Intelligence 38 (2010) 424–432

Contents lists available at ScienceDirect

Intelligence

Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England

Catherine M. Calvin [a], Cres Fernandes [b], Pauline Smith [b], Peter M. Visscher [c], Ian J. Deary [a,*]

[a] Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh 9JZ, UK
[b] GL Assessment, The Chiswick Centre, 414 Chiswick High Road, London, W4 5TF, UK
[c] Queensland Statistical Genetics, Queensland Institute of Medical Research, 300 Herston Road, Herston, Queensland 4006, Australia

ARTICLE INFO

Article history:
Received 21 January 2010
Received in revised form 21 April 2010
Accepted 21 April 2010
Available online 31 May 2010

Keywords:
Sex
Intelligence
Education
Cognitive Abilities Test
Key Stage 2

ABSTRACT

General cognitive ability (*g*) does not explain sex differences in academic test performance by the end of compulsory education. Instead, individual differences in specific reasoning abilities, after removing the effects of *g*, may contribute to the observed gender gaps. Associations between general or specific cognitive abilities, sex, and educational attainments were analysed in a cross-sectional study of 11-year-olds ($M=133.5$ months, $SD=3.5$), at an age before substantive gender-related selection-bias occurred. The 178,599 pupils (89,545 girls and 89,054 boys) attending English state schools represented 93% of the UK's local education authorities. In 2004 each student completed the Cognitive Abilities Test—Third Edition (CAT3), assessing verbal, quantitative, and nonverbal reasoning abilities. These data were linked to each child's attainment scores on national Key Stage 2 tests in English, mathematics and science. A sex difference in *g*, favoring girls, was statistically significant but of negligible effect size (Cohen's $d=.01$). Girls scored 26% of a *SD* higher than boys on a verbal residual factor, and boys scored 28% of a *SD* higher than girls on a quantitative residual factor, with negligible sex differences on a nonverbal residual factor (1% of a SD). In education, 10% more girls than boys achieved UK government targets in English. In mathematics and science, sex differences were less apparent at the government target grade (Level 4), although a 5% greater proportion of boys than girls performed at the highest level in mathematics (Level 5). General cognitive ability (*g*) was strongly related to an educational factor score ($r=.83$) as expected, and did not explain sex differences in academic performance. In general linear models, a verbal residual factor explained up to 29% of girls' higher English attainment, and better quantitative skills among boys explained 50% of their higher attainment in mathematics. Besides the significant contributions of specific cognitive abilities to gender differences in English and mathematics, there remains substantive variance of the educational gender gap left to explain.

1. Introduction

In national cohorts, a general cognitive ability factor (*g*), measured from childhood psychometric test scores, is strongly predictive of academic achievements (Bartels, Rietveld, Van Baal, & Boomsma, 2002; Deary, Strand, Smith, & Fernandes, 2007). The effect size of the association between intelligence and educational attainment is often reported as moderate-large (correlation of around .50; Neisser et al., 1996), and increases as samples become more representative of the general population (Sternberg, Grigorenko, & Bundy, 2001) or advance in school years (Bartels et al., 2002). Recently, in a national cohort of United Kingdom schoolchildren with well-standardised reasoning ability scores at 11 years and national exam scores in 25 subjects at 16 years, a large effect size was reported for the correlation between *g* and a latent trait of educational attainment (correlation of 0.81) (Deary, Strand, et al., 2007). Whereas

* Corresponding author.
E-mail address: I.Deary@ed.ac.uk (I.J. Deary).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.04.005

in most cases g has explained about one quarter of the variance in academic success (Spinath, Spinath, Harlaar, & Plomin, 2006), representative population data using latent traits thus show that g could contribute up to two-thirds (0.81^2) of this variance.

Since the general intelligence factor is a strong predictor of school grades, when there are group differences in educational attainment levels it is valid to look to g as an underlying factor. Government statistics report a continuing trend, evident since the 1990s (Spinath, Spinath, & Plomin, 2008), that girls outperform boys in English or reading from nine years or older (UK National Statistics, 2009; Perie, Moran, & Lutkus, 2005). Furthermore, nationally representative data from at least two countries indicate that, by the age of 15–16 years, girls on average attain higher grades in the large majority of academic subjects (Deary, Strand, et al., 2007; Gustaffson & Balke, 1993). However, there appear to be no reliable and substantial sex differences reported for a g factor extracted from well-standardised cognitive ability tests taken in childhood, which could make a contribution to girls' advantage at school. When significant sex differences in mean general cognitive ability are cited in the literature, the effect size is often small, or sampling bias is a confounding issue (Blinkhorn, 2005). Cohorts of schoolchildren that demonstrate valid representative sampling methods report negligible sex differences for either mean IQ scores (Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Strand, Deary, & Smith, 2006) or g (Deary, Irwing, Der, & Bates, 2007; Deary, Strand, et al., 2007; Deary et al., 2003; Strand et al., 2006). If there is no gender effect on mean general cognitive ability, then differences in learning outcomes of boys and girls may more likely be driven by sexually dimorphic social influences, subject-choice (Van Langen, Rekers-Mombarg, & Dekkers, 2006), and/or psychological factors such as internalised belief systems (Spinath et al., 2006) and motivation (Spinath et al., 2008). Also, as we discuss next, it is possible that sex differences observed in specific cognitive domains may yet explain some of the different academic outcomes of boys and girls.

