Intelligence and educational achievement

Ian J. Deary
Department of Psychology, University of Edinburgh, Scotland, UK

Steve Strand
Institute of Education, University of Warwick, UK

Pauline Smith, Cres Fernandes.
nferNelson, London, UK

Intelligence, in press.

Correspondence,

Professor Ian J. Deary
Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
Scotland, UK

Email I.Deary@ed.ac.uk

Abstract

This 5-year prospective longitudinal study of 70,000+ English children examined the association between psychometric intelligence at age 11 years and educational achievement in national examinations in 25 academic subjects at age 16. The correlation between a latent intelligence trait (Spearman's *g* from CAT2E) and a latent trait of educational achievement (GCSE scores) was .81. General intelligence contributed to success on all 25 subjects. Variance accounted for ranged from 58.6% in Mathematics and 48% in English, to 18.1% in Art and Design. Girls showed no advantage in *g*, but performed significantly better on all subjects except Physics. This was not due to their better verbal ability. At age 16, obtaining 5 or more GCSEs at grades A*-C is an important criterion. 61% of girls and 50% of boys achieved this. For those at the mean level of *g* at age 11, 58% achieved this; a standard deviation increase or decrease in *g* altered the values to 91% and 16%, respectively.

Introduction

Ability testing is one of the most widespread yet most controversial exports from academic psychology to the real world, intended to provide an objective measure of the individual differences in cognitive abilities that undoubtedly exist within society. Firm evidence that psychometric test scores accurately predict real-world success would have considerable import at the practical and the theoretical levels. It would justify the use of such tests as educational and occupational selection tools and as dependent variables in studies of possible genetic and neurophysiological correlates of cognitive ability differences. Predicting individual differences in educational outcomes was the raison d'être for the first broad test of cognitive ability (Binet, 1905; Zenderland, 1998). The discovery of general intelligence involved, in part, using individual differences in school examination scores (Spearman, 1904). Alongside occupational outcomes (Schmidt & Hunter, 1998) educational outcomes are the major target for the predictive validity of cognitive ability tests.

What, then, is the association between cognitive ability and educational achievement? There is broad agreement that there is a moderate to strong correlation between the two. Jencks et al.'s (1979, p. 102) detailed account of eight samples from six longitudinal studies reported correlations ranging from .40 to .63 between cognitive test scores and amount of education obtained. More recent overviews are provided by various authors and reach similar conclusions (Brody, 1992; Neisser et al., 1996; Jensen, 1998; Sternberg et al., 2001; Bartels et al., 2002b). For example, Mackintosh's (1998) survey reckoned that there is a correlation between .4 and .7 between IQ scores and school performance grades. More Specifically for the present study, Mackintosh stated that, "in Britain, the correlation between 11-year-old IQ scores and later educational attainment, including performance on school examinations at age 16, is about 0.5".

The present study will provide a better estimate of the true association between intelligence and education by having multiple cognitive tests as predictors and multiple educational outcomes, applied to a massive representative sample. This echoes Spearman's (1904) original examination of the correlation between the latent trait from several school examination results and the latent trait from discrimination tests. For example, using just two ability tests and scores from four school examination areas (languages, science, maths-physics, and humanities), there was a correlation of .53 between ability and education latent traits (Rinderman & Neubauer, 2004).

Another major issue addressed by the current study is the gender gap in educational outcomes. Boys perform less well in school assessments than girls, despite similar scores on cognitive tests (e.g. Fergusson & Horwood, 1997). The similarity of boys and girls on cognitive test scores at age 11 is quite well established. In studies involving an entire Scottish population (Deary et al., 2003), or massive representative samples from the United Kingdom (Strand, Deary, & Smith, in press), boys and girls at 11 years of age show no appreciable differences in mean general cognitive ability. However, girls score slightly higher on verbal ability and boys have a slightly larger standard deviation on general and specific ability scores. Here we aim to determine whether the sex difference in verbal ability (after accounting for *g*) explains any of the sex difference in school assessment performances.

