Chapter 2
What Is Fluid Intelligence? Can It Be Improved?

Patrick Kyllonen and Harrison Kell

Abstract General fluid intelligence (*Gf*) is the ability used in inductive and deductive reasoning, particularly with novel material. It can be contrasted with general crystallized ability (*Gc*) which reflects schooling and acculturated learning, and the two abilities have different developmental trajectories, with *Gf* peaking earlier in the lifespan. Gustafsson has made key contributions to our understanding of *Gf*. He (Gustafsson 1984) introduced hierarchical confirmatory factor analytic models to reconcile Thurstonian (non-hierarchical) and Spearman and Cattell-Horn (hierarchical) models of intelligence and in so doing identified *Gf* as a second-order factor which perfectly correlated with the third-order factor, general ability (*g*). This has important implications for understanding the nature of general cognitive skill. Subsequent research showed that *Gf* can be identified separately from g through variation in culture-related opportunities to learn (Valentin Kvist and Gustafsson 2008). *Gf* has served both as a predictor (Gustafsson and Balke 1993) and outcome (Cliffordson and Gustafsson 2008) in the developmental, cognitive training, cognitive aging, international comparative assessment, genetics, neuropsychopharmacological, human capital theory, and behavioral economics literatures. Understanding the nature of fluid intelligence and how to improve it has become a topic of renewed and general interest for optimizing human performance in school and in the workplace.

2.1 Introduction

General fluid ability (*Gf*) is commonly defined as the ability to solve problems in unfamiliar domains using general reasoning methods (Carroll 1993; Cattell 1963). It is typically contrasted with general crystallized ability (*Gc*), which is the ability to answer questions or solve problems in familiar domains using knowledge and strategies acquired through education, training, or acculturation. These two broad abilities are highly correlated, suggesting a common or general factor (*g*). One

P. Kyllonen (✉) · H. Kell
Educational Testing Service, Princeton, NJ, USA
e-mail: pkyllonen@ETS.ORG

M. Rosén et al. (eds.), *Cognitive Abilities and Educational Outcomes*,
Methodology of Educational Measurement and Assessment,
DOI 10.1007/978-3-319-43473-5_2

explanation for the *Gf*-*Gc* correlation is given by investment theory (Cattell 1987) which suggests that *Gf* is *invested* in learning so that the rate of learning different tasks depends on *Gf* (along with motivation and opportunities to learn). Therefore school achievement (*Gc*), reflecting the rate of learning, is related to *Gf*. Thorsen et al. (2014) found support for this idea, and further showed that *Gf* is invested in *Gc* not just initially, but continually throughout the school years. They showed that *Gf* measured in 3rd grade predicted *Gc* measured in 9th grade after controlling for *Gc* measured in 6th grade.

Stated this way it might be presumed that *Gf* is an important ability, as solving novel problems is a hallmark of intelligence. It is also essential to learning in school, performance on the job, and in life generally. Empirical studies confirm this intuition. *Gf*, as measured by various tests, has been shown in meta-analyses to predict school grades, and job training and job performance, particularly for high compared to medium complexity (Postlethwaite 2011). Studies based on representative samples show moderate to high correlations between *Gf* (measured by Raven's Progressive Matrices) and national achievement examinations (Pind et al. 2003). *Gf* predicts school achievement, and growth in school achievement, but is not by itself notably affected by the quality of schools (Finn et al. 2014). *Gf* also predicts life outcomes such as earnings, criminality, civic participation, and educational attainment (Borghans et al. 2008; Lindqvist and Vestman 2011). It also is an ability characterizing so called *super-forecasters* (Mellers et al. 2015), people who consistently make accurate predictions about future geopolitical events across a wide range of topics.

In this chapter we address several issues pertaining to general fluid ability. These are (a) what is it? How is it identified? (b) is it distinguishable from g? from *Gc*?, (c) does it improve, and can we improve it?

2.2 What Is an Ability?

The concept or construct of general fluid ability arises from the *abilities model*, or *psychometric model of intelligence* (e.g., Hunt 2011). It is based on the empirical observation that if a group of people are administered a variety of cognitive tests, such as samples of school tasks, intellectual puzzles, problems to solve, or in general, anything that requires learning, memory, or thought, then the most successful people on one test will tend to be the most successful on another, that is, there is a common factor. The factor model formalizes this relationship by positing a common, unobserved (latent) factor to account for correlations among test scores. If there are four tests, there are six ($n * [n - 1]/2$, $n = 4$) correlations among them. But a factor model posits a single latent factor (x, serving as an independent variable) with a regression coefficient (factor loading) for each of the four tests (y1–y4, serving as separate dependent variables), thereby accounting for the six observed data points with only four parameters, a savings of two parameters, and therefore a more parsimonious representation of their relationships. Fractal-like, this relationship can be repeated for clusters of tests; and empirically it turns out that there are clusters of tests (e.g., spatial tests, verbal tests) whose interrelationships cannot be completely accounted for by the general factor. This necessitates additional group factors, one for each cluster.

This model of accounting for test inter-correlations parsimoniously through the positing of latent factors is the fundamental basis by which we say that people have *abilities*. Abilities are the unobserved latent variables. There is a general ability (g) and there are group abilities. Carroll (1993, p. 23) refers to a cognitive ability as "an intervening variable, i.e., a calculation convenience, as it were, in linking together a particular series of observations." The abilities concept is not completely dependent on a factor model—other methods of representing the empirical fact of clustering of test-score variables, such as hierarchical clustering or multidimensional scaling also invoke an abilities explanation (Corno et al. 2002, p. 66; Snow et al. 1984). However, the factor model may more easily represent hierarchies of abilities. It was once thought that the general factor (Spearman 1927) and group abilities theories (Thurstone 1938) represented different abilities models, but Gustafsson (1984) showed that both views could be accommodated in the same hierarchical model which provided a better summary of their empirical relationships.

2.3 What Is General Fluid Ability?

Given that there is empirical clustering, what is the nature of those clusters? This can be explored by studying the common features of the tests that are the best representatives of those clusters, that is, the ones that have high factor loadings. A limitation is that factors from factor analysis simply represent what is common to a group of variables studied, and so if a kind of test is never studied then a factor underlying performance on such a test would never be identified; conversely, if a particular kind of test appears in many studies, a factor will be identified to account for that fact. However, reviews of studies that have been conducted on tests and their interrelationships reflect what many researchers have thought important enough to conduct a study on, and therefore the findings that emerge in such reviews will be important in the sense of warranting attention.