Studies of sex differences on mean scores of intelligence sub-tests, which reflect performance in specific cognitive domains, show diverse findings (Reynolds, Keith, Ridley, & Patel, 2008). One recent review reported that, on average, girls achieve higher total scores than boys on verbal reasoning tests, while more boys than girls are among the highest performers in quantitative and visual–spatial tests (Halpern et al., 2007). This is supported by a recent nationally representative cohort in which girls showed a small advantage over boys on mean verbal reasoning scores ($d = 0.15$), and they were more represented in the top 5% of performers compared to boys and, whereas there were negligible sex differences on mean nonverbal and quantitative reasoning test scores ($d = -0.03$ and 0.03 respectively), significantly more boys than girls were in the top 5% of performers in these domains (Strand et al., 2006). An over-representation of boys or girls in the extreme tail of the distribution can be caused by a difference in the mean and/or difference in variance between the sexes. However, a major issue in testing for sex differences in specific cognitive domains is that they may be obscured in mean subtest scores by the variance explained by g (Johnson & Bouchard, 2007). When statistical adjustment for the influence of g is made, residual factor scores from specific-reasoning ability tests show increased effect sizes of mean differences between girls and boys (Johnson & Bouchard, 2007). It is these factors which may contribute to the sex differences observed in educational scores.

In a recent five-year longitudinal study, verbal residual scores measured at 11 years were significantly higher for girls than boys, with an effect size of $d = 0.25$ (Deary, Strand, et al., 2007). However, linear regression models confirmed that this verbal residual factor did not make a substantial contribution to the higher educational achievements of girls by the time they reached 16 years. One explanation by the authors is that curriculum tests at age 16 rely heavily upon essay writing ability, at which girls show an advantage above boys (Halpern et al., 2007). Writing fluency and capability, not necessarily captured by standardized verbal tests, could explain the educational advantage held by girls in a broad range of subjects. Therefore, investigating sex effects on the association between residual factor scores and educational outcome may be more valid at a younger age of follow up, when school tests are less dependent on extensive essay writing.

The analysis of school test scores at a younger age might also benefit in reducing the risk of gender-related bias. For example, differential drop-out rates for boys and girls are reported to peak at 14 years of age, with four times as many boys as girls being excluded from schools in the UK (UK National Statistics, 2007). If children excluded from school were on average lower performers on cognitive tests, then such selection bias could inflate the mean performance of boys in cohort studies of older schoolchildren. Furthermore, the higher variance in cognitive ability in boys is consistent with more boys being represented in the extreme tails of the distribution (Johnson, Carothers, & Deary, 2008). Another potential source of gender-related bias could come from the academic choices of older schoolchildren. Boys and girls may differ according to social and psychological processes involved in subject selection at school, which generally begins in UK schools at age 14 years. Such subject choices will influence levels of success in end-of-school exams.

In the present study we have examined a large population-based cohort with educational outcomes and cognitive scores of 11-year-old children, at which age school exclusion rates are greatly reduced compared with later adolescence. The potential for sample bias in our cohort is also reduced by the fact that all students in England are assessed on the same three subjects: English, mathematics and science, and so there is no self-selection bias in relation to subject choice. Using data from the UK's Cognitive Abilities Test as an indicator of general intelligence we describe the means and standard deviations of general cognitive ability scores in girls and boys (cf. Deary, Strand, et al., 2007; Strand et al., 2006). We report the effect sizes of sex differences in specific cognitive domains, based on standardised and residual (removing g) scores from verbal, nonverbal and quantitative reasoning assessments. We examine sex differences in educational attainments at age 11 in the cohort (cf. UK National Statistics, 2005). The principal aims are to investigate: (1) sex differences in the residual scores of specific reasoning abilities: verbal, nonverbal, and quantitative; (2) whether specific cognitive specializations contribute to any gender gaps in the various educational achievements.

2. Methods

2.1. The dataset

The cognitive ability scores (CAT3) of 178,599 11-year-old schoolchildren in England from 2004, were linked to a database of their educational grades in English, mathematics and science subjects from the same year. This cohort represented 93% of the country's local education authorities, covered 1531 state and special education schools, and included 30% of all pupils in England that completed KS2 tests during 2004 ($n=592{,}000$). Among the sample there were 89,545 girls and 89,054 boys of equivalent age ($M=133.5$ months; $SD=3.5$).

2.2. General and specific cognitive reasoning

In recent years the Cognitive Abilities Test—Version 3 (CAT3) has been the most frequently used standardised test of reasoning ability in the UK. During the English autumn school term from September to December, 11- to 12-year-old state pupils complete Level D of the CAT3 which is standardised for this age group (Smith, Fernandes, & Strand, 2001). The battery involves nine time limited subtests with multiple-choice questions, each measuring verbal, quantitative, or nonverbal reasoning abilities. Oral instructions are given at the beginning of each subtest, reducing the demand on reading skills for adequate performance, and the complete battery lasts three sessions of between 50 min and 1 h. The verbal component of the test requires either inductive reasoning by conceptually linking words (*verbal classification*), or deductive reasoning by placing a correct word in the context of a sentence (*sentence completion*) or analogy (verbal analogies). The nonverbal battery of tests (*figure classification*, *figure analogies* and *figure analysis*) assesses the reasoning with, and manipulation of, visual patterns or representations. The quantitative component assesses reasoning ability with numbers (*number analogies*, *number series* and *equation building*), without putting demands on specialised mathematical knowledge. Each cognitive domain provides an age-standardised score based on normative test scores ($M=100$, $SD=15$). The CAT3 shows good reliability (Smith et al., 2001) and strong validity: performance on Level D highly correlates with scores from the Weschler Intelligence Scale for Children (third edition) for ages nine to 13 ($r=.82$) (Wright, Strand, & Wonders, 2005).