After 100 years of studying the cognitive ability-education association there is still a need for a definitive, prospective study, one which assesses initial cognitive ability and later educational attainment with comprehensive assessments. Here we report such a study, using a large, representative, five-year prospective examination of over 70,000 children in England. The present study asks: (1) what is the association between general cognitive ability and overall educational attainment in 25 different courses?; (2) what is the association between a latent cognitive ability trait (general intelligence or *g*) and a latent educational outcome trait?; (3) what is the effect of sex on examination performance, and is it accounted for by general cognitive and/or verbal abilities?; (4) in epidemiological terms, what is the effect size of cognitive ability on educational attainment?

Method

The Cognitive Abilities Test (CAT)

The CAT is the most widely used test of reasoning abilities in the UK, with close to one million school students assessed each academic year. The data reported here relate to the CAT second edition (CAT2E) (Thorndike, Hagen & France, 1986). CAT2E has ten separate sub-tests which are aggregated into three batteries, providing standardised measures of Verbal, Quantitative and Nonverbal reasoning abilities. An average of the three standardised scores, the mean CAT score, is also calculated. The tests are described in detail in Strand (2004) and in the test manual.

The CAT2E spans the age range 7:06-15:09 years and is divided into six levels of difficulty, Level A through to Level F. Level D is typically used in the first year of secondary school (year 7). The results are most often reported as standard age scores, with a mean of 100 and standard deviation of 15. The tests have very high levels of reliability. The internal consistency (KR21) estimates of the Level D batteries are: Verbal = .95; Quantitative = .90; and Non-verbal = .92 (Thorndike et al, 1986). The KR21s of the ten subtests range from .74 to .86 with a mean of .80 (Thorndike et al, 1986). The test-retest correlations are also high (Sax, 1984; Strand, 2004). For example, the Pearson correlation coefficients between age 10+ and 13+ (N = 10.644) were: Verbal = .87; Quantitative = .79; Non-verbal = .76; and Mean CAT score = .89 (Strand, 2004).

A third edition of the CAT (CAT3) was launched in June 2001 and is now more widely used than CAT2E. However an extensive equating study involving over 10,000 pupils allows CAT2E scores to be converted to CAT3 equivalents or *vice versa* (Smith, Fernandes & Strand, 2001).

Here, we have taken the age-standardised scores for the three batteries of the CAT2E and subjected them to principal axis factoring. Inspection of the scree slope clearly indicated a single factor. The first unrotated general ability factor (Spearman's *g*) accounted for 69.6% of the variance. A *g* factor score was computed and stored for each subject. Boys and girls did not differ significantly in *g* scores, even with about 35,000 in each group (Cohen's $d = 0.01$) (see also Strand, Deary & Smith, in press). We then created a residual verbal factor, unrelated to the *g* score, using linear regression with the Verbal reasoning standard age score as the dependent variable and *g* as the independent variable and saving the standardised residuals. Girls scored better in the residual verbal factor (Cohen's $d = 0.25$). The *g* and verbal scores have means of 1 and standard deviations of 0. The verbal score was used specifically in addition to *g* because it showed sex differences (Strand, Deary,& Smith, in press) and contributed additional variance to the educational outcomes.

National GCSE/GNVQ public examination results

Pupils in England sit national public examinations at age 15/16 years. The entire country's pupils sit the same examination, apart from the small number with moderate to severe learning difficulties who are educated in special schools. These are typically General Certificate of Secondary Education (GCSE) examinations, which are offered in a wide range of subjects and are graded from A* down to G (and U for ungraded) in each subject. For the purposes of the present analyses these are scored from 8 to 0, respectively. A small proportion of entries are also made for General National Vocational Qualifications (GNVQs). These can be awarded at Distinction, Merit or Pass levels. A detailed set of score equivalencies are defined by the government which allow GCSE/GNVQ results to be expressed on a common scale as 'points scores', and allow the calculation of the 'Best 8' performance score as an overall measure of attainment (Autumn Package, 2002). The 'Best 8' performance score gives the pupils' points score in the eight highest scoring units studied, whether they be GCSE full course, GCSE short course, GNVQ foundation or GNVQ intermediate examinations. Here, we examine the results from 25 different GCSE/GNVQ topics. We have divided them into groupings of arts and humanities, science, social sciences, and practical (Table 2). We also study the total points gained by each student, and the points gained from their 'Best 8' topics. In addition, we examine the binary variable of whether or not students gained 5 or more scores at grades A* to C, which is used widely in England as a performance criterion, e.g. for ranking school performances.

