Available online at www.sciencedirect.com
SCIENCE DIRECT®
Intelligence 31 (2003) 533–542
INTELLIGENCE

Population sex differences in IQ at age 11: the Scottish mental survey 1932

Ian J. Deary[a,*], Graham Thorpe[b], Valerie Wilson[b], John M. Starr[c], Lawrence J. Whalley[d]

[a]*Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland, UK*
[b]*Scottish Council for Research in Education, Edinburgh, Scotland, UK*
[c]*Department of Geriatric Medicine, University of Edinburgh, Edinburgh, Scotland, UK*
[d]*Department of Mental Health, University of Aberdeen, Scotland, Aberdeen, Scotland, UK*

Received 9 October 2002; received in revised form 18 March 2003; accepted 8 April 2003

Abstract

There is uncertainty whether the sexes differ with respect to their mean levels and variabilities in mental ability test scores. Here we describe the cognitive ability distribution in 80,000+ children—almost everyone born in Scotland in 1921—tested at age 11 in 1932. There were no significant mean differences in cognitive test scores between boys and girls, but there was a highly significant difference in their standard deviations ($P < .001$). Boys were over-represented at the low and high extremes of cognitive ability. These findings, the first to be presented from a whole population, might in part explain such cognitive outcomes as the slight excess of men achieving first class university degrees, and the excess of males with learning difficulties.

Keywords: Population sex differences; Scottish mental survey; Mental ability

1. Introduction

"There is a tendency for men to be 'more so' than females, whatever is being tested. Thus on intelligence tests, for instance, when groups of comparable men and women take tests,

* Corresponding author. Tel.: +44-131-650-3452; fax: +44-131-651-1771.
E-mail address: I.Deary@ed.ac.uk (I.J. Deary).

0160-2896/03/$ – see front matter
doi:10.1016/S0160-2896(03)00053-9

> they tend to gain mean scores which are similar, but the highest and lowest scorers are liable to be male. The finding is not confined to intelligence tests... There is a tendency for women students to gain proportionately more second class degrees—and, thus, fewer first and thirds—in many examination subjects... In the 'real world' situation, the same tendency holds: men rather than women are found at the extremes. There are more male geniuses, more male criminals, more male mental defectives" (Heim, 1970, pp. 136–137).

The interest in possible sex differences in IQ scores arises because low cognitive ability is a correlate of dependence within society and high cognitive ability is a correlate of achievements that are valued by society (Gottfredson, 1997; Lubinski, Webb, Morelock, & Benbow, 2001). There is unsettled debate about the cognitive differences of the sexes: whether their means or distributions differ, and whether there are sex differences in general and/or more specific cognitive skills (Mackintosh, 1996, 1998a, 1998b, pp. 182–188; Hedges and Nowell, 1995). Mackintosh's (1998a) resumé of the literature concludes that the sexes do not "differ greatly" in average IQ, but more so in some specific abilities. He concluded that the suggestion of males' greater cognitive variability was not proven. This suggestion has a long history. For example, Vernon (1940; Fig. 4 on p. 13 and Fig. 12 on p. 18) displayed data on 136 men and 114 women, tested on the Cattell Intelligence Test IIIA, which showed no sex difference in mean scores but more dispersed scores among men. Reports of the well-established finding within the United States that males show greater variability in some abilities (e.g., mathematics; Hedges & Nowell, 1995; Lubinski & Humphries, 1990) are not consistently found in other countries (Feingold, 1994). Jensen's (1998, pp. 531–543) overview of large cognitive test batteries in which there were large samples of the population found a near-zero sex difference in the general cognitive factor means. However, Mackintosh (1998b) agreed with Lynn (1998) that males score higher on the Wechsler tests.

The largest source of empirical evidence to date on sex differences in mental test scores is the reanalysis of six national U.S. studies conducted by Hedges and Nowell (1995). The studies involved people from age 15 to the early twenties. Sex differences in mean test scores were typically small, though the large subject numbers involved meant that most comparisons were significant at traditional P values. Female subjects scored better on reading comprehension, perceptual speed, and associative memory. Male subjects scored better on mathematics and vocational aptitude scales such as mechanical reasoning. In general, male subject participants had larger test score variances than female subject participants. With the exception of those test types on which female participants had higher mean scores, there tended to be more male participants in the highest scoring categories. This influential study had some limitations. None of the six participants is a complete population. Many of the tests were vocational aptitude tests or were related to school subjects. The ages of the participants tested meant that some samples involved people at stages where there would be dropout from universal education and there would be subject choice taking place with possible channelling of male and female subjects to different course choices. Therefore, issues of sampling and different male–female educational experiences might affect the mean and variance differences found.