Our sample has scores on the CAT3 that are normally distributed and range from 59 to 141. Mean performance on the different cognitive tests are consistent with normative data (means: verbal reasoning = 100.3, quantitative reasoning = 99.8, nonverbal reasoning = 100.8), although the variances are slightly less than what would be expected in the general population (*SD*: 14.2–14.5). This may indicate that extreme scorers are absent from our sample, perhaps because they are more likely to attend specialist (low scorers) or independent (high scorers) schools. However, the differences between the study samples' score distributions and normative data are extremely small according to effect sizes (Cohen's *d* between .01 and .05). Therefore, the sample is arguably well-representative of reasoning abilities among the general population of 11-year-olds in England.

In principal axis factoring (varimax) input of verbal, nonverbal and quantitative scores lead to the extraction of a single unrotated factor accounting for 81% of the variance, which is commensurate with Spearman's general intelligence factor (g). All three cognitive ability subtests loaded highly on g: verbal reasoning (.82), quantitative reasoning (.87) and nonverbal reasoning (.85). A g factor score ($M=0.00$, $SD=0.94$) was saved for each pupil in the dataset. Using linear regression we calculated three standardised residual scores to reflect specific cognitive reasoning domains, where g was the independent variable, and verbal, quantitative and nonverbal reasoning were dependent variables.

2.3. Educational attainment

Key Stage Two (KS2) national curriculum tests (Autumn Package, 2004) are completed by 10–11 year-old pupils in the English state school system in May of each year. These involve assessing comprehension and current knowledge in three principal subjects: English, mathematics and science. KS2 results can either have an effect at the individual level, whereby scores may be used to determine a pupil's future class or school placement, or at a group level, where performance levels of schools and local authorities are monitored by the national government to ensure educational standards are met. In the UK the government target is for every child to achieve a raw score of 27 or above in each KS2 subject, which is classified as a Level 4 attainment (Autumn Package, 2004). In 2004, 77% of all pupils tested attained Level 4 in English, 74% did so in mathematics, and 86% in science (UK National Statistics, 2005). A subject raw score of less than 21 is given the classification of Level 2; scoring in the range 21 to 26 is a Level 3 classification; a higher performance of 33 points or above is classified as Level 5. The data available to us were pupil performance levels in the three KS2 subjects (Levels 2, 3, 4 or 5), rather than their original raw scores. In our sample the relative number of pupils who attained Level 4 in each subject is 2–3% higher than those of the general population in 2004. However, proportions of pupils to achieve the highest performance level (Level 5) in our sample, is more similar to the national picture, indicating that lowest achieving rather than highest schools are more likely to be absent.

We conducted principal axis factoring (varimax) with the English, mathematics and science level scores, and extracted a single unrotated factor from the scree plot which accounted for 77.3% of the variance in scores. Each of the academic subjects loaded highly on this educational factor (e): English (.77), mathematics (.84) and science (.83). An e factor score was saved for each pupil for analytic purposes ($M=0.00$, $SD=0.93$).

Government-released data from 2004 reported that the Level 4 attainment rate in English KS2 test was higher among girls (83%) than boys (72%), whereas an equal number of boys and girls reached the government-set target level in the other two subjects (74% in mathematics; 86% in science). Our sample reflects a similar trend in educational performance by sex, in particular, that approximately 10% more girls achieved the expected level in English than boys (see Table 1). We observed a slightly better rate of attainment by boys in mathematics and science compared to government statistics from the same year (differences of 1.9% and 0.5% respectively).

Table 1
Percentages of boys and girls achieving the target performance grade (Level 4 or above), in each of three academic subjects, from (i) the study sample and (ii) UK government statistics from 2004.

Subject	Boys		Girls	
	Study sample	Government data	Study sample	Government data
English	75.5%	72%	85.5%	83%
Mathematics	78.2%	74%	76.3%	74%
Science	89.1%	86%	88.6%	86%

Note. Proportions of boys and girls achieving government-set educational targets are significantly higher in the study sample compared to national data (UK National Statistics, 2005), Pearson's chi-squared: $\chi^2(1)$, $p<0.001$.

3. Results

3.1. Cognitive ability and education

As expected, the general intelligence (g) and educational (e) factors were strongly correlated at age 11, with a coefficient of $r_S=.83$ (see Table 2). Therefore, cognitive ability accounted for at least 69% of the variance in an emergent factor of education attainment. In relation to academic subjects, g most strongly correlated with mathematics level scores ($r_S=.80$), followed by science ($r_S=.70$), and then English ($r_S=.66$). Among children who achieved the mean g score or higher, the large majority reached the government set target level in English, mathematics and science (97.3%, 98.9% and 99.7% respectively). Among those with a g score of one standard deviation below the mean or less, 49.9% attained Level 4 in science, 30.3% attained Level 4 English, and only 18.1 % achieved the expected level in mathematics.

Specific cognitive ability was therefore less associated with educational success after adjusting for the effects of g. There was a small positive correlation between the verbal residual factor and e ($r_S=.07$), and a small negative correlation between the nonverbal residual factor and e ($r_S=-.10$). No association was observed between the quantitative residual factor and e ($r_S=.00$). As might be expected, the verbal residual factor was most positively associated with success in English ($r_S=.23$, $p<0.001$), and the quantitative residual factor most strongly correlated with mathematics grade ($r_S=.13$, $p<0.001$) (data not shown). The non-verbal residual factor did not correlate positively with any educational outcome, and was most strongly negatively correlated with English performance ($r_S=-.20$, $p<0.001$).

