Gender differences in educational achievement to age 25

Sheree J. Gibb
David M. Fergusson
L. John Horwood

Christchurch Health and Development Study, Department of Psychological Medicine, University of Otago, Christchurch, New Zealand

Gender differences in educational achievement were examined in a cohort of 1265 individuals studied from birth to age 25. There was a small but pervasive tendency for females to score better than males on standardised tests and to achieve more school and post-school qualifications. The differences could not be explained by differences in cognitive ability as males and females had similar IQ scores. Teacher ratings of classroom behaviour revealed that males were more prone to inattentive, restless and distractible behaviours and aggressive, anti-social and oppositional behaviours than females. When the associations between gender and measures of educational achievement were adjusted for teacher ratings of classroom behaviour the gender differences were reduced substantially. These results suggest that one approach to reducing gender differences in educational achievement lies in improving classroom behaviour.

Over the last decade there has been evidence of a growing gender gap in educational achievement in a number of developed countries (for example, see Fergusson & Horwood, 1997; Hillman & Rothman, 2003; Praat, 1999; Thiessen & Nickerson, 1999; Weaver-Hightower, 2003). Educational statistics have indicated that females are outperforming males at all levels of the school system, attaining more school and post-school qualifications, and attending university in higher numbers (Alton-Lee & Praat 2001; House of Representatives Standing Committee on Education and Training, 2002; Mullis et al., 2003; Office for Standards in Education, 2003). Although males have traditionally outperformed females in mathematics and science, this advantage appears to be disappearing (Benbow & Stanley, 1980; Hyde, Fennema & Lamon, 1990; Hyde & Linn, 2006; Spelke, 2005). These findings have caused widespread concern about male educational achievement and have led to considerable speculation and discussion about the origins of gender differences in education. The literature relating to gender differences in educational achievement is vast and complex, and a large number of explanations have been offered but, for the most part, these explanations tend to cluster around three themes: biological factors; gender theory; and school factors.

Biological factors

A number of explanations attribute gender differences in educational achievement to biological differences between males and females. These explanations propose that gender differences in behaviour, skills and cognitive abilities are determined by biological factors such as brain organisation, hormones and genetics, and that these biologically determined differences in behaviour and abilities are responsible for gender differences in educational achievement. For example, Kimura and Hampson (1994) reported that fluctuations in testosterone in males and oestrogen in females were correlated with performance on a range of tests of cognitive ability. A number of studies have reported gender differences in brain structure and function (for reviews, see Cahill, 2006; Halpern, 1997) and in some cases this has been interpreted as evidence that gender differences in educational achievement are biologically determined (Biddulph, 1997; Gurian, 2001; Sax, 2005).

Gender theory

According to gender theory, males and females enter the educational system with different sets of behaviours, attitudes and values. These gendered behaviours, attitudes and values are the result of childhood socialisation in line with the cultural norms of masculinity and femininity (Biddulph, 1997; Epstein et al., 1998; Weaver-Hightower, 2003). It is proposed that, in educational settings, male behaviour, values and attitudes interfere with males' educational achievement. Explanations of this type frequently make reference to a 'laddish' or 'macho' male culture that comprises a constellation of behaviour that includes disruptive behaviour, anti-school attitudes and an interest in traditionally masculine subjects and pastimes (Francis, 1999). For example, Warrington, Younger and Williams (2000) found that boys were more likely than girls to be ridiculed by their peers for working hard at school, and frequently resorted to 'laddish' behaviour such as challenging authority, drawing attention to themselves and pretending not to care about schoolwork in order to gain acceptance from their peer group. Within the gender theory perspective, there are a range of complex and competing discourses regarding the interface between gender, education and society (for reviews, see Epstein et al., 1998; Weaver-Hightower, 2003). Discourses regarding boys' educational achievement tend to focus on the ways in which masculinities are constructed, sustained and reinforced in schools and in wider society. A number of themes can be extracted from these discourses, including

- the existence of multiple masculinities that constantly struggle and compete for dominance (hegemony) (Epstein, 1998; Weaver-Hightower, 2003)
- an emphasis of the importance of both micro-level factors within the school (playground interactions, curriculum materials) and macro-level factors in the wider society (economic changes, changing gender roles) in the construction and reinforcement of masculinities (Warrington et al., 2000; Weaver-Hightower, 2003)
- a recognition that hegemonic masculinity varies between social and ethnic groups and thus social and ethnic factors should not be ignored in discussions of masculinity (Gilbert & Gilbert, 2001; Jackson, 1998; Sewell, 1998).

School factors

A perspective focused on school factors attributes male educational underachievement to schools adopting learning and assessment procedures that are better suited to females than to males. These arguments assume that males and females possess different sets of behaviours, attitudes and learning styles and thus require different school and teaching practices to succeed (for review, see Murphy & Elwood, 1998). In some cases, these explanations claim that teaching and schooling has become 'feminised' and schools are no longer adequately addressing boys' educational needs. In a summary of this perspective, Delamont (1999) outlined the commonly cited ways in which schooling is considered to be feminised, including: school and classroom regimes that favour females; a lack of male teachers to act as academic role models for boys; a lack of toughness in discipline; a rejection of competition; and a bias towards feminism in curriculum materials.