The present study's datasets and the matching process

CAT is processed through a national service which provides computer scoring and analysis of pupils' answer sheets. Consent is asked from schools to retain these data for further research on monitoring CAT norms or developing indicators, and 94% of users consent to this use. CAT test scores from administrations in the 1997/98 academic year, mostly in September-November 1997, were matched to the national GCSE 2002 dataset. Thus a dataset of matched autumn 1997 CAT results and May 2002 GCSE results was created. After the matching was complete, individual records were given a unique ID code and all identifying information on pupils' names was deleted from the file to assure confidentiality for individual pupils and schools.

CAT scores were found for over 80,000 of the pupils in the national GCSE dataset, with pupils drawn from a total of 973 secondary schools across 103 Local Education Authorities. Table 1 compares the GCSE results at age 16 for the CAT-matched sample with the national averages derived from the full national dataset. The GCSE/GNVQ performance score for the CAT matched sample does not differ substantially from the national average in either mean or standard deviation. There are also are no significant differences in the proportion of higher grades awarded in the three core GCSE subjects of English, mathematics or science. Given the matched CAT sample constitutes over one-fifth (22%) of the total national dataset, this is not surprising. For the present analyses we selected only those pupils who took the CAT2E Level D test (the vast majority), and only those students who attended mainstream state secondary schools (again, the vast majority). This gave a maximum data set of N = 74,403 (37,509 girls, 36,894 boys). The difference between this and the number in Table 1 is accounted for by pupils who were missing one or more scores from the CAT.

Results

The association between intelligence and education

The correlations between the CAT's *g* factor and individual GCSE subject scores were all positive and medium to large in effect size (Table 2). The correlation with overall GCSE points score was .69, and with GCSE Best 8 was .72. In the arts and humanities group, the highest correlation was with English (.67). Most others were around .6, with Religious Education and Drama lower, around .5. In the science group, Mathematics correlated highest, at .77, and the individual sciences (Physics, Chemistry and Biology), which are taken by much smaller numbers of students, correlated around .5. This group is characterised by high ability and restriction of range; for example, in the pupils who took physics, the mean and SD of *g* are 0.96 and 0.64 compared with 0.0 and 1.0 in the entire sample. In the social sciences, Geography and History correlated above .6 and others around .5. In the practical subjects, all correlations were between .43 and .54, with Physical Education (r = .55) and Music (r =.54) highest. They are similar to the correlations with the CAT total score for previous data (Smith, Fernandes, and Strand, 2001) and for selected school subjects with these data (Strand, in press).

The next analysis addressed the question about the true correlation between the latent traits of ability assessed in the CAT2E and educational performance in GCSE exams. For this, we used structural equation modelling (EQS; Bentler, 1995) to estimate the correlation between a latent trait of mental ability and educational examination scores.

Students take different combinations of school subjects, so we analysed the largest possible complete data set. There were 13,248 students with full CAT2E data, and GCSE scores on English, English Literature, Mathematics, Double Science, French, and Geography. (All analyses were repeated with the 12,519 students who did History rather than Geography and there were trivial differences in the outcomes.) The analysis in the full sample demonstrated that there was a single factor underlying performance on the CAT's three batteries (see Method). The results from the six GCSE subjects were subjected to principal axis factoring. A single factor was found, accounting for 71.8% of the variance. Therefore, the simple model in Figure 1 was fitted to the data, with the question being: what is the correlation between the two latent traits? The fit of the overall model is not of prime consideration here. However, for the model shown in Figure 1 the average of the off-diagonal standardised residuals = .04, the chi square = 7799.0 (d.f. = 26, $p < .001$), and the comparative fit index = .920. Note the high values of the path coefficients between the manifest and latent traits. The correlation between *g* and general educational achievement was .81.

Sex differences in educational attainment

Girls performed better than boys on overall GCSE points scores, with a Cohen's *d* of 0.30 (Table 2). There were significant sex differences ($p < .001$) in all GCSE scores except Physics. Girls performed better in every topic except Physics. The effect sizes of the sex differences were often substantial. The Arts and Humanities group had Cohen's *d* values tightly grouped around 0.4. The practical subjects group had some large effects, for example Art and Design had a Cohen's *d* of 0.61. The Science and Social Science subjects showed smaller differences. The difference in Mathematics was very small (Cohen's *d* = 0.03).

