

The Stability of Individual Differences in Mental Ability from Childhood to Old

Age: Follow-up of the 1932 Scottish Mental Survey

Ian J. Deary¹, Lawrence J. Whalley², J. R. Crawford³, John M. Starr⁴

¹Department of Psychology, University of Edinburgh

²Department of Mental Health, University of Aberdeen

³Department of Psychology, University of Aberdeen

⁴Department of Geriatric Medicine, Royal Victoria Hospital, Edinburgh

Address for Correspondence,

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

Telephone +44 131 650 3452

Facsimile +44 131 650 3461

Email I.Deary@ed.ac.uk

The Stability of Individual Differences in Mental Ability from Childhood to Old**Age: Follow-up of the 1932 Scottish Mental Survey****Summary**

All Scottish children born in 1921 and attending school on June 1st 1932 ($N = 87,498$) undertook a validated test of psychometric intelligence, The Moray House Test. We followed up 101 of these people at age 77 and re-administered the same mental ability test. Concurrent validity data are provided for the Moray House Test at age 11 ($n = 1000$) and age 77 years ($n = 97$). The correlation between Moray House Test scores at age 11 and age 77 was 0.63, which adjusted to 0.73 when corrected for attenuation of ability range within the re-tested sample. This, the longest follow-up study of individual differences in human intelligence reported to date, shows that mental abilities are highly stable from childhood to late life.

Introduction

The stability of individual differences in human mental abilities is of scientific and popular interest (Jensen, 1980). In childhood it is of interest to discover whether educational initiatives can boost ability levels and whether environmental insults—such as poor nutrition or lead pollution—lower cognitive functions. In old age there is intense interest in whether mental ability differences earlier in life contribute to the risk of dementia and other syndromes of cognitive decline (Snowdon, Kemper, Mortimer, Greiner, Wekstein, & Markesbery, 1996).

There are reports of the stability of measures of mental ability: (a) within childhood (Humphreys, 1989); (b) from childhood to mid-adulthood (Kangas, & Bradway, 1971); (c) across young- to mid-adulthood (Eichorn, Hunt, & Honzik, 1981; Nisbet, 1957; Owens, 1966; Plassman, Welsh, Helms, Brandt, Page, & Breitner, 1995; Schwartzman, Gold, Andres, Arbuckle, & Chaikelson, 1987; Tuddenham, Blumenkrantz, & Wilkin, 1968); and (d) in old age (Mortensen, & Kleven, 1993). Table 1 summarises these studies, showing the duration across which stability of individual differences was assessed and the stability coefficients obtained. The studies collected in Table 1 show that intellectual ability differences become increasingly stable throughout childhood, and have high stability across many years of adulthood. Both the Concordia study (Schwartzman, Gold, Andres, Arbuckle, & Chaikelson, 1987) and the Intergenerational Studies (Eichorn, Hunt, & Honzik, 1981) found higher stability across adulthood for non-verbal/performance IQ-type abilities than for verbal abilities.

The value of knowing the stability of mental ability differences from early adulthood to old age was emphasised in two long-term follow-up studies. In the 'Nun Study' the linguistic complexity of hand-written autobiographies in early adulthood correlated with the incidence of dementia and mental ability level in late life (Snowdon et al., 1996). In a separate study, recruits from the American armed forces in the early 1940's were administered the Army General Classification Test and followed up 50 years later using

a brief telephone-administered cognitive interview (Plassman et al., 1995). This latter study made particular mention of the necessity yet rarity of having early life cognitive estimates in the interpretation of cognitive scores in old age. Therefore, though it is a research priority, we do not know the stability of psychometric intelligence differences from early to late life. The principal reason for this gap in our knowledge is the rarity of samples of the population who were tested in youth and then followed up in old age.

We now report the first follow-up study of human cognitive ability that extends from childhood (mean age 11 years) to old age (mean age 77 years), and is thus informative about the stability of mental functions across most of the human lifespan. Further improvements upon the best currently-available studies (Schwartzman et al., 1987; Plassman et al., 1995; Snowdon et al., 1996) include: (i) the use of the same validated mental test at baseline and follow-up using identical instructions; (ii) characterisation of the follow-up sample in terms of age, sex and initial IQ with respect to the entire relevant Scottish population; and (iii) concurrent validation of the mental test at first testing (age 11 years) and follow-up (age 77).