While there have been a large number of explanations of the origins of gender differences in educational achievement, few studies have examined the extent to which these differences are mediated by biological, sociocultural or school factors. A theme that permeates all explanations is that gender differences in educational achievement are largely a reflection of gender differences in classroom behaviour. This explanation is testable since it is possible to examine the extent to which gender differences in classroom behaviour explain gender differences in educational achievement. This approach was used by Fergusson and Horwood (1997) who examined the links between gender and educational achievement in a New Zealand birth cohort studied to the age of 18. That analysis showed the presence of small but pervasive differences in educational achievement including performance on standardised tests and achievement in school leaving examinations. These differences were explained in all cases by gender differences in teacher-reported classroom behaviour. Specifically, boys were described as being more prone to inattentive, distractible and restless behaviour in the classroom context and controlling for these tendencies virtually eliminated any association between gender and educational achievement.

A further theme in this literature has centred around the claim that, rather than considering overall gender differences in educational achievement, we should consider 'which boys' and 'which girls' are underachieving (Epstein et al., 1998; Kleinfeld, 1998; Praat, 1999; Tinklin et al., 2001). The implication of this claim is that social factors such as social class or ethnicity may modify the relationships between gender and educational achievement. This suggests that there is a need to test data for the presence of interactions between gender and social factors such as ethnicity and socioeconomic status, and to examine the nature of gender differences in educational achievement across different social groups.

The present study reports an extension of the study reported by Fergusson and Horwood (1997). It attempts to deal with the issues described above by using data gathered over the course of a 25-year longitudinal study to examine three issues:

- the size of the gender difference in educational achievement during the period from age 8 to age 25 years.

- the role of classroom behaviour as a mediating factor in the relationship between gender and educational achievement
- the extent to which relationships between gender and educational achievement are modified by social factors

Methods

The data used in this study were gathered as part of the Christchurch Health and Development Study (CHDS), a longitudinal study of a birth cohort of 1265 individuals born in Christchurch, New Zealand, in 1977. The cohort has been followed up at birth, four months, one year, and then at yearly intervals to age 16, and again at ages 18, 21 and 25. Data were gathered using various methods including: semi-structured interviews with participants and parents; teacher assessments; and standardised testing. The methodology and findings of the CHDS have been reviewed elsewhere (see Fergusson & Horwood, 2001; Fergusson, Horwood, Shannon & Lawton, 1989).

Outcome measures

Students in New Zealand attend primary school for eight years (Year 1–Year 8) and high school for a maximum of five years (Year 9–Year 13). The minimum school leaving age is 16. Most students turn 16 in Year 11 but the majority of students remain at school until at least the end of Year 12. During the time in which the CHDS cohort attended high school, the qualifications framework included:

- School Certificate examinations. At the end of Year 11, students were eligible (but not required) to sit School Certificate examinations. Most students sat School Certificate examinations in four to six subjects. A grade of 'C' or higher was required to pass a subject.
- Sixth Form Certificate. During Year 12, students who studied and passed an approved course (usually in five or six subjects) were awarded the Sixth Form Certificate.
- University bursary. In Year 13, students intending to attend university could sit University Bursary examinations. Those who attained a suitable passing grade were eligible for entrance to university.

Attainment of educational qualifications Data gathered during the 18-, 21- and 25-year interviews were used to determine attainment of the following school and post-school educational qualifications:

- Left school without qualifications. Participants who had left school by age 18 and had failed to gain any formal secondary school qualifications (School Certificate, Sixth Form Certificate or University Bursary) were classified as having left school without qualifications.
- Attained secondary school qualifications by age 25. Participants who had gained at least one formal secondary school qualification by age 25, either while they were at school or subsequently as an adult student, were classified as having gained secondary school qualifications.

- Attended university. At ages 21 and 25, participants were asked about their history of enrolment in tertiary education. Those who reported having ever been enrolled in a university course were classified as having attended university.
- Attained degree. At age 25, participants were asked about the qualifications they had attained since leaving school. Those who reported that they had completed a bachelor's or higher-level degree qualification from a university or equivalent tertiary institution were classified as having attained a degree.
- Overall highest educational qualification. Based on data collected in the 18-, 21- and 25-year interviews, participants' highest educational qualifications at age 25 were classified on a seven point scale where '1' denoted no qualifications and '7' denoted a university degree. This measure had a mean of 4.2 and a standard deviation of 2.2.

Standardised tests In addition to the attainment of educational qualifications, cohort members were also assessed on a series of standardised tests of academic achievement. For ease of comparison, all standardised test scores were scaled to a mean of 100 and a standard deviation of 10. The following standardised test scores were used:

- Word recognition. At ages 8, 10, 12 and 18 participants completed the New Zealand revision of the Burt Word Reading Test (Gilmore et al., 1981). Participants' total raw scores were given by the number of words correctly identified out of a possible total of 110. The KR20 reliabilities for these tests were all in excess of 0.97.
- Reading comprehension. At ages 10 and 12, participants' reading comprehension was assessed using tests based around the Progressive Achievement Test (PAT) of Reading Comprehension (Elley & Reid, 1969). For the 10-year measure test items spanned an 8-year-old's to 12-year-old's reading level. For the 12-year measure, test items spanned an 8-year-old's to 14-year-old's reading level. The reliability of both measures, assessed using coefficient alpha, was 0.83.
- Mathematical reasoning. At age 11 participants' mathematical reasoning skills were assessed using a test based around the Progressive Achievement Test (PAT) of Mathematical Reasoning (Reid & Hughes, 1974). Test items spanned difficulty levels from the 8- to 13-year age groups. The reliability of this measure was 0.87.