Girls and boys did not differ in *g*, but girls scored higher on the residual verbal factor. Might this explain the girls' better performance on GCSEs? The correlations between the residual verbal factor and GCSE subjects were all positive and of small effect size, demonstrating that verbal ability, orthogonal to *g*, contributes some additional variance to examination performance (Table 2). The correlation with overall GCSE points was .13. The only subject for which this ability had no contribution was Mathematics, where the correlation was zero. The largest correlations were in the Arts and Humanities group where correlations were around .2, and there were similar-sized correlations with History, Geography and Music.

The next step applied general linear modelling (ANCOVA) to the GCSE scores to discover the relative contributions of the predictors. Sex (male, female) was a fixed effect. The CAT's *g* and residual verbal factors were covariates. Age was also a covariate. The contributions from each of these effects to the overall GCSE scores and to each topic individually are shown in Table 3, and are expressed as partial eta squared values (proportion of variance accounted for). Age made very little contribution and is not discussed further. The *g* factor contributes 53.5% of the variance to GCSE Best 8 scores, and 49.2% of the variance to total GCSE scores. All of the contributions to individual courses are moderate to large in size. The highest contributions are to Mathematics (58.9%) and English (48.3%). The lowest contribution is to Art and Design (18.2%), and the practical subjects generally have lower values than the more traditional academic subjects, most of which have values greater than 40%. The residual verbal factor accounted for about an additional 3% of the variance in overall GCSE scores, and makes additional contributions between about 2% to 7% in the Arts and Humanities, and only 0% to less than 3% in the Sciences. Verbal ability did contribute a very small but significant amount to Physics performance, on which the sexes did not differ. The contribution to practical subjects is small, mostly around or below 1% of the variance. In the Social Sciences, History and Geography had contributions from the residual verbal factor of about an additional 4%.

Sex contributes 3.2% of the variance to overall GCSE scores and 3.7% to GCSE Best 8 scores (Table 3). The contributions of sex to the GCSE scores showed differences between subjects. The contributions to Arts and Humanities were between 3.4% and 7.4%. The contributions to Science were all very low, with only Chemistry and Biology above 1%, and Mathematics and Double science at or below 0.1%. The Social Sciences had contributions from sex at about an additional 1% to 2% of the variance. The practical subjects were mixed, with the notable results being the contributions of 5.7% to Design and Technology-Graphics and 8.0% to Art and Design. To quantify the extent to which the effects of sex were caused by girls' having better residual verbal scores, the same models were re-run for all subjects, but without the residual verbal factor in the model (results not shown). The results did produce some increase in the effect of sex but, in all cases, this is around or below 10% of the effect of sex seen in the model with the residual verbal factor. Therefore, girls' better scores on the verbal factor account for only a very small part of their better scores on GCSE examinations.

Intelligence and education: an epidemiological approach

Whether or not students gain five or more GCSE scores at grades from A* to C is important. For example, the Department for Education and Skills (DfES) use this binary variable as a key outcome in school performance tables. Such thresholds are also common for students personally, because they can be the bases of entry to further education and training. There were 39,193 students who met this criterion, and 30,646 who did not. Logistic regression was used with *g* scores as a predictor of this criterion. The exponent of β was used to obtain the odds ratio of meeting the criterion for a standard deviation increase in the *g* score. The result was that a standard deviation increase in *g* was associated with an increased odds of 7.32 ($p < .001$; 95% CI = 7.10 to 7.54). Practically, this means that, at a mean value of the *g* score, 58% achieved this criterion. On the other hand, 91% those scoring 1 SD higher on *g* achieved this, and only 16% of those scoring 1 SD lower on *g* did so. The predicted values from this univariate logistic regression model were used to conduct a Receiver Operating Characteristic curve analysis, with the criterion as the outcome (five grades at A* to C) and *g* as the predictor. The area under the curve was 0.859 ($p < .001$; 95%CI = 0.857 to 0.862). Values over 0.8 indicate good predictors.