One such review was Carroll's (1993) meta-analysis of 460 datasets. He found 241 instances where a reasoning factor (these subsequently were labeled fluid ability tests by Carroll) was identified. Carroll then classified these factors into three categories based on his analysis of their common and unique features. The 236 factors, i.e., the ones that occurred more than once and had relatively high loadings, fell into the categories of

(a) sequential or deductive reasoning factors (e.g., categorical syllogisms ["Some dogs don't bark, Fido is not a dog, can he bark?"], linear syllogisms [Fred is taller than Sam, but shorter than Joe, who's tallest?]), general verbal reasoning [e.g., deductive reasoning, as in identifying a logical conclusion based on verbally stated problem situation],

(b) inductive reasoning factors (e.g., rule discovery, number or letter series tasks, multiple exemplars tasks, matrix tasks, odd element or "odd man out" tasks, and analogies tasks); and
(c) quantitative reasoning factors (these are typically inductive or deductive reasoning tasks but involve quantitative elements, such as number series).

Carroll found that these three main categories of fluid reasoning tests were difficult to distinguish empirically, although there was some tendency for Induction and Sequential Reasoning to be more correlated with *Gf* factors, and Quantitative reasoning to be relatively more related to *Gc* factors. Commercial tests reflect this inconsistency. The Number Series test (Test 24) in the Woodcock-Johnson® III (WJ III®) (Woodcock et al. 2001) test battery is classified both as measuring *Fluid Reasoning* (*Gf*) and as measuring narrower *Mathematics Knowledge* and *Quantitative Reasoning* factors (Schrank 2006). The Woodcock Johnson classification is based on what has come to be known as the Cattell-Horn-Carroll (CHC) theory (McGrew 2005).

Carroll (1993) classified number series tests (Series Tasks more generally) as Inductive Tasks (p. 211), but also pointed out that they can be made more difficult and "thus classifiable as a quantitative task" (p. 213). Wilhelm (2006) reviewed Carroll's classification (and the separation of deductive and inductive reasoning) and pointed out that Carroll viewed series tests as tentative markers of induction due to the fact that analyses are often based on studies that have weak designs and show a single-factor solution. Wilhelm also pointed out that there is often a "content confound," with deductive tasks being primarily verbal, and inductive tasks often being spatial in content. This breakdown suggests that Carroll's Inductive-Deductive (Sequential)-Quantitative split is confounded with a Verbal-Spatial-Quantitative split. A study by Wilhelm (2000) (in German, but reviewed in Wilhelm 2006) suggested that a better representation of relationships is that there is a general fluid ability (*Gf*) factor with Verbal, Figural, and Quantitative sub-factors, and that Number Series was more closely aligned with two deductive measures, *Solving Equations* and *Arithmetic Reasoning,* than with other inductive measures with other contents, such as *Figural Classifications* and *Figural Matrices*.

2.4 Explorations of Measures of General Fluid Ability

In addition to identifying common features of tests that cluster in factor analyses, another approach taken to understand fluid ability has been to explore more systematically features of tasks that are good measures of (have a high correlation with) fluid ability. This approach must be taken with caution as particular measures have considerable task-specific variance (Gustafsson 2002). Nevertheless it can be informative to study several tests in such a way and examine potential underlying commonalities.

2.4.1 Complexity

An example is the complexity hypothesis (e.g., Stankov and Schweizer 2007), which is that because intelligence, particularly *Gf*, is the ability to deal with complexity in some sense, more complex tests must be better measures of intelligence. For example, Stankov and Schweizer (2007) defined complexity as the number of steps needed to reach a solution (their Swaps test), or the level of embedding of a rule used to sort number strings into two categories (their Triplets test), and found that thereby increasing complexity increased task difficulty and led to higher correlations with a *Gf* measure, Raven's progressive matrices. However, Gustafsson (1999) found that an equally plausible measure of complexity—having to switch problem solving sets on every item due to heterogeneous versus homogeneous item type groupings (using figural "odd man out," figure series, and Bongard figure classification tasks)—did not result in higher loadings. In fact, homogeneous groupings tended to lead to higher factor loadings, a result he attributed to the opportunities afforded by homogeneous grouping for within-task learning. He also invoked a working-memory explanation to account for his findings.

Informal approaches to defining complexity might be criticized [akin to Boring's (1923, p. 37) definition of intelligence "as what the tests of intelligence test"]. Halford et al. (1998) proposed a formal specification of the relationship between task features and cognitive complexity called *relational complexity*. Birney and Bowman (2009) applied this framework to both the Swaps and Triplets task as well as to a Latin Square and simple sentence comprehension task. However, they found no advantages to this particular framework over others in predicting *Gf* correlations.

2.4.2 Reasoning Ability and Working-Memory Capacity

The complexity hypothesis, and relational complexity, are essentially working-memory explanations. Working memory is the notion of a limited, short-term system in which temporary storage and information processing of the current focus of thought occurs (Baddeley 2003). The idea that individual differences in working memory capacity might underlie general fluid ability was originally suggested by Kyllonen and Christal (1990), who found correlations above $r = 0.80$ between latent factors of fluid ability and latent factors of working memory. The primary value of the study was to show that tests could be developed to measure working memory capacity using Baddeley and Hitch's (1974) simple definition (tasks that require simultaneous storage and processing of information, such as mental addition) and that capacity measures from those tasks correlated highly with established measures of fluid ability (such as sets, series, and matrices tests).

However, the question of whether the two factors are the same has been an issue addressed. Many factors influence the correlation estimate including measurement error (e.g., correlations between tests vs. latent factors), contents (e.g., verbal, spatial, numerical), item types (e.g., series, matrices, span), speededness, task-specific factors, strategies, and so on. If two large sets of reasoning and working memory measures do not confound these factors, then the latent factor inter-correlation will tend towards unity (e.g., Kyllonen 1995). On the other hand, if random working memory tests are correlated with random reasoning tests without attention to confounding of content, test type, and other factors, then the correlations will be lower. This finding has been evaluated in numerous studies since with a variety of methodologies (Conway et al. 2007; Unsworth et al. 2014). The general view now seems to be that the two factors are highly correlated but not the same (e.g., Kane et al. 2004). One meta-analytic estimate of that correlation is $\rho = 0.85$ (Oberauer et al. 2005), which is based on but higher than Ackerman et al.'s (2005) meta-analytic estimate of $\rho = 0.63$ (between working memory and *Gf*, controlling for content; p. 38). Much of the source for the distinction between typical reasoning and typical working memory tasks is in content (e.g., Wilhelm 2006) and to a lesser extent, speededness (Ackerman et al. 2005). Chuderski (2015) found that *Gf* tested under time pressure overlapped considerably with working memory (83 % variance) whereas *Gf* tested without time limits overlapped considerably less (58 %).