Cognitive ability

At ages 8 and 9 participants were assessed with the Revised Wechsler Intelligence Scale for Children (WISC-R) (Wechsler, 1974). Scores for Verbal IQ, Performance IQ and Total IQ were calculated using the methods described in the test manual. The reliabilities for these measures, assessed using split half methods, ranged from 0.87 to 0.95.

Classroom behaviour

At ages 6, 8, 10 and 12 years teachers were asked to provide reports of participants' classroom behaviour based on a behaviour inventory that combined items from the

Rutter, Tizard and Whitmore (1970) and Connors (1969) teacher questionnaires. Factor analysis of the item-level report data (Fergusson, Horwood & Lloyd, 1991) showed that it was possible to select items from these questionnaires that formed uni-dimensional scales reflecting two correlated dimensions of child disruptive behaviours: conduct problems—the extent to which children display antisocial, aggressive and oppositional behaviours; and attention problems—the extent to which children display inattentive, distractible or hyperactive behaviours in the classroom. At each age, conduct problem and attention problem scores were computed from the sum of the scale items for each factor. For ease of comparison these scores were scaled to a mean of 100 and a standard deviation of 10. The reliabilities of these scales, assessed using coefficient alpha, ranged from 0.88 to 0.95.

Demographic measures

Ethnicity At age 21 and age 25 participants were asked to identify their ethnicity. Participants who stated that they were of New Zealand Maori ethnicity at either age 21 or 25 were classified as Maori for the purposes of this study. Overall, 12.2 per cent of the total sample were classified as Maori.

Socioeconomic status The socioeconomic status of the family at birth was classified using the Elley and Irving (1976) scale of socioeconomic status for New Zealand. This scale classified families into 6 levels of socioeconomic status based on paternal occupation. For the purposes of this study, this scale was condensed to a three-level scale where '1' denoted managerial and professional occupations, '2' denoted clerical and technical occupations and '3' denoted semi-skilled and unskilled occupations. This measure had a mean of 2.1 and a standard deviation of 0.7.

Statistical methods

All statistical analyses were performed in SAS version 9.1 (SAS Institute 2003).

Differences in mean scores were tested for statistical significance using a *t*-test for independent samples (Tables 1, 2 and 3 below). Differences in percentage scores were tested for statistical significance using a chi-square test for independence (Table 1). In all cases effect size was measured by Cohen's *d* (Cohen, 1977).

Adjustment of mean scores for covariates was achieved by fitting multiple linear regression models in which the outcome measure was modelled as a function of covariates including gender, IQ and classroom behaviour. In general the model fitted was:

$$Yi = B0 + B1Gi + \sum BjZij + Ui$$

where Yi represented the individual's score on the outcome measure, Gi represented the gender of the ith individual, Zij were the relevant covariates and Ui was the disturbance term. In this model B1 provides an estimate of the net effect of gender on the outcome variable when other factors were taken into account. Adjusted means were computed using the least squares adjustment method.

Adjustment of percentage scores for covariates was achieved by fitting multiple logistic regression models in which the log odds of the outcome measure was

modelled as a function of covariates including gender, IQ and classroom behaviour. In general, the model used was

$$\text{Log}it[\Pr(Yi = 1)] = B0 + B1Gi + \sum BjZij$$

where log*i*t [Pr(Y*i* = 1)] is the log odds rating of achievement of the *i*th outcome and the other variables have similar interpretations to those in the multiple linear regression model described above. Adjusted percentages were computed using the method described by Lee (1981).

To simplify the covariate adjustment analysis and avoid issues of multi-collinearity resulting from the availability of multiple measures of child cognitive ability and behaviour, the following approach was adopted. First, a single measure of child cognitive ability was used based on the average of the child's total IQ scores at ages 8 and 9 years. Second, for each analysis, teacher ratings of child classroom behaviour at different ages were combined using a propensity score approach to provide a single score for attention problems and a single score for conduct problems. Specifically, for each outcome, behaviour scores were constructed as a weighted sum of the attention or conduct problem scores measured at ages 6, 8, 10 and 12, up to and including the age at which the outcome measure was assessed (or up to age 12 if the outcome measure was assessed after age 12). Weightings were given by the regression coefficients from a multiple regression model where the outcome measure was modelled as a function of the attention or conduct problem scores.

Finally, the regression models above were extended to incorporate measures of ethnicity and family socioeconomic status, and tests for gender interactions with these factors.

Sample size and sample bias

Sample sizes in this study varied from as low as 775 to as high as 1110. There are two reasons for this fluctuating sample size. Firstly, over the course of the study there has been a gradual loss of participants due to emigration from New Zealand, participant refusal and death. At age 25 the sample consisted of 1003 participants, which represented 79.3 per cent of the original sample. Secondly, for practical reasons it was possible to conduct standardised testing only on those participants resident in the Canterbury area at the time of interview. Therefore sample sizes for standardised test measures are lower than those for other measures. From ages 8 to 12 Canterbury residents represented approximately 80 per cent of the total sample. These variations in sample size raise issues about the extent to which the results may be affected by sample selection bias. To examine this issue, missing data were imputed using the 'PROC MI' procedure in SAS 9.1 (SAS Institute, 2003), under the assumption that the data were missing at random, and the data were re-analysed. The results of the analyses using the imputed data were essentially the same as those using the observed data, suggesting that the results of the current study were not substantially affected by selection bias. Unadjusted effect sizes (Cohen's *d*) for the observed data ranged from 0.13 to 0.63 with a median of 0.23. Unadjusted effect sizes for the imputed data ranged from 0.13 to 0.65 with a

median of 0.25. Due to the negligible differences between the results of the two analyses, this paper reports the analysis of the observed data only.