3.2. Sex differences in cognitive performance and educational attainment

Table 3 compares boys and girls on standardised and residual cognitive scores, academic subject level scores, and the g and e factors. The sex difference in general intelligence (g) was of negligible effect size (Cohen's $d=0.01$), although a slightly higher mean score by girls did achieve statistical significance in this large dataset. Girls performed on average 2.1 points above boys on verbal reasoning ($d=0.14$), and 0.4 points higher on nonverbal reasoning ($d=0.03$). Boys had a mean advantage of 1.4 points over girls on quantitative reasoning ability ($d=0.10$). There was significant heterogeneity of variance between boys' and girls' CAT3 scores: boys were 22% more variable than girls on quantitative reasoning, 14% more variable on nonverbal reasoning, and 12% more variable on verbal reasoning. Boys were 16% more variable on their g scores than girls. After controlling for the variance accounted for by g, sex differences showed increased effect sizes for the specific cognitive domains. For example, girls had a significantly higher mean score on the verbal residual factor ($d=0.26$), and boys' mean score on the quantitative residual factor was higher than the girls' ($d=0.28$). The nonverbal residual factor did not show a sizeable sex difference ($d=0.02$).

In English, girls performed significantly better than boys according to level scores ($d=0.33$) (see Table 3), and the odds of a girl attaining Level 4 English were nearly twice that of a boy (OR = 1.92, 95% CI [1.87, 2.00], $p<.001$). In science and math's performance, although boys performed on average higher than girls, these were negligible effect sizes ($d=0.08$, 0.03 respectively). A slightly larger proportion of boys than girls reached the expected level in science (OR = 1.06, 95% CI [1.03, 1.08], $p<.001$) and mathematics (OR = 1.11, 95% CI [1.08, 1.12], $p<.001$).

Table 4 shows the proportions of boys and girls who performed at each of the four educational grades in each subject. In English a higher proportion of girls reached the highest grade than boys (33.7% and 21.5%), and the proportions achieving the lowest grade in this subject were twice as much for boys as for girls. In mathematics a greater proportion of boys achieved the highest performance grade

Table 2
Correlations (r_S) among cognitive subtest and residual scores, educational levels, g and e.

Variable	1	2	3	4	5	6	7	8	9	10	11
Verbal	–										
Quantitative	.71	–									
Nonverbal	.69	.74	–								
g factor	.86	.92	.90	–							
Verbal R	.40	−.21	−.24	−.07	–						
Quantitative R	−.23	.38	−.19	.03	−.49	–					
Nonverbal R	−.23	−.17	.41	.01	−.54	−.40	–				
English	.71	.60	.52	.66	.23	−.06	−.20	–			
Mathematics	.67	.79	.69	.80	−.08	.13	−.06	.61	–		
Science	.69	.61	.61	.70	.12	−.08	−.06	.60	.67	–	
e factor	.79	.77	.70	.83	.07	*.00	−.10	.79	.89	.89	–

Note. R = residual factor score. Spearman's *rho* coefficients are all statistically significant at $p<0.001$ except that indicated with *.

Table 3
Performance of boys and girls on cognitive and education variables, and correlations between these variables.

Variable	Male–female comparisons			
	Boys' M (SD)	Girls' M (SD)	Effect size [a] (Cohen's d)	Variance ratio [b] (VR)
	$N=89{,}054$	$N=89{,}545$		
Cognitive standardised and factor scores				
Verbal	99.2 (14.9)	101.3 (14.1)	−0.14**	1.12**
Quantitative	100.5 (14.9)	99.2 (13.5)	0.10**	1.22**
Nonverbal	100.6 (14.7)	101.0 (13.8)	−0.03**	1.14**
g factor	−0.006 (0.975)	0.006 (0.904)	−0.01*	1.16**
Verbal R	−0.127 (1.025)	0.126 (0.958)	−0.26**	–
Quantitative R	0.144 (1.034)	−0.143 (0.944)	0.28**	–
Nonverbal R	−0.006 (1.036)	0.006 (0.963)	−0.02*	–
Educational level and factor scores				
English	3.902 (0.808) $n=88{,}005$	4.159 (0.746) $n=88{,}721$	−0.33**	1.17**
Mathematics	4.086 (0.831) $n=88{,}027$	4.018 (0.814) $n=88{,}672$	0.08**	1.04**
Science	4.323 (0.710) $n=88{,}157$	4.304 (0.708) $n=88{,}821$	0.03**	1.00*
e factor	−0.021 (0.94) $n=87{,}248$	0.021 (0.92) $n=88{,}132$	−0.05**	1.05**

Note. Verbal, nonverbal and quantitative are the standardised scores; Verbal R, nonverbal R and quantitative R are the residual factors of those subtest scores.
[a] Independent samples t-tests (2-tailed), $^*p<0.05$, $^{**}p<0.001$. Negative value indicates girls' advantage over boys.
[b] VR = male $(SD)^2$/female $(SD)^2$; Levene's test, $^*p<0.05$, $^{**}p<0.001$.

than among girls (34.8% and 29.7%). In science, although the difference in performance distributions of boys and girls reached statistical significance, at most this reflected a difference of 1.2%, where boys (44.8%) were very slightly more represented than girls (43.4%) at the highest performance grade.

Correlation coefficients were calculated between all cognitive and educational variables, for boys and girls separately (data not shown). The associations between the g and e factors were the same for both sexes, and there were no non-negligible differences in any of the correlations between cognitive and education variables according to sex.

3.3. g and sex in association with educational attainment

In General Linear Models (ANCOVA) to predict success in the three academic subjects, sex was a fixed effect and age and g were covariates in the basic model. Due to significant sex differences in verbal and quantitative residual scores, and their significant positive correlations with English and mathematics respectively, we ran further models adding each of these covariates in turn. All variables made statistically significant contributions to the models, even when these were small (partial eta squared of less than 0.1%). g made the most substantial contribution in each model (Table 5). For example, in the most explanatory models g predicted 64.4% of the variance in mathematics and over half the total variance in English (52.8%) and science (50.6%). Sex explained 4.8% of the variance in English in the basic model, and made very small contributions to mathematics (0.6%) and science performance (0.1%). Addition of the verbal residual factor to the basic model, to predict English, meant this covariate explained 14.8% of the variance, and resulted in an attenuation of the sex effect, by 29%. This means that girls' better scores on the verbal residual factor contributed to less than one third of their better attainment levels in English, which means that, factors that influence sex effects on English performance remain largely unaccounted for. The quantita-

Table 4
Distribution of educational performance levels attained by boys and girls in English, mathematics and science.