A multivariate logistic regression was run with the same criterion as the outcome and *g*, the residual verbal factor, sex, and age as the predictor variables. The resulting odds ratios, controlling for the other variables in the model, were as follows: *g* = 8.09 ($p < .001$; 95%CI = 7.84 to 8.34); residual verbal factor = 1.43 ($p < .001$; 95%CI = 1.40 to 1.46); sex = 1.96 ($p < .001$; 95%CI = 1.88 to 2.04); age = 1.014 ($p < .001$; 95%CI = 1.009 to 1.019). 61% of girls and only 50% of boys achieved the criterion. The predicted values from this multivariate logistic regression model were used to conduct a Receiver Operating Characteristic curve analysis, with the criterion at the outcome (five grades at A* to C) and *g*, the residual verbal factor, sex and age as the predictors. The area under the curve was 0.873 ($p < .001$; 95%CI = 0.870 to 0.875), only a little greater than when *g* was used alone; but significantly so, because the 95% confidence intervals of the two models do not overlap.

Discussion

Cognitive ability tests taken at age 11 correlate 0.81 with national school examinations taken at age 16. Thus, the main finding here is the large contribution of general mental ability to educational achievement overall, and to all 25 individual subjects, though these showed substantial differences. The additional contribution of a residual verbal ability was significant, about an order of magnitude less; its contribution ranged from zero in Mathematics to 7.1% of the variance in English. Girls performed significantly better than boys on all subjects except Physics. Adjusting for *g*, there was a contribution from sex to examination performance that was about the same magnitude as the residual verbal factor. The effect of sex, too, varied markedly in magnitude across examination subjects, and was not accounted for by the better verbal ability of girls. Obtaining 5 GCSEs at grades between A* and C is an important educational outcome in England. A student with an average cognitive ability test score has a 58% chance of obtaining this; a person one standard deviation higher in ability has a 91% chance.

The data used here have a number of strengths. The sample is very large and probably representative. The cognitive test battery assesses a range of abilities in ten individual subtests. The timing of the test was optimal: at the beginning of secondary education when there has not yet been subject-specific teaching in the educational outcomes at secondary level. The prospective longitudinal nature of the study is a strength. The range of educational topics assessed is comprehensive. The examinations are national; almost all schoolchildren in England take them, and their timing is such that no children have yet left school. Thus, the cognitive ability assessments and the later educational attainments were taken when almost all children were still in school.

A possible weakness is the lack of family background data. However, as Mackintosh (1998, p. 46) argues, the association between family background and educational attainments is not high enough to account for the association between cognitive ability and education. Moreover, both twin and adoption studies consistently find that the contribution of shared environment (principally family background) is small at the childhood ages relevant to the present study (Bartels et al., 2002a; Petrill et al., 2004). There are various possible causes of the cognitive ability-educational achievement association. Bartels et al. (2002b) found a strong genetic correlation between cognitive ability (measured at 5, 7, 10 and 12 years) and educational achievement at age 12. In an overview, Petrill and Wilkerson (2000) concluded that genetics and shared and non-shared environmental factors all influence intelligence and education, with genetics being important in the correlation between them, and non-shared environment being important in discrepancies between intelligence and educational attainments.

Whereas the correlations indicate that around 50% to 60% of the variance in GCSE examination points score can be statistically explained by the prior *g* factor, by the same token a large proportion of the variance is not accounted for by *g*. Some of the remaining variance in GCSE scores will be measurement error, but some will be systematic. Thus, non-*g* factors have a substantial impact on educational attainment. These may include: school attendance and engagement; pupils' personality traits, motivation and effort; the extent of parental support; and the provision of appropriate learning experiences, teaching quality, school ethos and structure among other possible factors (Strand, 2003; Petrides et al., 2005).