The findings in this literature are at least consistent with the idea that working memory capacity explains or contributes to differences in *Gf* task performance, but that other task effects (e.g., content, paradigm, speededness) also affect both *Gf* and working-memory task performance and may differentiate the two categories. It is not clear that factor analysis of tasks that are labeled as *Gf* or working memory tasks is the best method for addressing the issue. The systematic manipulation approach controlling for other factors such as what was done to explore the complexity hypothesis may be more appropriate. Analyses of Raven's Progressive Matrices have been conducted in this spirit.

2.4.3 Raven's Progressive Matrices

Progressive matrices, in particular, *Raven's progressive matrices*, has long been considered one of, if not the single best measure of general cognitive ability and general fluid ability (Gustafsson 1998). For example, Snow et al. (1984) summarized various studies of cognitive measures that showed Raven at the center of a multidimensional scaling representation, corresponding to having the highest general factor loading. This prompted Carpenter et al. (1990) to conduct a detailed information processing analysis of what the RPM measures, which they found to be the ability to encode and induce relationships between elements and to manage this in working memory. On the applied testing side, the literature findings concerning the matrix test led to the development and inclusion of a new matrix test for the WAIS-III and WAIS-IV IQ tests, the most widely used today. Embretson (2002) developed an adaptive version of a figural matrix test called the adaptive reasoning test (ART), based on the rules identified by Carpenter et al. (1990). Preckel and Thiemann (Preckel 2003; Preckel and Thiemann 2003) also have developed versions.

It is possible in principle to produce a matrix test with materials other than the simple geometric forms used in the Raven's Progressive Matrices (RPM) test. However, in practice the geometric forms work quite well, as Embretson (2002) showed, and as the adoption of a figural matrix test into the WAIS establishes. An advantage of the figural stimuli in the RPM, with regard to it being a good measure of *Gf*, is that their use and the rules that operate on them are novel, which means that cultural and educational effects are reduced relative to what they might be if for example numbers (and numerical relations) or words (and semantic relations) were used instead.

A question is whether inducing rules or keeping relations in mind (active in working memory) is the major source of difficulty on Raven's Progressive Matrices, or on inductive tasks in general. If there were many potential rules linking elements to one another then this might suggest that discovering those rules would be major source of difficulty. But Jacobs and Vandeventer (1972) examined 166 intelligence tests listed in the Buros (1965) Mental Measurement Yearbook and 35 additional tests in the ETS test collection library that fell into Guilford's (1967) category of *cognition of figural relations* (CFR), which overlaps considerably with Carroll's *Gf* category, subject to the stimuli being primarily figural (as opposed to verbal or numerical). Their purpose was to categorize all the rules (relationships between elements in the problem) that were used in such tests. They found that almost all relations between two elements in the 1335 item pool fell into 12 categories such as *shape change* (a change of form, e.g., square to circle, solid to dotted; 53 % of all relations); *elements of a set* (i.e., each element appears three times in a 3×3 matrix, 35 % of relations); movement in a plane (e.g., rotating 30°) (28 %).

Carpenter et al.'s (1990) analysis of Raven's Progressive Matrices (RPM) resulted in a set of just five rules, which exhaust those used in the RPM (and which map to Jacobs and Vandeventer's 1972, rules; see also, Diehl 2002; Embretson 2002). These were "constant in a row" (element is the same across columns) (53 %), "distribution of three" (element appears once in each row, once in each column) (35 %), "distribution of two" (same as distribution of three with one being a "null" element) (28 %), "pairwise progression" (an element changes across rows and columns) (26 %), and "figure addition/subtraction" (two entries visually sum to a third entry) (24 %). Carpenter et al. (1990) suggest that most problems can be described by identifying the specific elements within a 3×3 matrix cell, then applying one of the five rules to that element.

The Carpenter et al. (1990) study did not establish whether the rules and the processes were sufficient for creating a test that would behave well psychometrically. However, Embretson (2002) did this, and so did Diehl (2002) in her dissertation. Embretson (2002) developed a matrix test called the Adaptive Reasoning Test (ART) based on the five rules and on manipulating the number of elements and rules within an item. Embretson (2002) and Diehl (2002) also provided additional specifications for those rules, and developed a useful notation. In Embretson's (2002) notation, each element (e.g., shape) is represented by a letter, which can be subscripted to define attributes on that element (e.g., pattern, size, shading, thickness, orientation, number, color, etc.). In Embretson's (2002) system, any of 22 objects and 7 attributes appear in any "item structure," and an item structure was assumed to create an item with the same difficulty level, which analysis showed is reasonable.

Increasing the number of rules and the number of elements in a matrix reasoning test item increases processing time, beyond a simple additive function (Mulholland et al. 1980; Primi 2001). It also leads to increased demands on working memory (Primi 2001). Increasing difficulty by increasing the number of rules and elements is one of the main drivers of construct-relevant increased difficulty in matrix reasoning tests. On the other hand, perceptual complexity of the geometrical shapes in the matrix (which affects encoding difficulty) also affects item difficulty (Meo et al. 2007; Primi 2001), but does not affect how good a measure of *Gf* it is (Arendasy and Sommer 2012), and also tends to have a greater impact on the performance of females. This is also true of the adjacent fusion phenomenon in which adjacent elements become difficult to distinguish perceptually. This may not be a desirable item feature to include. Hornke and Habon (1986) and Arendasy and Sommer (2005) found that figural matrices with fused elements governed by different rules introduce a second dimension, and reduced *Gf/g* saturation (Arendasy & Sommer 2012).

In summary, several methods have been used to explore the nature of general fluid ability. These include defining common features of tests that cluster in factor analyses, and systematically manipulating features of tests to determine which might serve as *radicals* (construct-relevant difficulty-manipulating factors), or *incidentals* (construct-irrelevant factors) to use Irvine's (2002) terminology. Based on these analyses it appears that working-memory capacity is an important factor contributing to performance on tests of general fluid ability.