Results

Gender differences in educational achievement (ages 8–25)

Table 1 compares males and females on standardised tests of school achievement and measures of the attainment of school and post-school qualifications. For ease of comparison the standardised tests have all been scaled to a mean of 100 and a standard deviation of 10. Each comparison was tested for statistical significance using a *t*-test for independent samples (for continuous measures) or a chi-square test (for dichotomous measures), and the size of the effect was measured using Cohen's *d*. The table shows a pervasive tendency for males to score more poorly on standardised tests and to attain fewer qualifications than females. In almost all cases gender differences were statistically significant ($p<0.05$), with the only exception being the PAT Mathematical Reasoning test at age 11, where the difference was marginally significant ($p<0.10$). Effect sizes were typically small, with values for Cohen's *d* ranging from 0.13 to 0.31 with a median value of 0.17.

Gender differences in cognitive ability (ages 8–9)

One factor that could explain the tendency for males to perform more poorly on measures of educational achievement is gender-related differences in cognitive ability. Table 2 compares mean IQ scores on the WISC-R for males and females

Table 1 Gender differences in standardised test performance and educational attainment to age 25

Measure	*N*	*Males*	*Females*	*p*	*d*
Standardised tests[a]					
Burt Word Reading score age 8	881	98.4	101.6	<0.0001	0.31
Burt Word Reading score age 10	846	99.0	101.0	<0.005	0.21
Burt Word Reading score age 12	804	99.0	100.9	<0.01	0.19
Burt Word Reading score age 18	1008	99.1	100.9	<0.01	0.17
PAT Reading Comprehension age 10	847	98.7	101.3	<0.0005	0.25
PAT Mathematical Reasoning age 11	831	99.4	100.6	<0.10	0.13
PAT Reading Comprehension age 12	804	99.2	100.8	<0.05	0.16
Educational attainment					
Left school with no qualifications	1025	22.2%	16.4%	<0.05	0.15
Attained secondary school qualifications by age 25	1053	78.7%	84.8%	<0.05	0.16
Attended university	1001	37.8%	44.6%	<0.05	0.14
Attained degree	1003	22.3%	28.5%	<0.05	0.14
Overall highest educational qualification[b]	1001	4.0	4.4	<0.001	0.21

[a] For ease of comparison, standardised test scores have been scaled to a mean of 100 and a standard deviation of 10.

[b] Measured on a seven-point scale where 1 = no formal qualifications and 7 = university degree.

Table 2 Comparisons of male and female cognitive ability scores (WISC-R)

	N	*Males*	*Females*	*p*
Age 8				
WISC-R Verbal IQ	881	101.6	100.5	>0.30
WISC-R Performance IQ	881	102.8	102.8	>0.95
WISC-R Total IQ	881	102.3	101.5	>0.45
Age 9				
WISC-R Verbal IQ	811	101.8	100.2	>0.15
WISC-R Performance IQ	811	107.5	106.5	>0.35
WISC-R Total IQ	811	104.8	103.4	>0.20

at ages 8 and 9. Each comparison was tested for statistical significance using a *t*-test for independent samples. There were no significant differences between males and females on any of the IQ measures. However, males did score slightly higher than females on almost all of the IQ measures, with the only exception being Performance IQ at age 8 where male and female scores were equal.

Gender differences in classroom behaviour

Table 3 shows mean scores for males and females on two teacher-rated measures of classroom behaviour measured at ages 6, 8, 10 and 12. At each age, two measures were available: the extent to which participants engaged in distractible, restless and inattentive behaviour; and the extent to which they engaged in aggressive, anti-social or oppositional behaviour. For ease of comparison, all measures have been scaled to a mean of 100 and a standard deviation of 10. Each comparison was tested for statistical significance using a *t*-test for independent samples. The size of the effect was measured by Cohen's *d*. Table 3 shows that, at all ages, males were described by their teachers as being more likely to engage in both inattentive, rest-

Table 3 Gender differences in teacher ratings of classroom attention and conduct problems

Measure[a]	*N*	*Males*	*Females*	*p*	*d*
6 years					
Attention problems	1110	102.9	97.1	<0.0001	0.57
Conduct problems	1110	101.7	98.3	<0.0001	0.34
8 years					
Attention problems	1083	102.8	97.1	<0.0001	0.57
Conduct problems	1083	102.1	97.8	<0.0001	0.44
10 years					
Attention problems	1056	103.1	96.9	<0.0001	0.63
Conduct problems	1056	102.3	97.6	<0.0001	0.47
12 years					
Attention problems	1007	102.9	97.1	<0.0001	0.58
Conduct problems	1006	101.8	98.2	<0.0001	0.36

[a] For ease of comparison, all ratings of attention and conduct problems have been scaled to a mean of 100 and a standard deviation of 10.

less and distractible behaviour, and aggressive, antisocial and oppositional behaviour in the classroom. In all cases the difference between male and female scores was statistically significant. Effect sizes (Cohen's *d*) for inattentive, restless and distractible behaviours were moderate and ranged from 0.57 to 0.63, while effect sizes for aggressive, antisocial and oppositional behaviour were slightly smaller and ranged from 0.34 to 0.47.

