Intelligence 53 (2015) 59–64

Contents lists available at ScienceDirect

Intelligence

Interindividual differences in general cognitive ability from age 18 to age 65 years are extremely stable and strongly associated with working memory capacity

Michael Rönnlund [a,*], Anna Sundström [a,b], Lars-Göran Nilsson [c,d]

[a] *Department of Psychology, Umeå University, Sweden*
[b] *Centre for Demographic and Ageing Research, Umeå University, Sweden*
[c] *Aging Research Center, Karolinska Institutet, Sweden*
[d] *Umeå Center for Functional Brain Imaging, Umeå University, Sweden*

ARTICLE INFO

Article history:
Received 30 November 2014
Received in revised form 31 July 2015
Accepted 31 August 2015
Available online xxxx

Keywords:
General cognitive ability
Inter-individual differences
Stability
Longitudinal

ABSTRACT

The objective of the study was to examine the degree of stability of interindividual differences in general cognitive ability (g) across the adult life span. To this end, we examined a sample of men ($n = 262$), cognitively assessed for the first time at age 18 (conscript data). The sample was reassessed at age 50 and at five year intervals up to age 65. Scores from conscript tests at age 18 were retrieved and three of the subtests were used as indicators of g in early adulthood. At age 50–65 years, four indicators served the same purpose. At the 15-year follow-up (age 65) two working memory measures were administered which allowed examination of the relationship with working memory capacity. Results from structural Equation Modelling (SEM) indicated extremely high level of stability from young adulthood to age 50 (standardized regression coefficient = −95) as well as from age 50 to age 55, 60 and 65 with stability coefficients of .90 or higher for the for the latent g factor. Standardized regression coefficients between young-adult g and the g factor in midlife/old age were .95 from age 18 up to age 50 and 55, .94 up to age 60, and .86 up to age 65. Hence, g at age 18 accounted for 90–74% of the variance in g 32–47 years later. A close association between g and working memory capacity was observed (concurrent association: $r = .88$, time lagged association: $r = .61$). Taken together, the present study demonstrates that interindividual differences in g are extremely stable over the period from 18 to midlife, with a significant deviation from unity only at age 65. In light of the parieto-frontal integration theory (P-FIT) of intelligence, consistent with the close association between g and working memory capacity, midlife may be characterized by neural stability, with decline and decreased interindividual stability, related to loss of parieto-frontal integrity, past age 60.

1. Introduction

A general cognitive ability (g) factor has been proposed to account for the "positive manifold" observed across a variety of cognitive measures (Jensen, 1998; Carroll, 1993; Spearman, 1904). The g factor is important for certain aspects of life-success, as judged by associations between proxy measures (IQ scores) and later achievements, for example, educational attainment, job performance, and socioeconomic status (Deary, Whiteman, Starr, Whalley, & Fox, 2004; Strenze, 2007).

An important issue concerns to what degree interindividual differences in the g factor are stable over the life course. As noted by Deary, Whalley, Lemmon, Crawford and Starr (2000), it is of interest in childhood to discover whether educational interventions boost ability levels and whether adverse environmental factors lower cognitive functions. Stability estimates in the period from early adulthood to old age should similarly be informative of the extent to which individual differ in cognitive ageing. Given that individuals are differentially exposed to a myriad of factors that potentially influence cognition, one might expect the stability of individual differences from youth to late adulthood to be rather moderate.

However, contrary to this expectation, several studies indicate considerable stability of individual differences in IQ-test performance over long time periods (for an exception see Plassman et al., 1995). For example, in a study by Owens (1966), 96 freshmen at the Iowa State University took the Army Alpha, a group test originally developed to evaluate military recruits, when they were 19 years old and were retested 42 years later. The correlation between the initial and follow-up test scores was as high as .78, suggesting that about 60% of the variance in ability at age 61 was predictable from test scores age 19. More recently, Deary et al. (2000) examined a group of 101 adults who took the Moray House Test when they were 11 years old and were retested 66 years later, at age 77. The test–retest correlation was 0.63 and boosted to 0.73 when restriction of ability

* Corresponding author at: Department of Psychology, Umeå University, S-90187 Umeå, Sweden.
E-mail address: michael.ronnlund@umu.se (M. Rönnlund).

http://dx.doi.org/10.1016/j.intell.2015.08.011
0160-2896/

range at retest was adjusted for, suggesting that up to half of the variance in ability at age 77 was accounted for by adolescent performance. A further study involving the same intelligence test reported a coefficient of 0.54 from age 11 to age 90, with a value of 0.67 when restrictions in range were corrected (Deary, Pattie, & Starr, 2013). As discussed by the authors, the reported values may actually be an underestimate of stability of interindividual differences due to measurement error (i.e. lack of perfect reliability of the test).

In the present study we further examined the long-term stability of general intelligence, targeting the period from early adulthood to young-old age. We examined a group of men for whom we retrieved data from conscript testing at age 18. At age 50 study group took several cognitive tests as part of a longitudinal study on ageing and cognitive functions (Nilsson et al., 1997; 2004). A longitudinal follow-up of the tests at age 50 was furthermore made five, ten and fifteen years following the midlife assessment. Hence, the extent to which ability at age 18 was predictive of ability at age 50, 55, 60, and 65 years could be compared based on data from the same study sample and the repeated testing at higher ages allowed for examining the stability coefficients across a range in which cognitive functions generally dip longitudinally (e.g. Rönnlund, Nyberg, Bäckman, & Nilsson, 2005; Schaie, 1994). Importantly, even though different cognitive tests were administered in early and late adulthood, with three or more indicators of general ability at age 18 and at ages 50–65, the relationship between early and later cognitive ability could be examined at the latent ability level, rather than, as in the aforementioned studies, for manifest test scores which confound stability with measurement error, and hence tend to underestimate stability at the construct level. Finally, we examined the concurrent (at age 65) and time-lagged relationship between general ability and working-memory capacity, measures that were added to the battery at the fifteen-year follow-up. Working memory capacity has in prior studies (e.g. Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Kyllonen & Christal, 1990; Unsworth, Fukuda, Awh, & Vogel, 2014) been found to be substantially associated with general intelligence, and measures of fluid reasoning (Gf) in particular.