2.5 General Fluid Versus General Crystallized Ability

Spearman's (1924) general factor theory was challenged by Cattell (1963) (see also Cattell and Horn 1978) who proposed two general factors, general fluid and general crystallized. The distinction between fluid, crystallized, and general ability is often ignored in the literature (e.g., Herrnstein and Murray 1994; Jensen 1998; Nisbett et al. 2012; Rindermann 2007) as well as in the lay population which does not differentiate between them (Kaufman 2012). The proposal for a fluid-crystallized differentiation was based on the conceptual distinction between tasks reflecting the phenomenon in which "skilled judgment habits have become crystallized" due to schooling or prior learning experiences (Cattell 1963, p. 2) and those requiring adaptation to new situations (cf., Schank 1984). It was also based on the empirical finding that over the life span performance on *Gf* task peaks sooner and drops more rapidly than does performance on *Gc* tasks (discussed in more detail below, in "Fluid Ability and Age").

These findings motivated Cattell's (1987) investment theory which posits that in early life, a single general factor, *Gf*, "primarily associated with genetic factors and neurological functioning" (Valentin Kvist and Gustafsson 2008) is invested in and governs the rate of learning and "as a result of the fluid ability being invested in all kinds of complex learning situations, correlations among these acquired, crystallized abilities will also be large and positive, and tend to yield a general factor" (Cattell 1987, p. 139).

Carroll's (1993) meta-analysis identified a number of crystallized ability factors. These included language development, verbal and reading comprehension, lexical knowledge, foreign language aptitude and proficiency, listening and communication abilities, spelling, grammar, and phonetic coding, and cloze ability (the ability to infer correctly a missing word in a sentence or paragraph). It is clear that these are language-related factors.

Postlethwaite's (2011) meta-analysis showed even higher predictions of educational and workforce outcomes for crystallized than for fluid ability (Table 2.1). His classification scheme was based on McGrew's (1997) cross-battery classification system based on the Carroll-Horn-Cattell *Gf-Gc* model (however, his classification is subjective, a potential limitation to the study). The analysis found that crystallized ability predicted school grades, job training particularly for high compared to medium complexity jobs, and job performance, particularly higher skill jobs.

Although fluid and crystallized abilities often fail to be distinguished, there is growing recognition in both psychology (Ackerman 1996; Hunt 2011) and

Table 2.1 Correlations of *Gf*, *Gc*, and *g* with outcomes (Adapted from Postlethwaite 2011, Tables 6–14)

	Fluid ability[a]			Crystallized ability[b]			General ability[c]		
	k (N)	r	rho	k (N)	r	rho	k (N)	r	rho
Grades in school	67 (7991)	0.26	0.40	157 (199,642)	0.36	0.65	110 (29,739)	0.47	0.68
High School	26 (4134)	0.30	0.38	18 (2100)	0.43	0.53	32 (13,290)	0.53	0.65
College	41 (3857)	0.22	0.44	139 (197,542)	0.36	0.65	78 (16,449)	0.42	0.72
Job training	20 (3724)	0.25	0.54	114 (38,793)	0.38	0.70	24 (7563)	0.28	0.59
Low skill jobs[e]	11 (2658)[d]	0.23	0.44	29 (8152)	0.41	0.73	2 (156)	0.22	0.53
High skill jobs[f]	5 (569)	0.32	0.67	4 (596)	0.45	0.75	2 (2824)	0.22	0.57
Job performance	23 (3272)	0.14	0.27	199 (18,619)	0.23	0.49	86 (8070)	0.23	0.43
Low skill jobs[e]	2 (251)	0.01	0.01	108 (9307)	0.22	0.45	37 (3420)	0.20	0.37
High skill jobs[f]	2 (132)	0.31	0.64	27 (2214)	0.29	0.59	11 (861)	0.30	0.60

[a]Measured by tests such as Raven's Progressive Matrices and the Cattell Culture Fair test; [b]measured by tests such as the Mill Hill Vocabulary test, ASVAB and AFQT, Differential Aptitude Test; [c]measured by the g factor from the AFTQ and the GATB, Otis-Lennon, Stanford-Binet and others; [d]these are values for middle skill jobs because there were no low skill jobs; [e]O*NET Zone 1 and 2; [f]O*NET Zone 4 and 5
k number of studies; *N* number of test takers; *r* observed correlation weighted by N; *rho* observed correlation weighted by N, corrected for both range restriction and criterion unreliability
Adapted from Postlethwaite (2011), Tables 6–14

economics (e.g., Borghans et al. 2015; Heckman and Kautz 2014) that the two are highly overlapping (correlated) but nevertheless separate and distinguishable, and that crystallized abilities are more strongly predictive of school and workplace outcomes.

2.6 General Fluid Ability Versus General Ability

In a series of studies involving school children and adolescents taking various batteries of cognitive ability tests hierarchical factor analytic models were fit to the data using both exploratory (Schmid and Leiman 1957; Undheim 1981), and confirmatory approaches (Gustafsson 1984; Undheim and Gustafsson 1987). In these analyses general fluid ability (*Gf*) was found to be indistinguishable from general ability (*g*). At some level this is not a surprising result because descriptions of general fluid ability and general ability sound similar. Spearman (1904) suggested that general ability involved the "eduction of relations and correlates" which comes close to a description of the processing involved in the progressive matrices test, a prototypical measure of *Gf*, as discussed in the previous section.

As Valentin Kvist and Gustafsson (2008) pointed out, some studies have not replicated the *g* = *Gf* finding (e.g., Carroll 2003), and others have even argued that *Gc* is closer to *g*, on the basis of the centrality of *Gc* tests in particular test batteries. It would seem that this is a difficult issue to resolve given that the makeup of the test variable set will affect the location of factors designed to account for the relationships among those variables.

However, Valentin Kvist and Gustafsson devised a novel and compelling rationale for how the *Gf* = *g* hypothesis could be tested. The basic idea is that according to investment theory, *Gf* develops into a general factor because it drives knowledge and skill acquisition in diverse domains (e.g., vocabulary acquisition, rule induction), causing correlations between performances in those diverse domains. But that relationship assumes roughly equal learning opportunities. If there are differential opportunities to learn, say, between first and second language groups, then the relationship between *g* and *Gf* will be reduced. In their words:

> This suggests a way to test both the Investment theory and the hypothesis that *g* equals *Gf*, namely through investigating the effect of differential learning opportunities for different subsets of a population on the relation between *Gf* and *g*. From the Investment theory follows the prediction that within populations which are homogeneous with respect to learning opportunities there should be a perfect relationship between *Gf* and *g*, while for populations which are composed of subgroups who have had different learning opportunities, the relation between *Gf* and *g* should be lower (p. 425).