Adjustment of gender differences in educational achievement for classroom behaviour

Taken together, the results from Tables 1 to 3 suggest that differences in educational achievement to age 25 may be explained by gender-related differences in classroom behaviour, with higher rates of disruptive and antisocial classroom behaviour in males leading to poorer educational outcomes. To examine this possibility, regression models were fitted to the data in Tables 1 to 3. Continuous variables were analysed with linear regression models and dichotomous measures were analysed with logistic regression models (see Methods, above). All measures were adjusted for childhood IQ, teacher ratings of disruptive, restless and inattentive classroom behaviour from age 6 to age 12, and teacher ratings of aggressive, anti-social and oppositional classroom behaviours from age 6 to age 12 (or to the oldest appropriate age for outcome measures prior to age 12). The results led to the conclusions described in the next sections.

Standardised tests The analysis showed that in all cases controlling for IQ and classroom behaviour reduced the differences between means. After adjustment, values of Cohen's *d* ranged from 0.06 to 0.28 with a median value of 0.13. A number of comparisons still remained statistically significant, even after adjustment for IQ and classroom behaviour. These comparisons were all measures of educational achievement prior to age 12. Inspection of the covariates showed that, in all cases, IQ and measures of distractible, restless and inattentive classroom behaviour were significant predictors of educational achievement. Measures of aggressive, antisocial and oppositional classroom behaviour were also a significant predictor for three of the measures.

Attainment of school and post-school qualifications In all cases, controlling for IQ and measures of classroom behaviour up to age 12 again reduced the differences between males and females. After adjustment for IQ and classroom behaviour, values of Cohen's *d* ranged from 0.00 to 0.11 with a median value of 0.04. Controlling for IQ and classroom behaviour reduced the gender differences in achievement to the point where these differences were no longer statistically significant but the comparison for overall highest educational qualification still remained marginally significant ($p<0.08$). Inspection of the covariates indicated that attention problems were a significant predictor for university attendance, university degree attainment and overall highest educational qualification, while conduct problems were a significant predictor for attainment of secondary school qualifications. IQ was a significant predictor for all of the outcome measures.

Table 4 Gender differences in educational achievement, adjusted for IQ and teacher ratings of classroom behaviour

Measure	*N*	*Males*	*Females*	*p*	*d*	*Significant covariates*[a]
Standardised tests[b]						
Burt Word Reading score age 8	870	98.6	101.4	<0.0001	0.28	1, 2
Burt Word Reading score age 10	824	99.2	100.9	<0.005	0.17	1, 2, 3
Burt Word Reading score age 12	775	99.6	100.3	>0.20	0.07	1, 2, 3
Burt Word Reading score age 18	908	99.7	100.2	>0.30	0.06	1, 2
PAT Reading Comprehension age 10	825	99.0	101.0	<0.0001	0.21	1, 2
PAT Mathematical Reasoning age 11	803	99.6	100.6	<0.05	0.10	1, 2, 3
PAT Reading Comprehension age 12	777	99.9	100.0	>0.80	0.13	1, 2
Educational attainment						
Left school with no qualifications	917	17.7%	19.4%	>0.45	0.04	1, 3
Attained secondary school qualificationsby age 25	939	82.8%	82.1%	>0.70	0.02	1, 3
Attended university	898	39.4%	42.2%	>0.35	0.06	1, 2
Attained degree	899	25.4%	25.6%	>0.90	0.00	1, 2
Overall highest educational qualification[c]	898	4.1	4.3	<0.08	0.11	1, 2

[a] Significant covariates: 1 = Total IQ average 8/9 years; 2 = Attention problems at ages 6, 8, 10 and 12, measured up to and including the age of the outcome measure; 3 = Conduct problems at ages 6, 8, 10 and 12, measured up to and including the age of the outcome measure.

[b] For ease of comparison, all ratings of attentional and conduct problems have been scaled to a mean of 100 and a standard deviation of 10.

[c] Measured on a seven-point scale where 1 = no formal qualifications and 7 = university degree

Interactions with ethnicity and socioeconomic status One issue not dealt with by the preceding analyses is the extent to which educational achievement is determined by the interaction between gender and other variables such as ethnicity and socioeconomic status. To examine this possibility, each outcome measure was tested for the effects of gender, ethnicity, socioeconomic status and the interactions between gender and ethnicity and gender and socioeconomic status. While there were significant effects for ethnicity for some outcome measures (PAT Mathematical Reasoning age 11, leaving school without qualifications, secondary qualifications age 25, attended university, attained university degree, and overall highest educational qualification), in no case was there a significant interaction between gender and ethnicity (all $p>0.05$). Similarly, while there were significant effects for socioeconomic status on all outcome measures (all $p<0.01$), there were no significant interactions between gender and socioeconomic status (all $p>0.20$).

Discussion

From age 8 to age 25, this study found that there was a pervasive tendency for females to outperform males on measures of educational achievement. This trend was consistent across a variety measures of educational achievement including standardised tests, attainment of high-school qualifications, university attendance and university degree attainment. While gender differences in educational achievement

were pervasive, the effect sizes for these differences were small, with Cohen's *d* ranging from 0.13 to 0.31 with a median value of 0.17 but the small effect sizes translate into considerable differences on a population scale. For example, in New Zealand in 1996 (the year that many cohort members would have begun university study) there were 56 429 females enrolled at university and 49 261 males, a difference of 7168 individuals. Similarly, in 1995 (the year in which many cohort members would have left high school) 5457 males left high school without any formal qualifications, compared to only 4287 females. It is unlikely that these effects are specific to the cohort or time period studied for two reasons. Firstly, the finding is consistent with numerous other reports of male educational underachievement using different samples in a number of developed countries (Alton-Lee & Praat, 2001; Thiessen & Nickerson, 1999; Tinklin et al., 2001; Weaver-Hightower, 2003; Younger & Warrington, 2005). Secondly, the trend of male educational underachievement has been evident since at least the mid-1990s and continues to be reported in recent education statistics (Australian Bureau of Statistics, 1998; Freeman, 2004; Tinklin et al., 2001). Thus, the finding from the current study appears to be a reflection of an international trend of male educational underachievement that has been evident for at least the last decade.