There was a moderately large gender gap in educational outcomes in the present data, with different-sized effects in different school subjects. Girls have an advantage on verbal ability on the CAT (Strand, Deary & Smith, in press), and the residual verbal ability score used here did contribute to educational outcomes at about the same level as sex. However, sex differences in the educational outcomes were not accounted for by residual verbal ability scores. This accords with others' results, and it has been suggested that this is due in large part to the different classroom behaviours of boys and girls (Fergusson & Horwood, 1997). It is also possible that girls have a type of verbal advantage in GCSE that is not captured by CAT verbal subtests. Such abilities might include verbal fluency and/or the ability to express thoughts in connected prose, and/or better memory for information that was presented in verbal form (orally, by teachers or by reading textbooks). GCSE public examinations rely heavily on essays and other modes of assessment requiring extended writing. We also know that largest sex differences reflect girls' superiority in the area of writing. For example, Cole (1997) reported an analysis of multiple US national datasets for school students assessed at age 9, 14 and 17 on a wide variety of tests. There was no sex difference on Verbal reasoning/vocabulary (d = 0.05), a small advantage in verbal-reading (d = 0.20) a medium female advantage in verbal-language use (d = 0.40), and the largest difference for verbal-writing (d = 0.60). It is also possible that sex differences in GCSE reflect wider factors related to motivation and effort, such as girls' greater likelihood to complete and submit coursework (OHMCI, 1997), gendered patterns of subject choice (Arnot et al, 1996), and/or gendered allocation to tiered subjects (Elwood, 1995). Salisbury, Rees & Gorard (1999) provide a review of the literature in this area.

Spearman (1904) suggested that the general factor from school subjects would be almost perfectly correlated with general intelligence. In fact, the correlation is above 0.8, even when the cognitive ability measurements were taken five years before the school exams. Beyond that, in English school examinations there is further advantage to girls and to those with relatively strong verbal skills (after g is controlled). These data establish the validity of g for this important life outcome.

Acknowledgement—Ian Deary is the recipient of a Royal Society-Wolfson Research Merit Award. A version of this paper was presented by Ian Deary as the Spearman Centenary Lecture to the International Society for Intelligence Research in New Orleans, USA, in December 2004.

References

Arnot, J., David, M., & Weiner, G. (1996). *Educational reform and gender equality in schools*. Manchester: Equal Opportunities Commission.

Autumn Package. (2002). The Autumn Package. Accessed July 20, 2005 at: www.standards.dfes.gov.uk/performance.

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D. I. (2002a). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, *32*, 237-249.

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

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software, Inc.

Binet, A. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Annee Psychologique*, *12*, 191-244. (Translated in 1916 by E. S. Kite in *The Development of Intelligence in Children*. Vineland, NJ: Publications of the Training School at Vineland.)

Brody, N. (1992). *Intelligence*, 2nd edition. New York: Academic.

Cole, N. S. (1997). The ETS Gender study: How females and males perform in educational settings. Princeton, NJ: Education Testing Service.

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.

Elwood, J. (1995). Undermining gender stereotypes: examination and coursework performance in the UK at 16. *Assessment in Education*, *2*, 283-303.

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*, 83-96.

Jencks, C. (1979). *Who gets ahead?*: *The determinants of economic success in America*. New York, NY: Basic Books.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. New York: Praeger.

Mackintosh, N. J. (1998). *IQ and Human Intelligence*. Oxford: Oxford University Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*, 77-101.

OHMCI. (1997). *The relative performance of boys and girls.* Cardiff: Office of Her Majesty's Chief Inspector.

Petrides, K. V., Chamorro-Premuzic, T., Frederickson, N., & Furnham, A. (2005). Explaining individual differences in scholastic behaviour and achievement. *British Journal of Educational Psychology*, *75*, 239-255.

Petrill, S. A., Lipton, P. A., Hewitt, J. K., Plomin, R., Cherny, S. S., Corley, R., & DeFries, J. C. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental Psychology*, *40*, 805-812.

Petrill, S. A., & Wilkerson, B. (2000). Intelligence and achievement: a behavioral genetic perspective. *Educational Psychology Review*, *12*, 185-199.

Rinderman, H. & Neubauer, A. C. (2004). Processing speed, intelligence, creativity, and school performance: testing of causal hypotheses using structural equation models. *Intelligence*, *32*, 573-589.

Salisbury, J., Rees, G., & Gorard, S. (1999). Accounting for the differential attainment of boys and girls at school. *School Leadership & Management*, *19*, 403-426.

Sax, G. (1984). The Lorge-Thorndike Intelligence Tests/Cognitive Abilities Test. In Keysner, D. J., & Sweetland, R. C. (eds), *Test Critiques* (Volume 1, pp. 421-435). Kansas City, MO: Test Corporation of America.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Smith, P., Fernandes, C., & Strand, S. (2001). *Cognitive Abilities Test Third Edition: Technical Manual*. London: nferNelson.