Attainment reached	English		Mathematics		Science	
	Boys	Girls	Boys	Girls	Boys	Girls
Level 2	6.8% (5953)	3.3% (2967)	4.4% (3898)	4.3% (3801)	1.7% (1493)	1.5% (1360)
Level 3	17.7% (15,591)	11.1% (9859)	17.4% (15,278)	19.4% (17,181)	9.2% (8081)	9.9% (8792)
Level 4	54.1% (47,574)	51.8% (45,991)	43.5% (38,248)	46.6% (41,319)	44.3% (39,042)	45.2% (40,147)
Level 5	21.5% (18,887)	33.7% (29,904)	34.8% (30,603)	29.7% (26,371)	44.8% (39,541)	43.4% (38,522)

Note. Pearson's chi-square was used to compare the distribution of educational levels attained by boys and girls, per subject: English: $\chi^2(3)=4802$, $p<0.001$; mathematics: $\chi^2(3)=543$, $p<0.001$; science: $\chi^2(3)=62.4$, $p<0.001$.

Table 5
General linear models (η^2 values) of educational attainment in English, mathematics, and science at age 11.

	English models ($N=176{,}726$)			Mathematics models ($N=176{,}699$)			Science models ($N=176{,}978$)		
	1 [a]	2 [b]	3 [c]	1	2	3	1	2	3
g	.476	.525	.528	.639	.638	.644	.485	.506	.506
Sex	.048	.034	.040	.006	.005	.003	.001	.003	.003
Age	.018	.033	.037	.023	.023	.027	.011	.017	.017
Verbal *R*	–	.148	.161	–	.000	.007	–	.062	.043
Quantitative *R*	–	–	.021	–	–	.032	–	–	.000

Note. All models include sex as a fixed effect and age as a covariate. The values shown represent η^2. All effects are statistically significant, $p<0.001$ (even where the η^2 value is signified as 0.000).
[a] Model 1 includes *g* as a covariate.
[b] Model 2 includes *g* and the verbal residual score.
[c] Model 3 includes the following covariates: *g*, verbal residual score and quantitative residual score.

tive residual factor explained a significant 3.2% in the full mathematics model, and this resulted in an attenuation of the sex effect from the basic model, by 50%. Therefore, boys' better scores on the quantitative residual factor explain half of their better mathematics scores compared to girls.

As we observed different proportions of boys and girls in the top performance grades for English and mathematics (Level 5), we used multivariate logistic regression to evaluate the contribution of *g*, sex and age to attainment at this level, allowing for interaction effects between sex and the cognitive factors. When modelling for English attainment, the verbal residual factor was also included as a covariate. The number of children who reached the highest grade in English was 48,791, compared to 129,808 who did not. Our predictor model correctly classified 82.9% of schoolchildren's grades. All variables and interaction effects made statistically significant contributions to predicting a high performance in English, with the following odds ratios: $g=10.03$ (95% CI [9.72, 10.36], $p<0.001$), verbal residual $=2.28$ (95% CI [2.24, 2.33], $p<0.001$), sex (female) $=0.45$ (95% CI [0.44, 0.47], $p<0.001$), age $=1.13$ (95% CI [1.12, 1.13], $p<0.001$), Sex (female) $\times g=0.82$ (95% CI [0.78, 0.86], $p<0.001$), and Sex (female) $\times$ Verbal Residual $=0.94$ (95% CI [0.92, 0.97], $p<0.001$). The interaction effects of sex with *g* or the verbal residual factor are shown in Fig. 1a and c. These show that small sex differences in general intelligence are evident among high performers in English. Although there are negligible sex differences in *g* in the total sample, Fig. 1a shows that boys require a higher mean general intelligence score to excel in higher grade English compared to girls at age 11 years.

In modelling the predictors of top mathematics performance the quantitative residual factor replaced the verbal residual factor. In the study population 56,974 schoolchildren reached top grade mathematics, while 121,625 did not. Our model correctly classified 86.6% of children, and all of the variables and interaction effects were significant predictors. The odds ratios for predicting high attainment mathematics were: $g=29.70$ (95% CI [28.33, 31.14], $p<0.001$), quantitative residual $=1.44$ (95% CI [1.14, 1.47], $p<0.001$), sex (male) $=1.37$ (95% CI [1.31, 1.43], $p<0.001$), age $=1.12$ (95% CI [1.11, 1.12], $p<0.001$), Sex (male) $\times g=0.89$ (95% CI [0.83, 0.95], $p<0.001$), and Sex (male) $\times$ Quantitative Residual $=1.08$ (95% CI [1.05, 1.11], $p<0.001$). Fig. 1b and d shows the interaction effects of sex with *g*, and sex with the quantitative residual factor, in predicting higher grade mathematics. Fig. 1b shows that the Sex by *g* interaction effect on high performance mathematics is due to small sex differences in *g* among the low to moderate performers. Whereas, sex differences in the quantitative residual factor are greater among the highest performers in mathematics (Fig. 1d), and boys are increasingly more likely than girls to excel in mathematics as their quantitative residual scores increase.