The results of this study indicated that a male disadvantage in educational achievement was evident by age 8, and this disadvantage continued through to age 25. This finding raises the question of the origin of these gender differences in educational achievement. It can be suggested that gender differences in educational achievement arise from gender differences in cognitive ability, with lower cognitive ability in males leading to poorer educational achievement throughout the lifespan. The current study did not provide any support for this suggestion. There were no significant differences in IQ measured at ages 8 and 9, and there was a slight tendency for males to score higher than females on most of these tests. While IQ was a significant covariate for all of the regression models, adjusting gender differences in educational achievement for IQ had the effect of increasing, rather than decreasing, the gender gap. These results suggest that gender differences in educational achievement are not the result of gender differences in cognitive ability.

An alternative explanation is that gender differences in educational achievement are the result of gender differences in classroom behaviour. The findings of the current study support this explanation. From age 6 to age 12, males were described by their teachers as displaying significantly higher levels than females of distractible, restless, inattentive behaviour and aggressive, antisocial, oppositional behaviour. Adjustment of measures of educational achievement for gender differences in classroom behaviour eliminated many of the gender differences in educational achievement. The exceptions were measures of educational achievement prior to age 12 where adjustment for classroom behaviour reduced the gender gap but did not eliminate it. Taken together, these findings suggest that gender differences in educational achievement to age 25 were largely explained by gender differences in classroom behaviour during middle childhood. In support of this conclusion, there is a considerable body of literature reporting higher rates of disruptive, aggressive and inattentive behaviour in males inside and outside of the

classroom (Eme & Kavanaugh, 1995; for review, see Loeber et al., 2000; Warrington et al., 200) with some of this literature linking inattentive behaviour to poor educational achievement (Alexander et al., 1993; Fergusson & Horwood, 1995; Hinshaw, 1992; Tremblay et al., 1992). The link between male classroom behaviour and subsequent educational achievement may operate via one of two processes. Firstly, male disruptive behaviour during middle childhood may create a gender gap in educational achievement during middle childhood, with this gap being sustained through to young adulthood. Secondly, disruptive behaviour in middle childhood may be predictive of problematic behaviour in young adulthood that interferes with educational achievement at that time.

While the results highlight the role of disruptive classroom behaviour in male educational underachievement, they provide no information about the causes of gender differences in classroom behaviour. Thus, they cannot be seen as providing support for gender-role explanations of male educational underachievement as these explanations appeal to socially determined patterns of male socialisation to explain gender differences in behaviour. Evidence points to a range of factors contributing to the development of disruptive behaviour, including genetic, family and social factors (Eme & Kavanaugh, 1995; Leadbeater et al., 1995; Nix et al., 1999).

Some authors have proposed that social factors such as socioeconomic status and ethnicity may modify the relationship between gender and achievement (Alton-Lee & Praat, 2001; Coley, 2001; Dekkers et al., 2000). The current study provided no evidence to support this claim. There were no significant interactions between gender and socioeconomic status or gender and ethnicity for any of the outcome variables. Instead, the effects of gender and ethnicity were additive, with those who are male and of Maori ethnicity being the most disadvantaged, and those who are female and not of Maori ethnicity being the least disadvantaged. While educational achievement was significantly associated with ethnicity and socioeconomic status, the additive nature of these factors means that it is important to consider the full range of factors that contribute to educational disadvantage.

The findings from this study suggest that one of the ways in which male educational achievement could be raised is by improving classroom behaviour. While improving classroom behaviour is sometimes recommended as a strategy for tackling male underachievement (for example, see Education Review Office, 1999; Fergusson & Horwood, 1997; House of Representatives Standing Committee on Education and Training, 2002; Office for Standards in Education, 2003), these recommendations often fail to outline specific strategies and programs that may be useful. There is, however, a small but growing literature on school-based programs that are effective in reducing disruptive behaviour (for review, see Stage & Quiroz, 1997). In particular, group reward contingencies in which reinforcement depends on the collective behaviour of the group have been shown to be highly effective and are less time-consuming than individual reward contingencies. For example, the Good Behaviour Game (for reviews, see Tankersley, 1995; Tingstrom et al., 2006) is a group-oriented classroom reward system in which students are split into groups and the number of instances of specified negative target behaviour (for

example, calling out, talking, out-of-seat behaviour) is recorded for each group. All members of groups with behaviour counts under a specified acceptable limit receive a reward (for example, extra recess time or class privileges). The intervention has been shown to be highly effective at reducing problem behaviour: in some cases, reducing these behaviours by over 90 per cent (Harris & Sherman, 1973; Medland & Stachnik, 1972). Few studies have examined whether these programs also improve academic achievement. There is considerable scope for experimental research to determine the extent to which programs and strategies that reduce male disruptive classroom behaviour also reduce the gender gap in educational achievement.