Spearman, C. (1904). “General Intelligence” objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

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

Strand, S. (2003). *Getting the best from CAT: A practical guide for secondary schools*. London: nferNelson.

Strand, S. (2004). Consistency in reasoning scores over time. *British Journal of Educational Psychology, 74*, 617-631.

Strand, S. (in press). Comparing the predictive validity of reasoning tests and national end of Key Stage 2 tests: Which tests are the ‘best’? *British Educational Research Journal.*

Strand, S., Deary, I. J., & Smith, P. (in press). Sex Differences in Cognitive Abilities Test Scores: A UK National Picture. *British Journal of Educational Psychology*.

Thorndike, R. L., Hagen, E., & France, N. (1986). *Cognitive Abilities Test Second Edition: Administration Manual*. Windsor: nferNelson.

Zenderland, L. (1998). *Measuring Minds: Henry Herbert Goddard and the Origins of American Intelligence Testing*. Cambridge, UK: Cambridge University Press.

Table 1

Comparison of the matched sample against the population average for GCSE\GNVQ outcomes

Measure	CAT-matched sample	Full national dataset
GCSE\GNVQ 'Best 8' points: Mean (SD)	37.3 (13.4)	36.8 (14.0)
English: % entries graded A*-C	62%	61%
Maths: % entries graded A*-C	54%	53%
Science: % entries graded A*-C	54%	53%
Sample size	80,074	361,335

Table 2

Sex comparisons in cognitive ability and GCSE scores, and correlations between general cognitive ability (*g*) and the residual verbal factor and GCSE scores.

Cognitive Ability Test or GCSE Subject	Male-female comparisons			Correlations	
	Boys' mean (SD, N)	Girls' mean (SD, N)	*p* for difference (Cohen's *d*)	CAT *g*	CAT residual verbal[b]
Cognitive Abilities Test					
g factor[a]	-.007 (.965, 34,850)	.004 (.912, 35,680)	.096 (.01)	-	
Residual verbal factor[a]	-.126 (1.026, 34,850)	.122 (.958, 35,680)	<.001 (.25)	.00	-
Overall Score					
GCSE total points	39.5 (18.8, 36,894)	45.1 (18.2, 37,509)	<.001 (.30)	.69 (70,530)	.13
GCSE best 8	35.4 (13.5, 35,848)	39.4 (12.8, 36,759)	<.001 (.30)	.72 (68,904)	.14
Arts & Humanities					
English	4.52 (1.53, 34,947)	5.13 (1.41, 36,328)	<.001 (.41)	.67 (67,677)	.22
English literature	4.60 (1.61, 31,316)	5.22 (1.44, 34,317)	<.001 (.41)	.59 (62,416)	.20
Drama	4.97 (1.52, 4,022)	5.53 (1.38, 7,537)	<.001 (.39)	.47 (10,997)	.14
Religious Education	4.16 (2.07,	4.96 (1.93,	<.001 (.40)	.52 (13,572)	.16

	5,887)	8,211)			
French	4.04 (1.69, 17,876)	4.74 (1.69, 20,213)	<.001 (.41)	.64 (36,370)	.20
German	4.28 (1.62, 8,135)	4.93 (1.60, 9,491)	<.001 (.40)	.61 (16,638)	.18
Spanish	3.92 (1.82, 2,734)	4.71 (1.76, 3,983)	<.001 (.44)	.62 (6,501)	.18
Science					
Mathematics	4.48 (1.73, 35,371)	4.54 (1.71, 36,390)	<.001 (.03)	.77 (68,125)	.00
Double Science	4.52 (1.60, 30,831)	4.63 (1.61, 31,918)	<.001 (.07)	.68 (59,518)	.12
Single Science	2.95 (1.45, 2,661)	3.30 (1.48, 2928)	<.001 (.24)	.60 (5,331)	.02
Physics	5.83 (1.35, 1,555)	5.78 (1.39, 1,268)	.33 (-.04)	.50 (2,733)	.09
Chemistry	5.61 (1.33, 1,539)	5.92 (1.26, 1,272)	<.001 (.24)	.46 (2,720)	.08
Biology	5.84 (1.22, 1,568)	6.07 (1.21, 1,292)	<.001 (.19)	.51 (2,764)	.14
Social Sciences					
Geography	4.55 (1.77, 15,014)	4.88 (1.80, 12,430)	<.001 (.18)	.65 (26,081)	.16
History	4.62 (1.95, 11,697)	5.02 (1.91, 12,220)	<.001 (.21)	.63 (22,764)	.18
Business	4.49 (1.77, 6,591)	4.72 (1.74, 5,184)	<.001 (.13)	.56 (11,188)	.11
Information technology	4.43 (1.87, 5,747)	4.79 (1.83, 3,840)	<.001 (.19)	.47 (9,350)	.07