4. Discussion

In 11 year-old children from a large representative sample of the English population, general factors of cognitive ability and educational attainment correlated strongly (.83). The *g* factor was significantly associated with the outcome of three core academic subjects, contributing to 64% of the variance in mathematics scores, and around half of the total variance in English and science outcomes. Attainments in English and science were further predicted by a verbal residual factor (16% and 4% of the variance in scores, respectively), and a quantitative residual factor made a small but significant contribution to mathematics (3%). Girls performed ahead of boys in English and, whereas boys' mean scores were significantly above those of girls in science and math subjects, the statistical effect sizes were negligible. After controlling for the effects of *g* and the verbal and quantitative residual factors, gender continued to make a small contribution to English attainment scores (4%), but only negligibly to mathematics and science outcomes. Whereas we observed that girls' verbal reasoning ability was significantly higher than that of boys, this explained less than one third of their better English results. Quantitative reasoning explained up to a half of boys' better performance in mathematics, although the independent sex effect on mathematics performance was very small, explaining 0.3% of variance.

The main strength of this study is its nationally representative sample. The size of our dataset is extremely large, and represents approximately one third of all 11-year-old children in state school education in England in a single year. It covers nearly all local education authorities in the country. The sample reflects the general population's range of reasoning abilities, matching standardised norms on the cognitive tests. There is extremely low risk of ascertainment bias, given that the majority of school exclusions occur in later adolescence (UK National Data, 2007). The academic tests are delivered nationally each year, and every child in mainstream education completes tests in the same subjects, therefore there is no differential subject selection bias of girls and boys.

One weakness of our study is the cross-sectional design which compromises any interpretation of a causal association between cognitive ability and educational outcome. Despite this, the CAT3 battery has been validated as a stable measure of general intelligence (Wright et al., 2005). Furthermore, there is longitudinal evidence for the existence of a strong causal link between *g* and educational attainment, whereas a reciprocal relationship in which educational learning might also influence *g* is less supported (Watkins, Lei & Canivez,

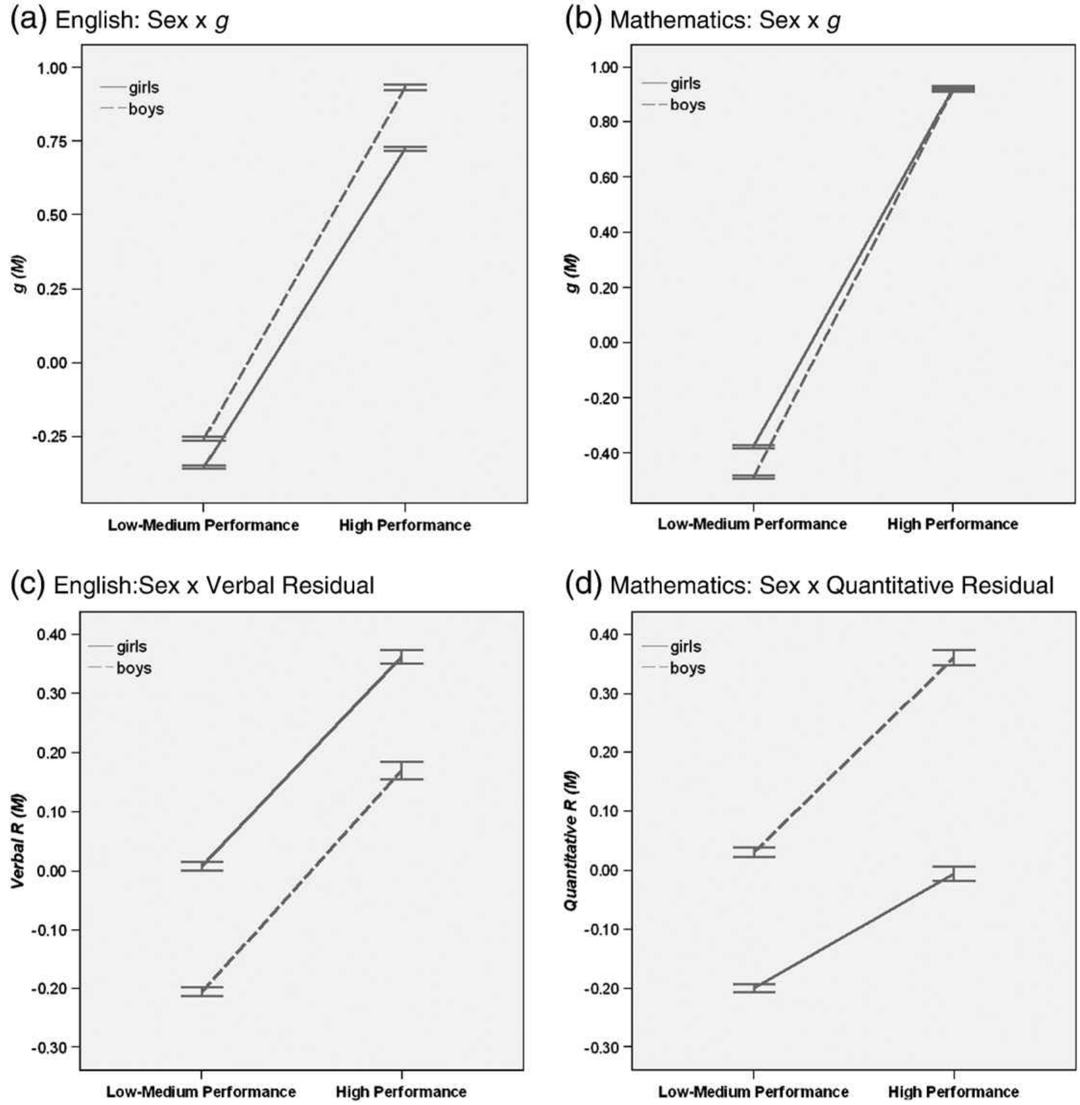

Fig. 1. Mean performance on cognitive factors by sex, among low-medium and high performers of English and mathematics.