A particular strength of this study is that it provides a life course perspective of educational achievement, with achievement measures from middle childhood through to young adulthood. The longitudinal design makes it possible to identify and control for factors that mediate educational achievement.

A limitation is that this study used data from a birth cohort of individuals born at a particular time and place. In particular, the cohort attended school in New Zealand during a period in which there was a very strong focus on tackling female educational underachievement (for example, see New Zealand Council for Educational Research, 1988). This raises issues about the capacity to generalise the findings to individuals born at different times and in different places.

These limitations notwithstanding, the findings of this study show a small but pervasive trend of male educational underachievement from childhood through to young adulthood that is largely explained by higher levels of disruptive and antisocial classroom behaviour in males. Thus, one approach to tackling the current gender gap in educational achievement may be through classroom-based programs that reduce disruptive behaviour and subsequently improve male academic achievement.

Keywords

gender differences	academic achievement	longitudinal studies
behaviour problems	education	academic awards

References

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-grade classroom behaviour: its short- and long-term consequences for school performance. *Child Development, 64*(3), 801–814.

Alton-Lee, A., & Praat, A. (2001). *Explaining and addressing gender differences in the New Zealand compulsory school sector.* Wellington: Ministry of Education.

Australian Bureau of Statistics (1998). *Educational attainment: gender differences in educational achievement.* Canberra: Australian Bureau of Statistics.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: fact or artifact? *Science, 210*(4475), 1262–1264.

Biddulph, S. (1997). *Raising boys.* Sydney: Finch Publishing.

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews: Neuroscience,* 7(6), 477–484.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Coley, R. J. (2001). *Differences in the gender gap: comparisons across racial/ethnic groups in education and work*. Policy information report. Princeton, NJ: Educational Testing Service.

Conners, C. K., (1969). A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry, 126*, 884–888.

Dekkers, H. P. J. M., Bosker, R. J., & Driessen, G W. J. M. (2000). Complex inequalities of educational opportunities: a large-scale longitudinal study on the relation between gender, social class, ethnicity and school success. *Educational Research and Evaluation, 6*(1), 59–82.

Delamont, S. (1999). Gender and the discourse of derision. *Research Papers in Education, 14*(1), 3–21.

Education Review Office (1999). *The achievement of boys*. Wellington: Education Review Office.

Elley, W. B., & Irving, J. C. (1976). Revised socio-economic index for New Zealand. *New Zealand Journal of Educational Studies, 11*, 25–36.

Elley, W. B., & Reid, N. A. (1969). *Progressive achievement texts: Teacher manual, reading comprehension, reading vocabulary*. Wellington: New Zealand Council for Educational Research.

Eme, R. F., & Kavanaugh, L. (1995). Sex differences in conduct disorder. *Journal of Clinical Child Psychology, 24*(4), 406–426.

Epstein, D. (1998). Real boys don't work: 'underachievement', masculinity and the harassment of 'sissies'. In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds), *Failing boys?: Issues in gender and achievement*. Buckingham, UK: Open University Press.

Epstein, D., Elwood, J., Hey, V., & Maw, J. 91998). Schoolboy frictions: feminism and 'failing' boys. In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds), *Failing boys?: Issues in gender and achievement*. Buckingham, UK: Open University Press.

Fergusson, D. M., & Horwood, L. J. (1995). Early disruptive behavior, IQ, and later school achievement and delinquent behavior. *Journal of Abnormal Child Psychology, 23*(2), 183–199.

Fergusson, D. M., & Horwood, L. J. (1997). Gender differences in educational achievement in a New Zealand birth cohort. *New Zealand Journal of Educational Studies, 32*(1) 83–96.

Fergusson, D. M., & Horwood, L. J. (2001). The Christchurch Health and Development Study: Review of findings on child and adolescent mental health. *Australian and New Zealand Journal of Psychiatry, 35*(3), 287–296.

Fergusson, D. M., Horwood, L. J., & Lloyd, M. (1991). Confirmatory factor models of attention deficit and conduct disorder. *Journal of Child Psychology and Psychiatry, 32*(2), 257–274.

Fergusson, D. M., Horwood, L. J., Shannon, F. T., & Lawton, J. M. (1989). The Christchurch Child Development Study: A review of epidemiological findings. *Paediatric and Perinatal Epidemiology, 3*(3), 278–301.

Francis, B. (1999). Lads, lasses and (new) labour: 14–16-year-old students' responses to the 'laddish behaviour and boys' underachievement' debate. *British Journal of Sociology of Education, 20*(3), 355–371.

Freeman, C. E. (2004). *Trends in educational equity of girls and women*. Washington, DC: National Center for Educational Statistics.

Gilbert, P., & Gilbert, R. (2001). Masculinity, inequality and post-school opportunities: disrupting oppositional politics about boys' education. *International Journal of Inclusive Education, 5*(1), 1–13.

Gilmore, A., Croft, C., & Reid, N. A. (1981). *Burt word reading test New Zealand revision: Teachers manual*. Wellington: New Zealand Council for Educational Research.

Gurian, M. (2001). *Boys and girls learn differently! A guide for teachers and parents*. San Francisco: Jossey-Bass.

Halpern, D. F. (1997). Sex differences in intelligence: implications for education. *American Psychologist, 52*(10), 1091-1102.

Harris, V. W., & Sherman, J. A. (1973). Use and analysis of the 'Good Behavior Game' to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis, 6*(3), 405–417.