Background to the present study

We used data from the 1932 Scottish Mental Survey to investigate the stability of psychometric intelligence differences across a gap of 66 years. The Scottish Mental Survey 1932, under the auspices of the Scottish Council for Research in Education (SCRE), sought to quantify the numbers of people in Scotland who were "mentally deficient". It was broadened to "obtain data about the whole distribution of the intelligence of Scottish pupils from one end of the scale to the other" (Scottish Council for Research in Education, 1933). On June 1st 1932, all children at school in Scotland and born in the calendar year 1921 undertook a group-administered mental ability test, including some practice items. Children were tested in classrooms by teachers who followed detailed printed instructions. The number of children tested was 87,498

(44,210 boys and 43,288 girls). A very small number of private schools and those children absent owing to sickness were the only 1921-born children not tested.

The group mental ability test used in the Survey is referred to in the original publication as the 'Verbal Test' (SCRE, 1933). The test comprises a variety of types of item 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. The test was closely related to the Moray House Test No. 12 which was used in 'eleven-plus' examinations in England. We shall hereafter refer to the test as the Moray House Test. The scores on the Moray House Test in 1932 were validated by individually re-testing a representative sample of 1000 of the children (500 boys, 500 girls) on the Stanford Revision of the Binet-Simon scale. Those who administered the individual Stanford-Binet tests had special training in mental testing.

Method

Subjects

The Scottish Council for Research in Education made the complete data set for the 1932 Scottish Mental Survey available to the authors. From January to May 1998 we traced local (North-East Scotland) survivors of the 1932 Scottish Mental Survey. With the approval of the Grampian Ethics of Research Committee and family doctors, we contacted 199 survivors randomly selected from the Community Health Index (the local register of people's allocations to family physicians in the UK's national Health Service) and 35 other locals who volunteered on hearing media reports of the study. Of the 234 potential subjects, 208 people agreed to a full physical and mental health assessment and 73 agreed, in addition to the aforementioned health checks, to re-take the Moray House Test precisely 66 years to the day after the first sitting. The 73 attended a group testing session at a large public hall in Aberdeen town centre on June 1st 1998, the 66th

anniversary of the original testing session. The hall was specially furnished with desks and chairs for a group examination. Some 28 other subjects attended on dates up to five months later at times convenient to them. All but one—who attended in a wheelchair—of the 101 re-tested subjects were ambulant. None of the 101 was suffering from any major physical or mental illness or medication known to affect cognitive functioning.

Mental test and procedure

The Moray House Test was administered in a group fashion using the same instructions as those used in 1932 (SCRE, 1933). Forty five minutes were allowed for the completion of the test. Only two of the tests' items required minor altering from the 1932 version. A question involving shillings and pence was altered to feet and inches because, whereas money altered to a decimal format in the United Kingdom in 1971, the measurement of distance is still principally duodecimal, especially among old people. Another question archaically referring to "vitamine" was changed to read "vitamins." In addition to the newly-gathered 1998 test scores, all subjects' Moray House Test scores were identified from the records of the 1932 sitting.

On a separate occasion, 97 of the 101 subjects completed a further mental ability test to provide concurrent validity for the 1998 scores on the Moray House Test. The test was Raven's Progressive Matrices (Raven, Court, & Raven, 1977), a non-verbal pattern-completion test. Raven's Matrices test is a good indicator of general intelligence (Spearman's g ; Carroll, 1993). The Raven test was administered to subjects individually by trained researchers using a time limit of 20 minutes.

Results

The mean score on the Moray House Test for the 101 subjects in 1932 was 43.5 (S.D. = 11.8), and for the same subjects in 1998 was 54.2 (11.8) (Table 2). Mean scores for men and women were very similar at age 11 but, at age 77, men scored higher than women by about three points. A mixed model analysis of variance of Moray House Test scores

was carried out with time as a repeated measure (1932 versus 1998) and sex as a between subjects factor. The effect of time was highly significant, with people scoring better at age 77 years than at age 11 ($F = 111.8$, $d.f. = 1,99$, $p < .001$). There was no significant overall effect of sex on test scores ($F = 0.3$, $d.f. = 1,99$, $p = .6$). There was a statistical trend in the interaction between time and sex, with men tending to gain higher scores over the 66 year gap between tests ($F = 2.7$, $d.f. = 1,99$, $p = .1$).

The population mean for the Moray House Test in 1932 was 34.5 (34.5 for 44,210 boys, and 34.4 for 43,288 girls), and the standard deviation was 15.5 (15.9 for boys, and 15.0 for girls). Thus, the sample re-tested in 1998 had a 1932 Moray House Test score mean that was 0.59 standard deviation units higher than the population mean—equivalent to 8.9 IQ points on a standard IQ-type scale with $m = 100$, $s = 15$. The standard deviation in the re-tested sample was only 76% of that found in the population.