2007). It is important to remember that the CAT3 and KS2 tests do not measure the same thing: CAT3 assesses transferable reasoning abilities that can be utilised for a wide-range of problem-solving tasks, whereas KS2 tests indicate specific knowledge acquired within a particular school curriculum (Strand, 2006). The strong association between our cognitive and educational factor scores ($r = .83$) replicates that reported by Deary, Strand, et al. (2007) in their 5-year longitudinal study (.81), supporting our view that cognitive ability as a stable underlying trait, may have strongly predicted academic success in our cohort. We did not have access to the raw scores of the sub-tests within English, mathematics and science and, instead, based our analysis on fairly crude categories of general performance levels in these three subject areas. This may have weakened any sex effects on educational performance that were observed.

We replicated an emerging trend in the literature of negligible sex differences in childhood on mean scores of general cognitive ability, or *g*. After adjusting for the influence of *g* on the variance in cognitive subtest performance, we found small sex differences at age 11 in verbal and quantitative residual factors extracted from performance scores on well-standardised cognitive ability tests, whereas none existed on the nonverbal residual factor derived from tests demanding visuo-spatial manipulation and reasoning. The effect size for girls' higher verbal residual score ($d = .26$) is equivalent to that previously reported for this age group (Deary, Strand, et al., 2007; Van der Sluis et al., 2008), and is larger than the gender effect on verbal reasoning out with adjusting for the general ability factor, as predicted by Johnson and Bouchard (2007). Quantitative ability scores, that showed negligible sex differences on mean performance, indicated a male advantage once the effects of *g* on their variance were removed, with $d = .28$. This finding has previously been observed in a study sample of 10- to 11-year-olds, for which a latent quantitative reasoning factor, derived from a higher-order model of intelligence, was higher for boys than girls (Keith, Reynolds, Patel, & Ridley, 2008). In the complete cohort of 6- to 59-year-olds, from where this sample came, it was evident that the male advantage on the quantitative trait continued to increase with age. Such cognitive specializations could reflect differences in male and female neuroanatomical function and structure (Haier, Jung, Yeo, Head, & Alkire, 2005; Narr et al., 2007; Schmithorst & Holland, 2007). Unlike evidence from a few studies that have reported on sex differences in a latent visual–spatial factor (Keith et al., 2008; Reynolds et al., 2008), we did not observe a male advantage on the nonverbal residual factor. This inconsistent finding may be due to the different methods use to derive latent versus residual factor scores, or differences between the contents of the specific tests used across the studies. However, in at least one of the previous studies, the male advantage in visual–spatial reasoning, whilst controlling for general ability, was not statistically significant among 10- to 11-year-olds, but only became so in older age bands (Keith et al., 2008).

Despite the female advantage in a verbal residual factor at age 11 years, this only explained just over a quarter of girls' superior performance in academic English at the same age. Deary, Strand, et al. (2007) suggested that girls might perform better in the majority of educational tests at age 16 because of a verbal advantage that is not detected in cognitive tests at age 11. One aspect of school examination at age 16 years in the UK is the strong reliance upon essay writing, a verbal ability that is reported to show the largest sex differences in favour of girls (Willingham & Cole, 1997). Furthermore, trend analysis of the widening gender gap in educational attainment at age 16 reveals that changes introduced to the UK's national curriculum in 1988, increased written coursework to the examination system, which more girls than boys have a preference for over examination testing (Machin & McNally, 2005). If writing skills contribute substantially to the gender gap in educational outcomes, this would highlight a weakness by educational departments to assess academic attainment fairly, without gender bias. Our educational data were measured using tests that are less dependent on writing technique and, whereas, girls excelled in English at age 11, this was not to the same effect size as at 16 years ($d = 0.41$) (Deary, Irwing, et al., 2007). Therefore, there is still the possibility that educational testing methods may favour girls to a greater degree than boys in certain populations. Alternatively, it may be that a range of behavioural, psychological and/or social factors contribute to the gender effects on educational tests we observed at 11 years, and to the small to medium effect sizes of sex differences in education in later adolescence. For example, in a recent 9-year-old twin cohort, ability self-perceptions were reported to contribute incrementally to girls' higher attainment in English tests and boys' better performance in mathematics tests (Spinath et al., 2008). Such variables may become more or less prominent throughout child development and learning, which longitudinal studies can address.

Boys' mean performance in mathematics and science was marginally higher than girls, and so there was relatively little gender effect for the cognitive residual factors to explain. Nevertheless, we observed a significant interaction effect between the quantitative residual factor and sex in predicting high mathematics performance, so that, after controlling for g, quantitative reasoning was a better predictor of mathematics performance in boys than girls. This may indicate sex differences in the strategies used to complete numerical reasoning tasks, perhaps reflecting differences in neuroanatomical resources (Kellor & Menon, 2009). However, a closer examination of the nature of the quantitative tests is warranted, and it should be addressed whether sex differences on individual tests are better explained by a broader cognitive factor (consistent with higher-order models of general intelligence), for example, working memory.