In the present study we examine sex differences in IQ test scores in a whole population. Most previous studies of sex differences in cognitive abilities have sampling problems, as discussed by Jensen (1998) and Hedges and Nowell (1995). Samples are often relatively small or highly selected, and therefore could be biased: their selection might be a function of confounders of ability levels and spread. Here we compare the means and variabilities of boys and girls on a valid cognitive ability test taken at age 11 years by almost everyone in Scotland born in 1921 in the Scottish Mental Survey 1932 (SMS1932). The mental test was designed by Professor Sir Godfrey Thomson, who also chaired the SMS1932's statistical committee. The SMS1932 noted a larger standard deviation of test scores among boys, but any more detailed examination of sex differences in IQ variability was prohibited because analyses were performed by hand (Scottish Council for Research in Education (SCRE), 1933, pp. 103–108). The present analysis has the following advantages over existing reports: an entire population was tested; at the time of testing everyone was still in education and there was no curriculum choice; and the test used was not aimed at educational attainment or vocational aptitude.

2. Participants and method

The SMS1932 was conducted under the auspices of the Scottish Council for Research in Education (Deary, Whalley, Lemmon, Crawford, & Starr, 2000; SCRE, 1933). It aimed to quantify the numbers of people in Scotland who were "mentally deficient." Its remit was broadened to "obtain data about the whole distribution of the intelligence of Scottish pupils from one end of the scale to the other" (SCRE, 1933). It took place on June 1st, 1932 (SCRE, 1933). A small amount of testing took place a few days later. The SMS1932 team planned to administer a cognitive ability test to everyone in the Scottish nation born in 1921. Deaf and blind children, and those absent from school on the testing day, were not tested. Other exceptions were a few private schools and some schools that received inadequate numbers of tests. The total number tested in the SMS1932 was 87,498, over 95% of the relevant population. Three tests were administered to each pupil. The first two tests were pictorial in nature, designed for illiterate pupils.

2.1. First Picture Test

This test is contained on a single page. At the top are line drawings of a door, table, chair, wheelbarrow, and a garden-roller. Under each, respectively, are the numbers 1, 2, 3, 4, and 5. Below this, there are 40 further line drawings of these objects in five rows of eight objects each. The pupil's task was to write the correct number under each object. The first four were used as practice. Thereafter, the pupils were given 1 min to complete as many as possible. Most pupils were expected to score perfectly on this test. The task is similar to the processing involved in the Wechsler Digit Symbol test, but with different stimuli.

2.2. Second Picture Test

This test, with nine items, is also contained on a single page. Each item has three line drawings to the left of a vertical line, and five line drawings on the right. For example, the first item, used for practice on the test, contains drawings of snowdrops, daffodils, and crocuses on the left of the vertical line, and a bottle, a cup, a rose, a brush, and scissors to the right. The pupil's task was to find the item on the right hand side of the line "that is most like the first three." This is a classification task. After completing the first two items, pupils were given 2 min to complete the remaining seven items. Like the First Picture Test, it was intended to be easy.

2.3. Moray House Test (MHT)

We described the Moray House Test—the main test used in the SMS1932—previously (Deary et al., 2000). In the original SCRE (1933) publications it was always referred to as the "Verbal Test," because it required literacy and numeracy to understand and complete the items. The test was closely related to the Moray House Test No. 12 that was used in secondary school selection in England. It had a total possible score of 76 and was administered with a time limit of 45 min. Teachers read precise instructions, provided by SCRE, which may still be consulted. There were eight practice items. The test comprises a variety of item types, as follows: following directions (14 items), same–opposites (11), word classification (10), analogies (8), practical items (6), reasoning (5), proverbs (4), arithmetic (4), spatial items (4), mixed sentences (3), cypher decoding (2), and other items (4). The test has 71 numbered items, 75 items in total, and the maximum possible score is 76.