They administered a battery of 15 tests, measuring *Gf*, *Gc*, general visualization (*Gv*), and general speediness (*Gs*) to 3570 18–60 year olds, mostly men, registered at a Swedish employment office. For the purposes of hypothesis testing the sample was divided into native speakers (N = 2358), European immigrants (N = 620) and non-European immigrants (N = 591). Hierarchical models were fit to the data with first order *Gf*, *Gc*, *Gv*, and *Gs* factors, and a second order *g* factor. When the analyses were conducted within groups, the correlation between *g* and *Gf* was 1.0, as expected, due to roughly equal opportunities to learn. But when the data were pooled, which put together groups with very different learning opportunities, then the correlation between *g* and *Gf* was 0.83. As Valentin Kvist and Gustafsson (2008) point out, the result "provides support for the Investment theory, and for the hypothesis that *Gf* is equivalent to *g*…however…only when the subjects have had approximately equally good, or equally poor, opportunities to develop the knowledge and skills measured" (p. 433).

The Valentin Kvist and Gustafsson (2008) finding is an important one for understanding the relationship between *g* and *Gf*. It also is reminiscent of an argument made almost 40 years ago by Zigler and Trickett (1978) who proposed that IQ tests measure three distinct components, formal cognitive processes, school learning, and motivation. If the school learning or motivation components are unequal, then IQ tests are poor measures of cognitive processing ability (Brent Bridgeman [personal communication, March 30, 2016] pointed this out).

2.7 Does Fluid Ability Change? Can It Be Improved?

This section addresses the issues of the natural change in *Gf* over the lifespan as well as the secular effect or Flynn effect, which is the change in population cohort *Gf* over time. The section also addresses ways to improve *Gf*, either through school, direct training, or pharmaceutically.

2.7.1 Fluid Ability and Age

It is largely accepted that fluid ability decreases, and crystallized ability increases, with age (Hunt 2011). Fluid ability is generally held to peak in the early to mid-20s before declining and crystallized ability to peak in the early 30s and remain fairly stable into the early 60s before declining (e.g., McArdle et al. 2002). Despite the general acceptance of these trends, interpretations of age-related changes in fluid ability are complicated by several factors.

The first factor is differences in research designs. It is mainly cohort-sequential designs that have found that fluid ability peaks in the early 20s (Horn 1989; Tucker-Drob and Salthouse 2011), yet these designs are prey to cohort effects, including different cohorts being differentially affected by the Flynn Effect (the secular growth in *Gf* scores over the past half century); longitudinal research suggests that fluid ability does not decline until at least the early 60s (Schaie 2012). Indeed, a recent longitudinal study actually found an *increase* in fluid ability of about 15 points from age 12 to age 52 (Schalke et al. 2013). Longitudinal studies, however, are susceptible to selective attrition, which can distort findings of construct-level changes over time. Additionally, in order to properly compare the results of cross-sectional and longitudinal studies it is important that these investigations use the same cognitive tests, otherwise differences attributed to changes in fluid ability over time may be confounded with test differences that do not reflect true differences at the construct level. It has been claimed (e.g., Horn and Noll 1997) that longitudinal studies have tended to use tasks that are better characterized as indicators of crystallized than fluid ability and after accounting for this the results of the two research traditions are well-aligned in demonstrating the early-in-the-lifespan decline in fluid ability.

A second methodological factor that must be accounted for is where the studies occurred. Much of the cross-sectional and longitudinal research that is taken to be key to understanding age-related differences in fluid ability is conducted in Western, English-speaking countries, especially the United States. For example, Schaie's (2012) results are based on the Seattle Longitudinal Study and many of the studies conducted by John L. Horn and colleagues drew participants largely from the United States; Deary et al.'s (1998) sample was Scottish. Cultural and environmental influences may moderate the association between differences in age and differences in fluid ability; Schalke et al.'s (2013) study finding an increase in fluid ability scores between 12 and 52 was conducted in Luxembourg. Instructive are investigations finding that socioeconomic status modifies the relationship between fluid and crystallized ability (Schmidt and Crano 1974) and the heritability of IQ scores (Harden et al. 2007). More attention should be paid to within- and between-country differences in changes in fluid ability over time, especially in non-Western countries. In pursuit of discovering some "universal" law (cf. Danziger 2009) governing the association between age and fluid ability it is certainly possible to average over all countries, all cultures, and all socioeconomic strata, but in terms of interpretability and usefulness doing so would be comparable to averaging a person's blood pressure readings over the course of her entire lifetime to derive her "true" blood pressure (cf. Sechrest 2005).

Finally, differences in test-taking motivation over time may further complicate interpreting studies demonstrating age-related declines in fluid ability. Research in the past several years has made clear the importance of accounting for motivation when examining test scores conducted in low-stakes settings (e.g., Duckworth et al. 2011; Liu et al. 2012). Classic accounts of proper procedures for conducting tests of individual differences (Fiske and Butler 1963; Terman 1924) emphasize the need to make the testing situation as similar to an experimental one as possible, the goal being to eliminate all between-subjects variance in all influences on test scores—except variance in the construct that test is intended to measure. It is important to remember that *an individual's* test scores are the result of many variables (e.g., eyesight, psychomotor process), but the goal when measuring individual differences is for *differences* in individuals' scores to solely reflect variance in the construct being assessed (Cronbach 1971).

When cognitive ability tests are given under high-stakes conditions it is assumed that motivation does not play a role in determining differences in test-takers' scores because they are all putting forth maximal effort (Sackett 2012). This is not the case when tests are taken under low-stakes conditions, however, where there is little incentive for individuals to put forth their full effort. It is important to ask why older test-takers in-particular would be fully motivated to perform well on tests whose defining characters include being content- and context-free (Ackerman 1999). After exiting the fairly homogenous compulsory school curriculum, people are able to exert more control over their environments and select themselves into situations that will give them access to content they find (relatively) interesting. Presenting older adults with content of no apparent real-world significance and (that is likely not intrinsically interesting to them), after they have been able to avoid such content for decades, and then asking them to fully engage with that content without offering any major incentives does not appear to be a recipe for eliciting maximal effort. By definition (Carroll 1993), cognitive ability constructs are maximal performance variables that can only be measured when test-takers are fully motivated; to the extent older adults are not fully motivated during fluid ability testing the construct validity of these assessments must be questioned.