2. Method

2.1. Participants

The included participants were enrolled in the Betula prospective cohort study, a longitudinal study of memory and health in Umeå, Sweden. The study started in 1988 and involves randomized sampling from the population register in Umeå community (Nilsson et al., 1997; 2004). At the second test occasion (T2; 1993–1995), considered as baseline in the present study, three subsamples were involved: Sample 1 (S1; 40–85 years), Sample 2 (S2; 35–80 years), and Sample 3 (S3; 40–85 years). For participants in S1 and S3, a longitudinal follow-up was made five (1998–2000), ten (2003–2005), and fifteen years (2008–2010) later.

Following approval from a regional ethic committee, we retrieved cognitive test scores for a subset of the participants in S1–S3 from the archive containing information gathered at draught boards (The Swedish military archives). Targeted were those participants expected to have taken the conscript tests during a period from 1954, when standardized scores were registered for each subtest and for total performance, and up to 1967. For 432 of the 435 participants (99.3%) the data were successfully retrieved. The present study sample involved 262 participants in S1 and S3 who were 45, 50, or 55 years (approximately 1/3 at each of the age levels) at baseline.

2.2. Measures

2.2.1. General cognitive ability — Swedish enlistment battery

Depending on birth cohort, each participant had taken one of two versions of the Swedish Enlistment Battery (SEB). The first, SEB-1954 ($n = 82$), involved five tests: Instructions, Concept discrimination, Technical comprehension, Levers, and Multiplication. Descriptions of the tests and standardized loadings (SL) on a general and specific ability factors are adopted from Carlstedt (2000; see Table 1, p. 35 for loadings based on confirmatory analyses, that also included other tests). Labelling of factors are adopted from Schneider and McGrew (2012), were g = general ability, Gf = fluid reasoning, Gc = comprehension-knowledge; Gv = visual processing, and Glr = Long-term storage and retrieval.

The Instructions test, intended to measure the primary factor Induction, a narrow ability within Gf (Schneider & McGrew, 2012) mainly reflected g (SL on $g = 0.77$; Gc = 0.19). This test contained verbal instructions to make markings on the answer sheet that fulfilled the conditions provided by the instructions. Item difficulty was manipulated by complexity of the instructions and by distractive negations or conditional clauses. Concept Discrimination involved classification of words (SL on $g = 0.67$; SL on Gc = 0.30). Technical comprehension involved a set of illustrated technical and physical problems and reflected g, Gc, as well as Gv; Carlstedt, 2000). Levers was a mechanical reasoning test with main loading on Gv (SL = 0.58; SL on $g = 0.43$). Multiplication, finally, was found to reflect a "Math" factor (SL = 0.54) apart from g (SL = 0.67).

In the second version, SEB-1959 ($n = 180$), that replaced the SEB-1954, Instructions, Concept discrimination and Technical comprehension were retained. Multiplication was excluded, and the Levers test was replaced by Paper form board. The latter test involved judgments of which target object out of four would be correctly put together by a set of disarranged parts of objects and reflected g and Gv (Carlstedt, 2000).

Regardless of SEB version, a standard-nine score (M = 5, SD = 2) was provided for each subtest together with a total score based on the sum of the standardized subtest scores. In the present analyses we used a composite measure created by averaging the standard-nine scores of the 4–5 (depending on SEB version) subtests. Further, we used the three tests common to both SEB versions (Instructions, Concept discrimination, and Technical comprehensions) as indicators of a general ability factor. A principal components analysis (PCA) of all data available for SEB-1959 (S1-S3; $n = 147$) showed that the foregoing three test were those with the highest loadings on a single component with eigenvalue >1 (0.88 for Instructions, 0.82 for Concept discrimination, and 0.76 for Technical comprehension) that accounted for 57.0% of the variance in test scores.

2.2.2. General cognitive ability — the Betula battery

Judgments based on content, including a wish to include verbal as well as nonverbal materials and a wish to tap different factors at the stratum II-level (Carroll, 1993), served as a basis for selecting four indicators of general ability. The first indicator was raw scores on the WAIS-R Block Design test (BDT; Wechsler, 1981) which may be regarded to reflect g/Gv. Cronbach's α for the BDT in a large Swedish sample was .82 (Rönnlund & Nilsson, 2006a). The second measure was scores on a 30 item multi-choice vocabulary (VOC) test (Dureman, 1960) which should reflect g/Gc. Split-half (Spearman–Brown) reliability of .82 was reported for this measure (Rönnlund & Nilsson, 2006a). The third measure, action recall (ARC) was computed as the sum of two free recall trials involving 16 action phrases (e.g. "Lift the book") enacted at study, and 16 action phrases with no enactment at study (Rönnlund, Nyberg, Bäckman, & Nilsson, 2003), regarded to reflect g and Glr (associative memory and free recall). Split-half coefficient for the separate trials/conditions were .63 and .62 (Rönnlund & Nilsson, 2006a) and a five-year test–retest correlation of the sum of these was $r = .60$ in the present sample. Finally, a measure of word fluency (WFL) was computed as the sum of two phonemic fluency tests that involved oral generation of as many words as possible during one minute. The restrictions were i) words with initial letter A and ii) words with initial letter M containing five letters. The measure should reflect g/Glr