Together, the First and Second Picture Tests and the Moray House Test were collectively known as the "Group Test." In the description of compiling and pilot testing the Group Test there is no mention of selecting items to minimise differences between the sexes (SCRE, 1933, pp. 55–88).

The scores on the Moray House Test in 1932 were criterion validated by SCRE's individually retesting a representative sample of 1000 of the children (500 boys, 500 girls) on the Stanford Revision of the Binet–Simon scale during the Summer of 1932. The coefficients of validation were .81 for boys and .78 for girls (SCRE, 1933, p. 100). The MHT shows a stability coefficient of .63 to .65 (above .7 when restriction of range is taken into account) between age 11 and age about 80 ($n = 101$ and $n = 493$; Deary et al., 2000; Deary, Whiteman, Starr, & Whalley, in press). At age 80, the criterion validity coefficient between MHT and Raven's Progressive Matrices is greater than .7 ($n = 543$; Deary et al., in press).

2.4. Data processing and statistical analyses

The SMS1932 data were recorded in hand-written ledgers. The ledgers from the Fife, Angus, and Wigtown areas of Scotland are absent. For this project, the available data were transferred to computer files using Microsoft Excel and thereafter to SPSS version 11. Data copied were name (removed from subsequent records), school, region, date of birth, and scores on the First and Second Picture Tests and the Moray House Test. All data were entered twice, and the two records were compared for identity. Any mismatches were compared with the original, handwritten ledgers. This resulted in records for 86,520 participants. Of these, 81,182 had a score recorded on the First Picture Test, 81,118 had a score on the Second Picture Test, and 81,140 had a score on the Moray House Test.

Scores on the First and Second Picture Tests and the Moray House Test were summed. This total score was converted to an IQ-type score with a mean ± S.D. of 100 ± 15. In addition, these were coded to allocate people to the following IQ bands: 50 to < 60, 60 to < 65, and thereafter in bands of 5 up to a top band of 130 to 140. For the present study the following exclusions applied to subjects. Only those with nonzero scores on the First and Second Picture Tests were selected. Five people with incorrectly recorded scores on the First Picture Test (scores>40) and three on the second Picture Test (scores >9) were omitted. This left 79,376 (39,343 girls and 40,033 boys) with total score information.

3. Results

The mean IQ score, based on the total score of the Picture and Moray House Tests, was 100.64 for girls and 100.48 for boys. This difference of 0.16 (95% CI = – 0.037 to 0.367) was nonsignificant, despite the massive numbers tested, $t(79,374) = 1.6$, $P = .11$. The standard deviation was 14.1 for girls and 14.9 for boys. Levene's test for comparing variances was significant, $P < .001$, i.e., boys and girls differed significantly in variability.

Participants were allocated to IQ score bands as described above. The absolute number and the relative percentages of boys and girls in each IQ band is shown in Fig. 1. The proportions

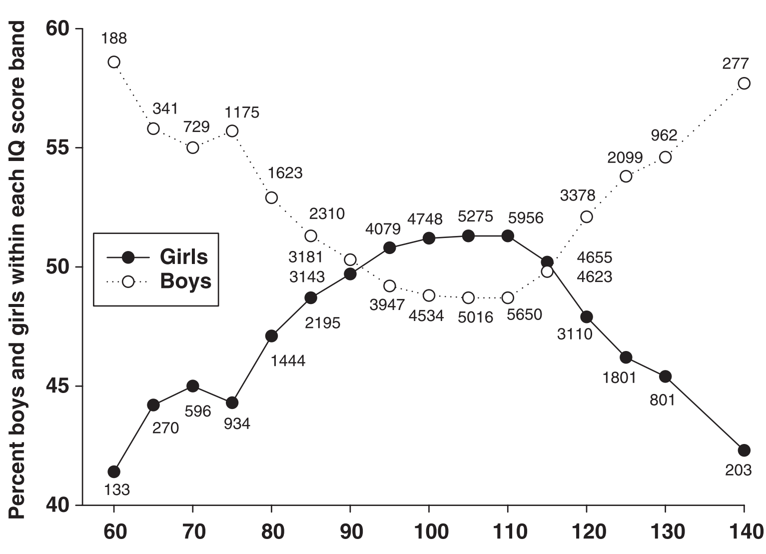

Fig. 1. Numbers and percentages of boys and girls found within each IQ score band of the Scottish population born in 1921 and tested in the Scottish Mental Survey in 1932 at age 11. The *y* axis represents the percentage of each sex in each 5-point band of IQ scores. Numbers beside each point represent the absolute numbers of boys and girls in each 5-point IQ score band.