This line of reasoning suggests that more consideration should be given to the extent to which test-taking effort plays a role in the observed decline of fluid ability with age. If fluid ability tests are consistently administered under low-stakes conditions, the major incentive for test-takers to do well on them may be reducible to their internal sense of competitiveness, perhaps manifesting in the need to "demonstrate one's full potential" or simply outscore fellow test-takers; giving test-takers an achievement motivation inventory (Freund and Holling 2011) could potentially allow for control of some of this construct-irrelevant variance. Providing test-takers with extrinsic incentives (Liu et al. 2012) is another option for inducing motivation. However, finding which incentives are most effective can be challenging [e.g., calling a test a game can increase motivation (Bridgeman et al. 1974)], and incentive manipulations can have different effects on easy versus difficult tasks (e.g., Harkins 2006).

A related topic is the hypothesis that much of the age-related loss of fluid intelligence is attributable to a declining ability to maintain focused attention. Horn (2008) summarizes several converging lines of evidence supporting this hypothesis from studies on vigilance, selective attention, Stroop (i.e., tasks requiring one to name the color of words, such as the word blue presented in a red font, where the perceptual and semantic information conflict), and distracted visual search tasks. These findings complement more recent research (e.g., Burgess et al. 2011; Melnick et al. 2013) indicating the ability to suppress distracting information and maintain concentration in the face of interference is associated with better performance on fluid ability tasks. If older adults already have more difficulty concentrating, it should come as no surprise that they score more poorly on tasks that demand intense concentration but are of no intrinsic interest to them and scores on which have no impact on their lives once they exit the testing session.

2.7.2 *Fluid Ability and the Flynn Effect*

The Flynn Effect is a label for the phenomenon of rising cognitive ability test scores over the past century, at an average rate of 0.3 points per year (Hunt 2011). These gains have been demonstrated across industrialized countries and age groups and are primarily observed on components of tests that are categorized as tapping fluid ability (e.g., progressive matrices); scores on crystallized ability tests have either remained constant or declined (Flynn 2007). Many different reasons for the Flynn Effect have been put forth, ranging from increased test sophistication to better nutrition to safer environments, but a definitive explanation has not been identified (Hunt 2011). There is evidence that the Flynn Effect has ceased, at least in some countries—yet there is also evidence that it continues, even among those with cognitive abilities in the top 1 % of the distribution (Wai and Putallaz 2011).

The Flynn Effect has important implications for increasing fluid ability because the rate at which cognitive test score gains have occurred suggest they must be environmental in origin. This implicates learning processes. Flynn (2007), Flynn & Weiss (2007) suggests that the rising scores are partially due to the fact that individuals have learned to think progressively more "scientifically" over the past century. That is, individuals know to consistently map concrete objects (e.g., dawn and dusk) onto higher-order relations (e.g., "separates night from day"), rather than simply thinking of those objects in terms of their immediate properties (e.g., "time of day", "degree of brightness"; Flynn and Weiss 2007). Although this shift in thinking may have occurred largely implicitly due to increased exposure to more formal education and more complex environments, clearly it has occurred through learning processes, suggesting it can be explicitly taught. Intriguingly, this theory calls into question the stark distinction between fluid and crystallized ability, as it posits that increases in fluid ability are rooted in *knowledge* that approaching problems using abstract reasoning tends to be an effective strategy.

2.8 Can Fluid Ability Improvement Be Accelerated?

2.8.1 *Through Schooling*

There is evidence that fluid intelligence can be improved. Ceci (1991), Ceci and Williams (1997) identified several different types of evidence consistent with the idea that schooling raises IQ. Some of these are simply observational—higher test scores accompany more time in school where differential attendance is due to starting school late, attending intermittently, dropping out before graduation, or conversely, staying in school longer to avoid the draft during the Vietnam war years. Another type of evidence is the summer slump where scores go down during the several months of summer vacation, which suggests that cognitive growth is not solely due to maturation.

Maturation and schooling are confounded, and so one approach to disentangle them is to estimate the effect of maturation by comparing ability or achievement scores of same-grade students who vary in age (because a grade will consist of individuals who range from old for the grade, such as those almost old enough to be eligible for the higher grade, to young for the grade, that is, those almost eligible to be held back to a lower grade). The difference in test scores between the relatively old and relatively young students within a grade (or the slope of the test score on age regression line) provides the age or maturation effect on test scores. Then separately comparing test scores of the oldest in a lower grade with the youngest in the next higher grade, a regression discontinuity, will provide the effect of schooling on test scores. This general approach has been used in several studies (e.g., Cahan and Cohen 1989; Stelzl et al. 1995), with a finding that the effect of schooling on test scores is twice as strong as the effect of age (Cliffordson 2010).

Another approach has been to investigate the effects of compulsory education on IQ by comparing same age students who differ in schooling due to variation in age entry requirements or mandatory attendance. An example of the latter is a study by Brinch and Galloway (2011) who noted that in the 1960s mandatory school attendance was changed from seventh to ninth grade in Norway. Different Norwegian communities enforced the change at different times, and so it was possible to compare the effects of school attendance in communities that were largely similar, in effect a natural experiment. An abilities test given to 19 year olds as part of mandatory military service allowed for an estimate of 3.7 IQ points per year. (Note that this is a convenient shorthand in this literature to express schooling's effect on the convenient, well known IQ scale; it does not imply that growth is linear across grades, as there is not sufficient data to make such a claim.)

Differential effects of type of schooling have also been investigated. Gustafsson (2001) compared performance on a mandatory military enlistment test battery given to 18 year old males who had previously gone through different tracks in secondary school (e.g., technical, natural science, vocational). The battery comprised measures of *Gf*, *Gc*, and *Gv*. He controlled for initial differences in grades (following a common pre-upper-secondary curriculum) and socioeconomic status. He found that students who had completed academic tracks had higher *Gf* scores, technical and science tracks had higher *Gv* scores, and both effects were stronger than track effects on *Gc*. Becker et al. (2012) showed similar effects of academic versus vocational tracking in Germany.