Information Technology short course	3.79 (1.88, 4,144)	4.39 (1.90, 5,217)	<.001 (.32)	.48 (8,931)	.11
Practical					
Art & Design	4.59 (1.62, 6,486)	5.54 (1.49, 9,397)	<.001 (.61)	.43 (15,104)	.09
Music	5.11 (1.97, 2,221)	5.48 (1.70, 3,224)	<.001 (.20)	.54 (5,208)	.16
Physical Education	4.87 (1.53, 9,802)	5.03 (1.65, 4,696)	<.001 (.10)	.55 (13,846)	.07
DT-Food	4.14 (1.68, 3,605)	4.84 (1.65, 10,626)	<.001 (.42)	.52 (13,493)	.08
DT-Graphics	4.24 (1.80, 8,387)	5.06 (1.67, 6,745)	<.001 (.47)	.45 (14,328)	.07
DT-Resistant Materials	4.18 (1.67, 11,361)	4.88 (1.69, 3,499)	<.001 (.42)	.48 (14,059)	.06
DT-Textiles	3.68 (1.86, 201)	5.00 (1.65, 6,557)	<.001 (.75)	.52 (6,390)	.09

Note. [a]These variables are in standard units (mean = 0; SD = 1). [b]This is the residual of the CAT verbal score after regression with the CAT *g* factor entered as the independent variable; Ns for this column are the same as those for the column immediately to the right.

Table 3
General linear modelling of GCSE scores. Values shown are eta squared for each fixed effect and covariate in the model.

Subject	Eta squared values from general linear model				N
	CAT *g*	CAT residual verbal[b]	Age	Sex	
Overall Score					
GCSE total points	.492	.027	.000	.032	70,530
GCSE best 8	.535	.032	.001	.037	68,904
Arts & Humanities					
English	.483	.072	.001	.065	67,677
English literature	.383	.049	.001	.060	62,416
Drama	.226	.018	.001	.034	10,997
Religious Education	.303	.032	.000[a]	.065	13,572
French	.448	.053	.000	.074	36,370
German	.402	.040	.000[a]	.062	16,638
Spanish	.413	.040	.001[a]	.073	6,501
Science					
Mathematics	.589	.000	.000[a]	.001	68,125
Double Science	.465	.026	.000	.000	59,518
Single Science	.361	.017	.000[a]	.004	5,331
Physics	.244	.005	.000[a]	.000[a]	2,733
Chemistry	.215	.002	.000[a]	.019	2,720
Biology	.264	.013	.000[a]	.012	2,764
Social Sciences					
Geography	.443	.039	.001	.011	26,081
History	.406	.041	.001	.014	22,764
Business	.336	.026	.000[a]	.011	11,188

Information Technology	.228	.005	.000[a]	.016	9,350
Information Technology short course	.234	.009	.000[a]	.030	8,931
Practical					
Art & Design	.182	.001	.001	.080	15,104
Music	.288	.016	.001[a]	.008	5,208
Physical Education	.303	.013	.005	.000[a]	13,846
DT-Food	.294	.015	.000[a]	.045	13,493
DT-Graphics	.211	.001	.000[a]	.057	14,328
DT-Resistant Materials	.223	.004	.001	.021	14,059
DT-Textiles	.263	.008	.001[a]	.012	6,390

Note. [a]These effects have p values > .01 (even in some cases where the eta squared value is signified as .000). All other effects have p values ≤ .01. [b]This is the residual of the CAT verbal score after regression with the CAT g factor entered as the independent variable.

Figure 1

Structural equation model (confirmatory factor analysis) to examine the correlation between latent traits of intelligence and educational achievement.