of girls and boys in each band were significantly different, $\chi^2(15) = 147.9$, $P < .001$. This is not just a reflection of different numbers of boys and girls in the population: girls represent 49.6%, and boys 50.4%, of the participants providing these data. In the IQ bands that cover the range 90 to < 115, girls are found in slight excess, a difference of about 2%. At the extremes, boys are over-represented. In the IQ band from 50 to < 60 boys make up 58.6% of the population, a gender gap of 17.2%. In the IQ band 130 to < 140 boys make up 57.7% of the population, a gender gap of 15.4%. The gradation between the extremes appears regular: as the population moves away from the extremes the sex difference in proportions steadily lessens with, eventually, a slight excess of girls in the average score range.

4. Discussion

The SMS1932 data provide the most comprehensive test to date of the possibility of sex differences in IQ at age 11. There was no difference in the mean, despite there being about 40,000 children in each sex. The standard deviation was significantly different between boys and girls. There was a slight excess of girls in the IQ 90–115 region. There was a more marked excess of boys at the extremes, with approximately 1.4 boys to every girl in the IQ 50 to < 60 and 130 to < 140 bands. We are not aware of any selection bias among the small percentage of children not appearing for the SMS1932 that might have reduced the number of girls at the extremes and inflated their numbers in the middle range of scores.

The possibility that the construction of the test itself introduces the observed larger spread of IQ for boys was considered. The key issues in 1932 were to construct a test of good face validity for assessing average intelligence. While the potential for sex bias in the detail of the test was recognised, this was in a muted way, and does not appear to have significantly influenced judgement about which test items to include and which not (SCRE, 1933). To suppose that item bias gives rise to the effect seen here does, however, raise the question of how, and that is less easily seen. Had it simply been a question of one sex scoring uniformly higher than the other, then the presence of a group of items biased in favour of that sex could be suspected. A more complex, and consequently more unlikely, form of bias is required to explain the pattern actually found. No simple one is readily apparent. One candidate, for example, would be if a group of items of otherwise high facility were more difficult for boys than for girls whereas, conversely, a group of low facility items were easier for boys. This is quite an involved scenario and not likely to have occurred in practice, although equally unlikely to have been noticed if it had. There are item-level data for 500 boys and 500 girls (SCRE, 1933, pp. 87–88). Analyses of these data do not, though, reveal any bias that might have created the observed sex difference. Whereas there are some sex differences in item facilities, they are balanced between the sexes and show no link to particular item content-types or item facility at all.

Any practical implications of this difference in IQ spread between the sexes are dependent upon the reliability and predictive validity of such test scores. The Moray House Test's long-term stability and criterion validity were described above. The Moray House Test is a significant predictor of longevity (Whalley & Deary, 2001). Mental tests in general were used widely because of their success in discriminating those with learning difficulties from those without (Zenderland, 1998). Psychometric intelligence tests have validity as moderately strong predictors of educational and occupational success (Neisser et al., 1996; Schmidt & Hunter, 1998; Smith, Fernandes, & Strand, 2001, pp. 42–49). People identified as having exceptionally high cognitive ability at age 13 are more than 50 times as likely as the general population to take doctoral-level degrees (Lubinski et al., 2001). Therefore, it is possible that an excess of one population group with extreme IQ scores might in part explain sex differences in real-life attainments.