Concurrent validity of the Moray House Test

From the re-test in 1998 the mean score on Raven's Standard Progressive Matrices was 28.6 (S.D. = 8.5) (Table 2). Men and women had similar scores ($t = 1.3$, $d.f. = 95$, $p = .2$). Table 2 shows concurrent validity coefficients for the Moray House Test at both the 1932 and 1998 test waves. As described above, after taking the Moray House (group-administered) Test in 1932, one thousand children (500 boys, 500 girls) were tested individually on the Stanford revision of the Binet intelligence test battery. The correlation between the Moray House Test and the Stanford-Binet test was 0.81 for the boys and 0.78 for the girls. In the 1998 re-tested sample, the correlation between Moray House Test raw score and raw score on Raven's Standard Progressive Matrices was 0.55 (Table 2). All coefficients have p values of $< .001$. The coefficients of men and women did not differ significantly.

'Lifetime' stability of mental ability differences

The Pearson r correlation between the Moray House Test scores in 1932 and 1998 was 0.63 ($p < .001$). The 95% confidence limits on this correlation are from 0.50 to 0.74. This raw correlation is an underestimate of the true correlation in the population because of the attenuation of the re-tested sample with respect to variance on the 1932 Moray House Test scores. The disattenuated correlation across the 66 year gap, allowing for the restricted range of the sample, is 0.73. This corrected coefficient, too, is an underestimate of the true value because the correlation is further attenuated by measurement error on both testing occasions. The stability coefficients are similar for men and women (Table 2). The correlation between the Moray House Test scores in 1932 and the Raven test taken in 1998 was 0.49 ($p < .001$), not significantly different from the correlation between Raven and the Moray House Test taken in 1998.

Discussion

To our knowledge, this is the longest follow-up study of human psychometric intelligence differences reported to date. The interval between the two testing sessions comprises most of the normal human lifespan. The present study has attractive aspects of design which rarely occur together in other studies: the same test was used at first test and follow-up; the test had concurrent validation at both test sessions; the original and follow-up samples can be compared quantitatively with the entire age-relevant population; men and women were tested; and the initial and follow-up tests were conducted with identical delays for all subjects. The corrected correlation between the Moray House Test scores at age 11 and 77 years informs us that, for community-resident old people in relative good health, psychometric intelligence differences show high stability across most of the human lifespan. The majority—just—of the variance in test scores at age 77 is to be found as early as age 11 years. The possibility that men and women might show different patterns of change for the Moray House Test scores needs establishing in a larger study. The larger gain in the scores of men is in accord with the

results of Kangas and Bradway (1971) who reported data on 48 people studied using the Stanford Binet and/or Wechsler Adult Intelligence Scale at ages 4, 14, 30 and 42 years.

The new data provided here have implications for both applied and basic science. In the field of cognitive gerontology, the results underscore the importance of taking into account the premorbid mental ability of a person in the investigation of cognitive decline or dementia in old age (Deary, 1995). Moreover, the results validate the assumption implicit in premorbid ability estimates: that in the absence of disease processes we might expect broad stability of individual differences in mental abilities across the human lifespan.

In the field of differential psychology the genetic and environmental sources of this remarkable stability of individual differences in human intelligence must be sought. The stability exists in the face of a change in the genetic contribution to intelligence differences over the lifespan: counter-intuitively, perhaps, genes account for more of the variance in intelligence differences in old age than in childhood and young adulthood (Plomin, Pedersen, Lichtenstein, & McLearn, 1994). And we must seek an account of the half of the variance that is not stable over the lifespan. Some of this will be error variance, that properly belongs to stability. Some of it will be found in the genetic differences, coming in to play in later life, that protect us from cognitive decline or accelerate it (MacLulich, Seckl, Starr, & Deary, 1998). And some of it will be found in individual differences in the slings and arrows of fortune. The present results help us to apportion these sources of variance more clearly.

References

- Carroll, J. B. (1933). *Human mental abilities: A survey of factor-analytic studies*.
Cambridge, UK: Cambridge University Press.
- Deary, I. J. (1995). Age-associated memory impairment: A suitable case for treatment?
Ageing and Society, 15, 393-406.