It is likely that our verbal, nonverbal and quantitative factor scores did not adequately reflect the full range of specific cognitive specializations that explain sex differences in educational outcomes at age 11. When latent variables are extracted from broader cognitive test batteries a different pattern of specific cognitive domains are represented. For example, in a U.S. study of 2375 boys and girls aged between six and 18 years of age, scores from the Kaufman Assessment Battery for Children-version II extracted five latent cognitive factors, including crystallised and fluid intelligence, visual–spatial reasoning, and short- and long-term memory. This reported sex differences on the visual–spatial factor, revealing boys' higher performance equivalent to six IQ-points' difference at age nine to 11 years (Reynolds et al., 2008). In a study of 522 adults who completed the Dutch WAIS-III a verbal first order factor was extracted, but this revealed no sex differences, whereas significant sex effects were observed for working memory and perceptual organization factors favoring men, and on a perceptual speed factor favoring women (Van der Sluis et al., 2006). Without the subtest scores from all nine of the verbal, nonverbal and quantitative tests of the CAT3 we were unable to extract latent cognitive factors which may perhaps have better explained the sex differences we observed in educational attainment at age 11. To take this further, if raw scores rather than performance levels from the educational tests had also been available, it would then have been appropriate to use structural equation modelling (SEM) to analyse sex differences in the associations between the higher-order structures of both cognitive and educational measures. In studies investigating the effects of broad cognitive abilities on educational attainments, and applying SEM techniques to scores on extensive test batteries, direct effects on mathematics attainment are shown for the broad factors of fluid and crystallised intelligence, as well as processing speed (Taub, Keith, Floyd, & McGrew, 2008). On the other hand, reading attainment, for example, may be more dependent on different patterns of broad factors, including comprehension-knowledge, short-term memory, auditory processing, long-term retrieval and processing speed (Evans, Floyd, McGrew, & Leforgee, 2001).

In a population of 11-year-old schoolchildren we confirm that sex differences on educational attainment cannot be explained by general cognitive ability. However, we present evidence that girls and boys show some small advantage in verbal and quantitative reasoning abilities respectively, and these cognitive specialisations contribute somewhat to sex differences in school grades.

Acknowledgements

The authors wish to thank the UK Government's Department of Children, Schools and Families for granting permission to use the Key Stage 2 educational data. The work was undertaken by The University of Edinburgh Centre for

Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC), Engineering and Physical Sciences Research Council (EPSRC), Economic and Social Research Council (ESRC) and Medical Research Council (MRC) is gratefully acknowledged.

References

Autumn Package 2004. (2004). Retrieved April 24, 2008, from http://www.standards.dfes.gov.uk/performance/ap/

Bartels, M., Rietveld, M. J., Van Baal, G. C., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research*, *5*(6), 544−553.

Blinkhorn, S. (2005). Intelligence: A gender bender. *Nature*, *438*(7064), 31−32.

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother–sister differences in the g factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence*, *35*, 451−456.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13−21.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish Mental Survey 1932. *Intelligence*, *31*, 533−542.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2001). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*(2), 246−262.

Gustaffson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*(4), 407−434.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *Neuroimage*, *25*(1), 320−327.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Shibley Hyde, J., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1−51.

Johnson, W., & Bouchard, T. J. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence*, *35*, 23−39.

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, *3*, 518−531.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence*, *36*(6), 502−525.

Kellor, K., & Menon, V. (2009). Gender differences in the functional and structural neuroanatomy of mathematical cognition. *NeuroImage*, *47*(1), 342−352.

Machin, S., & McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy*, *21*(3), 357−372.

Narr, K. L., Woods, R. P., Thompson, P. M., Szeszko, P., Robinson, D., Dimtcheva, T., et al. (2007). Relationships between IQ and regional cortical gray matter thickness in healthy adults. *Cerebral Cortex*, *17*(9), 2163−2171.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77−101.

Perie, M., Moran, R., & Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress: Three decades of student achievement in reading and mathematics* Retrieved November 11, 2009, from http://nces.ed.gov/nationsreportcard/pdf/main2005/2005464.pdf

Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence*, *36*(3), 236−260.

Schmithorst, V. J., & Holland, S. K. (2007). Sex differences in the development of neuroanatomical functional connectivity underlying intelligence found using Bayesian connectivity analysis. *Neuroimage*, *35*(1), 406−419.

Smith, P., Fernandes, C., & Strand, S. (2001). Cognitive abilities test third edition: Technical manual. London: GL assessment.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, *34*, 363−374.

Spinath, B., Spinath, F. M., & Plomin, R. (2008). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality*, *22*, 211−229.

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, *47*, 1−41.

Strand, S. (2006). Comparing the predictive validity of reasoning tests and national end of key stage 2 tests: which tests are the 'best'? *British Educational Research Journal*, *32*(2), 209−225.

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

Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *Social Psychology Quarterly*, *23*(2), 187−198.

UK National Statistics (2005). First Release: National Curriculum Assessments at Key Stage 2 Retrieved April 24, 2008, from http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000581/SFR22-2005v3.pdf

UK National Statistics (2007). First release: Permanent and fixed period exclusions from schools and exclusion appeals in England 2005/2006 Retrieved August 26, 2009, from http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000733/index.shtml.

UK National Statistics (2009). First Release: National Curriculum Assessments at Key Stage 2 in England Retrieved August 26, 2009, from http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000865/index.shtml

Van der Sluis, S., Derom, C., Thiery, E., Bartels, M., Polderman, T. J. C., Verhulst, F. C., et al. (2008). Sex differences on the WISC-R in Belgium and The Netherlands. *Intelligence*, *36*, 48−67.

Van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J. C., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence*, *34*, 273−289.

Van Langen, A., Rekers-Mombarg, L., & Dekkers, H. (2006). Sex-related differences in the determinants and process of science and mathematics choice in pre-university education. *International Journal of Science Education*, *28*(1), 71−94.

Watkins, M. W., Lei, P., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*, 59−68.

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment.* Mahwah, N.J.: Erlbaum.

Wright, I., Strand, S., & Wonders, S. (2005). Estimation of premorbid intellectual abilities in children Retrieved April 1, 2008, from http://www2.warwick.ac.uk/fac/soc/cedar/staff/stevestrand/wrightstrandwonderecp.pdf