The sex difference in the lower ability range reflects the greater numbers of boys than girls with low educational attainments; some of these boys would have met contemporary criteria for learning difficulties. This excess of young boys is consistent with sex differences reported in numerous epidemiological surveys (Rutter et al., 1990), though the sources of such differences are disputed. Mackintosh (1998b) attributes the excess of lower-ability boys to their specific problems with reading. Recurring explanations in the clinical, sociological, and educational literature include the failure of more young boys than girls to acquire internal social controls, to become disruptive in the classroom, and to fail to cooperate with learning and assessment procedures. Such "conduct disordered" boys would have been expected to perform poorly in the SMS1932. This type of account does not satisfactorily explain the origins of sex differences in poor performance (it simply replaces one association with another), nor does it explain why boys also perform so much better at the higher range of ability. Thus, although the variability in Fig. 1 looks systematic that does not mean there is likely be a single cause operating at all ability levels. Heim (1970) outlined possible explanations for the excess of men at the highest IQ scores that ranged from biological to social. For example, the excess of high ability boys might be a result of sex-based socialisation differences in the cultivation of the human repertoire of problem solving abilities (Sternberg, 1985). Boys may have been given more opportunities to develop high-level reasoning skills.

The differences in the proportions of boys and girls at the higher end of the SMS1932 scores might in part explain any excess high-level cognitive achievements in men. A quantifiable and consistent example is that men obtain slightly more first class university degrees. Alice Heim's high-level tests of intelligence—the AH5 and AH6—showed greater standard deviations in male students (Mellanby, Martin, and O'Doherty, 2000; Watts, 1953), and it was one of her research team (Watts, 1953, pp. 355–356) who suggested there might be a link between the difference in IQ variability between men and women and the fact that men gain more first- and third-, but fewer second-, class degrees than women. University degree class is relevant in this context because it is one of the few quantifiable and systematically and comprehensively recorded (HESA, 2002; Smith and Naylor, 2001a, 2001b) life outcomes with which high cognitive test scores might be associated. There is research interest in attempting to explain the patterning of sex differences in degree classes among men in the UK, especially at the elite universities of Oxford and Cambridge where men obtain up to 60% more first class degrees than women (Leman, 1999; McCrum, 1994; Smith and Naylor, 2001a, 2001b; Surtees, Wainwright, and Pharoah, 2002).

The Higher Education Statistics Agency (http://www.hesa.ac.uk/holisdocs/pubinfo/stud.htm) has collected data from all UK universities since 1994. More women graduate with university undergraduate degrees than men: the proportion increased steadily from 1.05:1 in 1994–1995 to 1.25:1 in 2000–2001. Women perform better in degrees overall: over the 7 years of HESA statistics, a mean of 48.5% of women obtained upper second class degrees, whereas only 39.5% of men did. Men are more likely to achieve third class or pass degrees: in the data from 1996–1997 until 2000–2001 the mean percentage of men was 9.47, and of women was 5.10. Against this consistent evidence of women's being substantially more successful overall in higher education, men exceed women among the small number of people who obtain first class degrees. From 1994–2001, a mean of 8.20% of male graduates obtained a first class degree, whereas 6.99% of women did. Thus, whereas the proportion of men to women is 1:1.23 for upper second class degrees and 1.86:1 for third class degrees, that for first class degrees is 1.17 to 1. The sex differences in degree classes are very similar for all the 7 years of data, and are not accounted for by observed variables such as degree subject choice or social class (Smith & Naylor, 2001a). At the more academically elite universities, the ratio of men to women for first class degrees is higher (Surtees et al., 2002).

The SMS1932 data apply to people whose main educational and occupational attainments occurred from just before the Second World War until retirement in the mid-1980s. Therefore, the distributions of IQ scores represent this specific cohort without necessarily generalising to prior or subsequent cohorts. However, it is notable that the sex differences in patterning of degree classes in UK Universities might have been similar in the period 1920–1950 (Watts, 1953) to what it was during the 1990s (HESA, 2002). The data here apply to a particular mental test, and not to abilities in which the sexes, especially at older ages, are known more clearly to differ Mackintosh (1998b). Nevertheless, the MHT has proven to be a strong correlate of verbal and nonverbal abilities in later life (Crawford, Deary, Starr, & Whalley, 2001; Deary et al., in press). The SMS1932 data apply to childhood and so do not address the debates concerning sex differences in abilities in later adolescence, adulthood, and old age (Hedges & Nowell, 1995). For example, Lynn (1994, 1998) suggests that any male–female mean difference becomes more marked after age 16, though Mackintosh (1996, 1998a, 1998b) and Jensen (1998) disagree that such a mean difference exists at all. There is considerable uncertainty about whether men or women fare better with respect to cognitive ageing (Meinz & Salthouse, 1998).