- Eichorn, D. H., Hunt, J. V., & Honzik, M. P. (1981). Experience, personality, and IQ: adolescence to middle age. In D. H. Eichorn, J. A. Clausen, N. Haan, M. P. Honzik, & P. H. Mussen (Eds.), *Present and Past in Middle Life*. New York: Academic Press.
- Humphreys, L. G. (1989). Intelligence: three kinds of instability and their consequences for policy. In R. L. Linn (Ed.), *Intelligence*. Urbana: University of Illinois Press.
- Jensen, A. R. (1980). *Bias in mental testing*. London: Methuen.
- Kangas, J., & Bradway, K. (1971). Intelligence at middle age: a thirty-eight-year follow-up. *Developmental Psychology*, 5, 333-337.
- MacLulich, A. M. J., Seckl, J. R., Starr, J. M., & Deary, I. J. (1998). The biology of intelligence: from association to mechanism. *Intelligence*, 26, 63-73.
- Mortensen, E. L., & Klevan, M. (1993). A WAIS longitudinal study of cognitive development during the life span from ages 50 to 70. *Developmental Neuropsychology*, 9, 115-130.
- Nisbet, J. D. (1957). Intelligence and age: retesting with twenty-four years' interval. *British Journal of Educational Psychology*, 27, 190-198.
- Owens, W. A. (1966). Age and mental abilities: a second adult follow-up. *Journal of Educational Psychology*, 57, 311-325.
- Plassman, B. L., Welsh, K. A., Helms, M., Brandt, J., Page, W. F., & Breitner, J. C. S. (1995). Intelligence and education as predictors of cognitive state in late life: A 50-year Follow-up. *Neurology*, 45, 1446-1450.
- Plomin, R., Pedersen, N L., Lichtenstein, P., & McClearn, G. E. (1994). Variability and stability in cognitive abilities are largely genetic in later life. *Behavior Genetics*, 24, 207-215.
- Raven, J. C., Court, J. H., & Raven, J. (1977). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: Lewis.

- Schwartzman, A. E., Gold, D., Andres, D., Arbuckle, T. Y., & Chaikelson, J. (1987). Stability of intelligence—A 40 year follow up. *Canadian Journal of Psychology, 41*, 244-256.
- Scottish Council for Research in Education. (1933). *The Intelligence of Scottish children: A national survey of an age-group*. London: University of London Press.
- Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the Nun Study. *Journal of the American Medical Association, 275*, 528-532.
- Tuddenham, R. D., Blumenkrantz, J., & Wilkin, W. R. (1968). Age changes on AGCT: A longitudinal study of average adults. *Journal of Consulting and Clinical Psychology, 32*, 659-663.

Acknowledgements

The study was supported by a grant to LJW from Henry Smith's Charities. Helen Francis, Patricia Whalley and Mariesha Struth assisted in collection and collation of data. We are indebted to the Scottish Council for Research in Education—especially Graham Thorpe and Professor Wynne Harlen—for providing data from the 1932 Scottish Mental Survey.

Table 1

Summary of some key studies of the stability of individual differences in psychometric intelligence.

Study	Mean initial age (years)	Mean follow-up age	Correlation	Test used
Humphreys (1989)	2	9	.56	Wechsler Preschool and Primary Scale of Intelligence & Wechsler Intelligence Scale for Children
	9	15	.47	
	2	15	.78	
Kangas & Bradway (1971)	4	42	.41	Stanford Binet
	14	42	.68	
	30	42	.77	
Eichorn et al. (1981)	17-18	36-48	.83 (men) .77 (women)	Stanford Binet or Wechsler Bellevue (initial) & Wechsler Adult Intelligence Scale (follow-up)
Plassman et al. (1995)	Approx. 18	Mid 60s	.46	Army General Classification Test (initial) & Telephone Interview for Cognitive Status (follow-up)
Owens (1966)	19	50	.79	Army Alpha
	50	61	.92	
	19	61	.78	
Nisbet (1957)	22	47	.48	Simplex Group Test
Schwartzman et al. (1987)	25	65	.78	Revised Examination 'M'
Tuddenham et al. (1968)	30a	43	.64-.79	Army General Classification Test
Mortensen et al. (1993)	50	60	.94	Wechsler Adult Intelligence Scale
	60	70	.91	
	50	70	.90	
Deary et al. (present study)	11	77	.63	Moray House Test

aSubjects were probably seven years younger than this making the follow-up interval 20 rather than 13 years.

Table 2

Descriptive statistics of, and correlations among, mental tests for all subjects and separately for men and women.

	Mean (SD)	N	MHT 1932	MHT 1998
All subjects				
MHT 1932	43.5 (11.8)	101		
MHT 1998	54.2 (11.8)	101	.63	
Raven	28.6 (8.5)	97	.45	.55
Men				
MHT 1932	43.1 (13.3)	49		
MHT 1998	55.7 (11.8)	49	.62	
Raven	29.8 (8.7)	47	.43	.53
Women				
MHT 1932	43.9 (10.3)	52		
MHT 1998	52.8 (11.8)	52	.68	
Raven	27.5 (8.3)	50	.50	.55

Note. All correlations are $P < .01$. MHT = Moray House Test; Raven = Raven's Standard Progressive Matrices.