Cliffordson and Gustafsson (2008) treated test scores from measures of *Gf*, *Gc*, and *Gv* as dependent variables similar to Gustafsson (2001), but they included age and amount of schooling at the time of testing as predictors (they also included controls for socioeconomic status, background, and school grades). They found results generally consistent with previous findings, with the effect of schooling double the effects of age, a schooling effect of approximately 2.7 IQ points per year, and differences between tracks in expected directions, such as social science and economics tracks having the highest effect on the *Gf* measure (4.8 points), the technology track having the highest effect on the Technical Comprehension test (3.4 points), and only Natural Science having an effect on the *Gv* measure (1.6 points).

2.8.2 Through Working Memory Training

An intriguing experiment published several years ago had individuals practice a working memory task known as the dual n-back task (Jaeggi et al. 2008). Participants were simultaneously shown a square appearing in one of 6 locations on a computer screen, and heard a letter (e.g., “C”) at the same time. After 3 s, they were shown another square, and heard another letter. This sequence repeated indefinitely. The task was to indicate independently whether the square location and the letter were the same as they were on the previous trial (“1-back”). If they answered correctly, then the question was made more complex by asking whether the two items were the same as they were 2 trials back. The task continued to adapt (1-back, 2-back, 3-back, etc.) according to whether the respondent was correct or not. Participants engaged in this task for anywhere between 8 and 19 training sessions, and were given *Gf* pretests and posttests (e.g., Raven’s Progressive Matrices). The researchers found that treated participants (compared to no-treatment controls) performed significantly better on the *Gf* measures as a result of working memory training.

This study has been replicated a number of times and a recent meta-analysis suggested that there was consistent evidence that several weeks of working memory training, specifically based on the n-back task, transfers to fluid ability tasks (Au et al. 2015). However, another meta-analysis suggested that while working-memory training did produce reliable short-term improvements in working-memory skills, there was no evidence that working-memory training transferred to other skills such as *Gf*, attention, word decoding, or arithmetic (Melby-Lervåg and Hulme 2012). There also seems to be little evidence that “brain training” tasks of a more commercial variety transfer to fluid tests (Owen et al. 2010).

2.8.3 Through Pharmaceutical Agents

Use of pharmaceutical agents to enhance intelligence is a growing area of research (Dance 2016). One wakefulness promoting agent in particular, modafinil, which is FDA approved for treating sleeping disorders, such as narcolepsy, shift-work sleep disorder, and general sleepiness is known as a smart drug for non-sleep-deprived individuals (Geggel 2015). A recent meta-analysis on its effects showed that modafinil enhanced attention, executive functions, learning, and memory, but did not affect creativity or working memory (Battleday and Brem 2015).

2.8.4 Through Attentional Control

A common theme of many efforts to enhance fluid ability is a focus on increasing concentration and attentional control (Nisbett et al. 2012). This accords well with Horn’s (2008) hypothesis that declines in these abilities explain much of the age-related decay in performance on fluid tasks and that age is associated with deteriorating performance in jobs with intense attentional demands (Kanfer and Ackerman 2004; Sells et al. 1984). To what extent should these findings inform how we conceptualize the construct of fluid intelligence? Should we consider attentional control and concentration "part of" the fluid ability construct or simply "channels" that assist or undermine its deployment? If the former, this implies that given unlimited time individuals should be able to complete fluid ability tasks of any difficulty level, since by removing time constraints individual differences in concentration and vigilance would be eliminated. This seems absurd, however, as it further implies that individuals' basic problem-solving abilities do not practically differ once differences in their concentration have been accounted for—yet it seems unlikely that all individuals could, for example, derive complex mathematical formulae given even unlimited time.

If the ability to maintain concentrated attention for long periods is not taken as being an aspect of fluid intelligence but simply a facilitator of it this implies that many of the efforts to enhance fluid ability do not actually do so but instead merely allow people to more fully take advantage of their current abstract reasoning skills. Assume that performance on a reasoning test is a function of both current abstract reasoning skills and ability to maintain concentrated attention for long periods, which could be a kind of motivation or personality effect. Perhaps one of the major reasons that fluid ability scores increase with each passing grade but decline with age after leaving school is that schooling implicitly trains people to concentrate their attention for long periods of time on content that they do not necessarily find particularly interesting—and the effects of this training decay after individuals have completed their compulsory schooling and are able to exert more control over their environments and choose content they find more intrinsically interesting to interact with. This line of reasoning suggests that training non-cognitive skills such as self-regulation and self-discipline (Nisbett et al. 2012) could increase scores on fluid ability tasks—but also that such training does not enhance individuals' fluid ability itself, merely the extent to which they are able to take advantage of it.

2.9 Conclusions

General fluid ability is an important and influential concept in psychology, in education, and in policy. The purpose of this chapter was to address the issue of its nature, its measurement, how and whether it is distinguishable from other abilities, such as crystallized ability and general ability, and how it can be improved. Jan-Eric Gustafsson has made key contributions to our understanding of fluid ability with respect to all these topics.

In this chapter we reviewed what we know and what we are learning about fluid intelligence. Fluid ability is the ability to solve problems in novel contexts, using deductive or inductive reasoning such as in letter series problems, or with progressive matrices problems. It is contrasted with crystallized ability, which reflects the ability to apply knowledge acquired in school or through acculturation, as reflected in vocabulary and reading comprehension tests. Fluid ability is sometimes empirically indistinguishable from general cognitive ability, although this depends on test takers having roughly comparable opportunities to learn. Fluid ability peaks earlier than crystallized ability over the lifespan. Test scores on measures of fluid ability have increased in successive cohorts over the past 50 years, a phenomenon known as the Flynn effect, although there is some indication that this is no longer happening, at least in the most developed countries. Fluid ability is highly correlated with working memory capacity, and there is some suggestion that working-memory training, particularly on the n-back task, may transfer to performance on fluid ability tasks. There is evidence from various sources that schooling may improve fluid ability, although much of the evidence is based on observational data. There is also some evidence that particular school tracks, such as academic, and social science, may be particularly associated with improvements in fluid ability. There also is some, albeit mixed evidence that pharmaceutical agents, particularly a wakefulness promoting agent, modafinil, improve fluid ability. There are other influences on test scores besides abilities, such as motivation and attention, and these may be the factors responsible for some of the improvements in fluid ability test scores due to schooling, training, and other variables.