It will be interesting to develop these findings: to discover whether this sex difference in a nation's intellectual capital is maintained in later year-of-birth cohorts; to explore further any practical implications of these subtle sex differences in mental ability; and to seek their causes in social and biological factors.

Acknowledgements

This research was supported by grants from the Scottish Executive Health Department Chief Scientist Office and the Biotechnology and Biological Sciences Research Council SAGE initiative. Ian Deary is the recipient of a Royal Society–Wolfson Research Merit Award. Lawrence Whalley holds a Wellcome Trust Career Development Award.

References

Crawford, J. R., Deary, I. J., Starr, J., & Whalley, L. J. (2001). The NART as an index of prior intellectual functioning: A retrospective validity study covering a 66 year interval. *Psychological Medicine*, *31*, 451–458.

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*, 49–55.

Deary, I. J., Whiteman, M. C., Starr, J. M., & Whalley, L. J. (in press). The impact of childhood intelligence in later life. *Journal of Personality and Social Psychology*.

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles*, *30*, 81–92.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45.

Heim, A. (1970). *Intelligence and personality*. Harmondsworth, UK: Penguin.

HESA (2002). http://www.hesa.ac.uk/holisdocs/pubinfo/stud.htm.

Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.

Leman, P. J. (1999). The role of subject area, gender, ethnicity and school background in the degree results of Cambridge University undergraduates. *Curriculum Journal*, *10*, 231–252.

Lubinski, D., & Humphries, L. G. (1990). A broadly based analysis of mathematical giftedness. *Intelligence*, *14*, 327–355.

Lubinski, D., Webb, R. M., Morelock, M. J., & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology*, *86*, 718–729.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, *17*, 257–271.

Lynn, R. (1998). Sex differences in intelligence: A rejoinder to Mackintosh. *Journal of Biosocial Science*, *30*, 529–532.

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, *28*, 559–571.

Mackintosh, N. J. (1998a). Reply to Lynn. *Journal of Biosocial Science*, *30*, 533–539.

Mackintosh, N. J. (1998b). *IQ and human intelligence*. Oxford: Oxford University Press.

McCrum, N. G. (1994). The academic gender deficit at Oxford and Cambridge. *Oxford Review of Education*, *22*, 369–397.

Meinz, E. J., & Salthouse, T. A. (1998). Is age kinder to females than to males? *Psychonomic Bulletin & Review*, *5*, 56–70.

Mellanby, J., Martin, M., & O'Doherty, J. (2000). The 'gender gap' in final examination results at Oxford University. *British Journal of Psychology*, *91*, 377–390.

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.

Rutter, M., MacDonald, H., Lecouteur, A., Harrington, R., Bolton, P., & Bailey, A. (1990). Genetic factors in child psychiatric disorders, 2: Empirical findings. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *31*, 39–83.

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.

Scottish Council for Research in Education (1933). *The intelligence of Scottish children: A national survey of an age-group*. London: University of London Press.

Smith, J., & Naylor, R. (2001a). Determinants of degree performance in UK universities: A statistical analysis of the 1993 student cohort. *Oxford Bulletin of Economics and Statistics*, *63*, 29–60.

Smith, J., & Naylor, R. (2001b). Dropping out of university: A statistical analysis of the probability of withdrawal for UK university students. *Journal of the Royal Statistical Society A*, *164*, 389–405.

Smith, P., Fenandes, C., & Strand, S. (2001). *Cognitive abilities test 3: Technical manual*. Windsor, UK: NFER-Nelson.

Sternberg, R. J. (1985). *Beyond IQ*. New York: Cambridge University Press.

Surtees, P. G., Wainwright, N. W. J., & Pharoah, P. D. P. (2002). Psychosocial factors and sex differences in high academic attainment at Cambridge University. *Oxford Review of Education*, *28*, 21–38.

Vernon, P. E. (1940). *The measurement of abilities*. London: University of London Press.

Watts, K. P. (1953). Influences affecting the results of a test of high grade intelligence. *British Journal of Psychology*, *44*, 359–367.

Whalley, L. J., & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal*, *322*, 1–5.

Zenderland, L. (1998). *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing*. Cambridge, UK: Cambridge University Press.