Hillman, K., & Rothman, S. (2003). *Gender differences in educational and labour market outcomes*. Melbourne: Australian Council for Educational Research.

Hinshaw, S. P. (1992). Externalizing behaviour problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*(1), 127–155.

House of Representatives Standing Committee on Education and Training (2002). *Boys: Getting it right*. Canberra: Commonwealth Government.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin, 107*(2), 139–155.

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science, 314*(5799), 599–600.

Jackson, D. (1998). Breaking out of the binary trap: boys' underachievement, schooling and gender relations. In In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds), *Failing boys?: Issues in gender and achievement*. Buckingham, UK: Open University Press.

Kimura, D., & Hampson, E. (1994). Cognitive pattern in men and women is influenced by fluctuations in sex hormones. *Current Directions in Psychological Science, 3*(2), 57–61.

Kleinfeld, J. (1988). *The myth that schools shortchange girls; Social science in the service of deception*. Washington, DC: Women's Freedom Network.

Leadbeater, B. J., Blatt, S. J., & Quinlan, D. M. (1995). Gender-linked vulnerabilities to depressive symptoms, stress, and problem behaviors in adolescents. *Journal of Research on Adolescence, 5*(1), 1–29.

Lee, J. (1981). Covariance adjustment of rates based on the multiple logistic regression model. *Journal of Chronic Disease, 34*, 415–426.

Loeber, R., Burke, J. D., Lahey, B. B., Winters, A., & Zera, M. (2000). Oppoistional defiant and conduct disorder: A review of the past 10 years, part I. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(12), 1468–1484.

Medland, M. B., & Stachnik, T. J. (1972). Good-Behavior Game: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 5*(1), 45–51.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*. Chestnut Hill, MA: Boston College.

Murphy, P., & Elwood, J. (1998). Gendered learning outside and inside school: influences on achievement. In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds), *Failing boys?: Issues in gender and achievement*. Buckingham, UK: Open University Press.

New Zealand Council for Educational Research (1988). *Equal opportunity in education for girls and women*. General overview. *In Royal commission on social policy, The April report, 3* (part 2). Wellington: Government Printer.

Nix, R. L., Pinderhughes, E. E., Dodge, K. A., Bates, J. E., Pettit, G. S., & McFadyen-Ketchum, S. A. (1999). The relation between mothers' hostile attribution tendencies and children's externalizing behavior problems: the mediating role of mothers' harsh discipline practices. *Child Development, 70*(4), 896–909.

Office for Standards in Education (2003). *Boys' achievement in secondary schools*. London: OFSTED Publications Centre.

Praat, A. (1999). Gender differences in student achievement and in rates of participation in the school sector, 1986–1997: a summary report. *The Research Bulletin, 10*, 1–11.

Reid, N. A., & Hughes, D. C. (1974). *Progressive achievement texts: Teacher manual, mathematics*. Wellington: New Zealand Council fo Educational Research.

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health and behaviour*. London: Longmans.

SAS Institute (2003). *SAS 9.1* [computer software]. Cary: SAS Institute Inc.

Sax, L. (2005). *Why gender matters: What parents and teachers need to know about the emerging science of sex differences*. New York: Doubleday.

Sewell, T. (1998). Loose canons: Exploding the myth of the 'black macho' lad. In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds), *Failing boys?: Issues in gender and achievement*. Buckingham, UK: Open University Press.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60*(9), 950–958.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education. *School Psychology Review, 26*(3),333–368.

Tankersley, M. (1995). A group-oriented contingency management program. *Preventing School Failure, 40*(1), 19–24.

Thiessen, V., & Nickerson, C. (1999). *Canadian gender trends in education and work*. Ottawa: Human Resources Development Canada Applied Research Branch.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The Good Behavior Game: 1969–2002. *Behavior Modification, 30*(2), 225–253.

Tinklin, T., Croxford, L., Ducklin, A., & Frame, B. (2001). *Gender and pupil performance in Scotland's schools*. Edinburgh: Edinburgh University Press.

Tremblay, R. E., Masse, B., Perron, D., Leblanc, M., Schwartzman, A. E., & Ledingham, J. E. (1992). Early disruptive behvior, poor school achievement, delinquent behavior, and delinquent personality: longitudinal analyses. *Journal of Consulting and Clinical Psychology, 60*(1), 64–72.

Warrington, M., Younger, M., & Williams, J. (2000). Student attitudes, image and the gender gap. *British Educational Research Journal, 26*(3), 393–407.

Weaver-Hightower, M. (2003). The 'boy turn' in research on gender and education. *Review of Educational Research, 73*(4), 471–498.

Wechsler, D. (1974). *Wechsler intelligence scale for children—revised*. New York: Psychological Corporation.

Younger, M., & Warrington, M. (2005). *Raising boys' achievement*. London: Department for Education and Skills.

Acknowledgements

This research was funded by grants from the Health Research Council of New Zealand, the National Child Health Research Foundation, the Canterbury Medical Research Foundation and the New Zealand Lottery Grants Board. Sheree Gibb was supported by a University of Otago PhD scholarship.

Authors

Sheree Gibb is a PhD student working on a study of gender differences in income and educational achievement using the data from the Christchurch Health and Development Study, University of Otago.

David Fergusson is principal investigator of the Christchurch Health and Development Study, University of Otago.
Email: dm.fergusson@otago.ac.nz

John Horwood is deputy director of the Christchurch Health and Development Study, University of Otago, and provides bio-statistical support.