparent change in direction in the 1990s could then be argued to have been a result of the implementation of different government policies to increase the participation and the mathematics achievement level of girls by the State and Federal Governments. Alternatively, it could be argued that the level of performance of the boys has declined more than that of the girls over time, as can be seen from Table 1 there is a general decline in achievement of all Australian students over the 30-year period.The possibility of a significant decline in the level of achievement in mathematics of boys, rather than a noticeable gain in the achievement of girls must be of some concern for Australian education.

Until, there is clear evidence of statistically significant differences between the sexes in mathematics achievement and of sufficient magnitude to warrant resources being diverted to addressing the problem, there would appear to be more pressing problems to be considered in Australian education than the nature and origin of differences between the sexes in mathematics achievement at the lower secondary school stage. Moreover, such programs as have been undertaken in Victoria of establishing separate classes for boys and girls in mathematics at the lower secondary school stage would seem to be a misdirection of effort and resources until it can be shown that a substantial problem exists.

This comment should not be taken to divert attention from the problem of the lower participation of girls in advanced mathematics classes at senior secondary school level.

The failure to detect significant differences between the sexes in mathematics achievement in this study which uses the largest and most carefully designed samples available in Australia over a period of more than 30 years has several important implications for educational research in Australia. There is a need to :

employ large well designed random samples and to attain high response rates in order to

- test relationship that are widely assumed to exist;
- use sound statistical procedures to test for the statistical significance of differences;
- report and discuss the magnitudes of effect sizes and the pattern of results, as well as the use of procedures of inferential statistics;
- monitor in a systematic way changes in mathematics achievement over time, as well as changes in the performance of clearly identified subgroups of students; and
- maintain the view that all students can learn mathematics successfully irrespective of sex, race, ethnic background and social class, at all stages of schooling and in particular at the lower and middle secondary school levels.

## REFERENCES

Adams, R. J. & Khoo, S.T. (1993). *Quest- The Interactive Test Analysis System.* Hawthorn, Victoria: ACER.

Baker, D. P. & Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education, 66*(2), 91-103.

Barnes, M & Horne, M. (1996). Gender and mathematics. In P Sullivan, K Owens & B Atweh, (eds) *Research in Mathematics Education in Australasia 1992- 1995.* University of Western Sydney, Macarthur: Mathematics Education Research Group of Australasia. pp. 51-87.

Brick, J. M., Broene, P., James, P. & Severynse, J. (1997). *A User's Guide to WesVarPC.* (Version 2.11). Boulevard: Westat, Inc.

Byrne, B. M. (1989). *A Primer of LISREL Basic Applications and Programming for Confirmatory Factor Analytic Models.* New York: Springer-Verlag.

Bryk, A. S., Raudenbush, S. W. and Congdon, R. (1992). *Hierarchical Linear Models.* Newbury Park, CA: Sage.

Carss, M. C. 1980. Girls, mathematics and language: Some observations from classrooms. In Adelaide College of the Arts and Education (eds.).*Improving maths for girls: Report of a conference held at Raywood inservice centre, 30 May-1 June, 1980.*

Cohen, J. (1992). A power primer. *Psychological Bulletin,* 112 (1), 155-159.

Finn, J. D. (1997). Hypothesis testing. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (2nd ed.),Oxford: Pergamon, pp. 556-561.

Hambleton, R. K.& Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement, 14 (2),* 75-96.

Jöreskog, K. L. & Sörbom, D. (1992). *LISREL8 and PRELIS2 for DOS, DOS Extender and Windows.* Chicago: Scientific Software Incorporated.

Keeves, J. P. (1968). V*ariation in mathematics education in Australia*: Hawthorn, Victoria: ACER.

Keeves, J. P. (1973). Differences between the sexes in mathematics and science courses. *International Review of Education, 19,* 47-63.

Leder, G. C. (1992). Mathematics and gender: Changing perspectives. In D A Grouws (ed.). *Handbook of Research on Mathematics Teaching and Learning, New* York: Macmillan, pp. 597-622.

Leder, G. & Forgasz, H. (1992). Gender: A critical variable in mathematics education. In B Atweh & J Watson (eds), *Research in Mathematics Education in Australasia 1988-1991,* Brisbane: MERGA, pp. 67-95.

Lokan, J., Ford, P. & Greenwood, L. (1996). *Maths and Science on the Line: Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study.* Melbourne: ACER.

Megarry, J. (1984). Sex, gender and education. In S Acker (ed.). *Women and Education: World Yearbook of Education,* London:Kogan

Moss, J. D. (1982). *Towards equality: progress by girls in mathematics in Australian secondary schools.* Hawthorn, Victoria: ACER.

Rosier, M. J. (1980). *Changes in Secondary School Mathematics in Australia, 1964-1978.* Hawthorn, Victoria: ACER.

Sontag, L. M. (1984). Vertical equating methods: A comparative study of their efficacy. *DAI, 45-03B*, p. 1000.

Spearritt, D. (1997). Factor Analysis. In J P Keeves. (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (2nd ed.), Oxford: Pergamon, pp. 528-539.

Tilahun Mengesha Afrassa. (1998). Mathematics Achievement at the Lower Secondary School Stage in Australia and Ethiopia: A Comparative Study of Standards of Achievement and Student Level Factors Influencing Achievement. Unpublished PhD thesis, The Flinders University of South Australia, Adelaide, Australia.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement.* Chicago: Mesa Press.

## Acknowledgment

The first author was sponsored by The Flinders University of South Australia Overseas Postgraduate Research Scholarship and the Flinders University Research Scholarship.