Fluid ability is now a firmly established construct in education, psychology, and the social sciences more generally. It is likely to continue to draw research attention into the foreseeable future just as it has over the past 50 years.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process personality, interests, and knowledge. *Intelligence, 22*, 229–259.

Ackerman, P. L. (1999). Traits and knowledge as determinants of learning and individual differences: Putting it all together. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 437–460). Washington, DC: American Psychological Association.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30–60.

Arendasy, M., & Sommer, M. (2005). Using psychometric technology in educational assessment: The case of a schema-based isomorphic approach to the automatic generation of quantitative reasoning items. *Learning and Individual Differences, 17*(4), 366–383.

Arendasy, M., & Sommer, M. (2012). Using automatic item generation to meet the increasing item demands of high-stakes assessment. *Learning and Individual Differences, 22*, 112–117.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review, 22*(2), 366–377.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839. doi:10.1038/nrn1201

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Battleday, R. M., & Brem, A. K. (2015). Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: A systematic review. *European Neuropsychopharmacology, 25*, 1865–1881.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*, 682–699. doi:10.1037/a0027608

Birney, D. P., & Bowman, D. B. (2009). An experimental-differential investigation of cognitive complexity. *Psychology Science Quarterly, 51*(4), 449–469.

Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008, February). *The economics and psychology of personality traits* (NBER Working Paper No. 13810). Cambridge, MA: National Bureau of Economic Research.

Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2015). What do grades and achievement tests measure? Unpublished manuscript, Human Capital and Economic Working Group, University of Chicago, Chicago, IL. Retrieved from https://hceconomics.uchicago.edu/sites/default/files/3_BGHH-Grades-Achievement-Measure-Paper_2015-12-30a_sjs.pdf

Boring, E. G. (1923). Intelligence as the tests test it. *New Republic, 36*, 35–37.

Bridgeman, B., Strang, H. R., & Buttram, J. (1974). “Game” versus “test” instructions for the WISC. *Journal of Educational Measurement, 11*, 285–287.

Brinch, C. N., & Galloway, T. A. (2011). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences of the United States of America, 109*(2), 425–430. doi:10.1073/pnas.1106077109

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General, 140*(4), 674–692.

Buros, O. K. (Ed.). (1965). *The sixth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press. ISBN 0-910674-06-X.

Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60*(5), 1239–1249.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404–431.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Boston, MA: Pergamon.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action.* New York: North-Holland.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15*, 139–164.

Ceci, S. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*(5), 703–722.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52*, 1051–1058.

Chuderski, A. (2015). The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure. *Personality and Individual Differences, 85*, 98–104.

Cliffordson, C. (2010). Methodological issues in investigations of the relative effects of schooling and age on school performance: The between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. *Educational Research and Evaluation: An International Journal on Theory and Practice, 16*(1), 39–52.

Cliffordson, C., & Gustafsson, J.-E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence, 36*, 143–152.

Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (Eds.). (2007). *Variations in working memory*. Oxford, England: Oxford University Press.

Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., et al. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ: Lawrence Erlbaum.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.

Dance, A. (2016). A dose of intelligence. *Nature, 536*, S2–S3.

Danziger, K. (2009). The holy grail of universality. In T. Teo, P. Stenner, A. Rutherford, E. Park, & C. Baerveldt (Eds.), *Varieties of philosophical and practical concerns* (pp. 2–11). Concord, Ontario, Canada: Captus University Publications.

Deary, I. J., Starr, J. M., & MacLennan, W. J. (1998). Fluid intelligence, memory and blood pressure in cognitive aging. *Personality and Individual Differences, 25*(4), 605–619.

Diehl, K. A. (2002). Algorithmic item generation and problem solving strategies in matrix completion problems. Unpublished doctoral dissertation, University of Kansas.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 7716–7720.

Embretson, S. E. (2002). Generating abstract reasoning items with cognitive theory. In S. Irvine & P. Kyllonen (Eds.), *Generating items for cognitive tests: Theory and practice*. Mahwah, NJ: Erlbaum.

Finn, A. S., Kraft, M., West, M. R., Leonard, J. A., Bish, C., Martin, R. E., … Gabrieli, J. D. E. (2014). Cognitive skills, student student achievement tests and schools. *Psychological Science, 25*, 736–744.

Fiske, D. W., & Butler, J. M. (1963). The experimental conditions for measuring individual differences. *Educational and Psychological Measurement, 26*, 249–266.

Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn effect*. Cambridge, MA: Cambridge University Press.

Flynn, J. R., & Weiss, L. G. (2007). American IQ gains from 1932 to 2002: The WISC subtests and educational progress. *International Journal of Testing, 7*, 209–224.

Freund, P. A., & Holling, H. (2011). How to get really smart: Modeling retest and training effects in ability testing using computer-generated figural matrix items. *Intelligence, 39*, 233–243.

Geggel, L. (2015). 'Smart drug' modafinil actually works, study shows. *Live Science*. Retrieved April 1, 2016 from http://www.livescience.com/51919-modafil-improves-attention-learning.html

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.

Gustafsson, J.-E. (1998). Hierarchical models of individual differences in cooperative abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35–71).

Gustafsson, J.-E. (1999). Measuring and understanding G: Experimental and correlational approaches. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 275–289). Washington, D. C.: American Psychological Association.

Gustafsson, J.-E. (2001). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *International Education Journal, 2(4)*, 166–186.

Gustafsson, J.-E. (2002). Measurement from a hierarchical point of view. In H. L. Braun, D. G. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 73–95). Mahwah, NJ: Erlbaum.

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

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803–865.

Harden, K. P., Turkheimer, E., & Loehlin, J. C. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics, 37*(2), 273–283.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91*(3), 436–455. doi:10.1037/0022-3514.91.3.436

Heckman, J. J., & Kautz, T. (2014). Achievement tests and the role of character in American life. In J. J. Heckman, J. E. Humphries, & T. Kautz (Eds.), *The myth of achievement tests: The GED and the role of character in American life* (pp. 3–56). Chicago, IL: University of Chicago Press.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York, NY: The Free Press (Simon & Shuster).

Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29–73). Urbana, IL: University of Illinois Press.

Horn, J. L. (2008). Spearman, g, expertise, and the nature of human cognitive capability. In P. C. Kyllonen, R. D. Roberts, & L. Stankov (Eds.), *Extending intelligence: Enhancement and new constructs* (pp. 185–230). New York: LEA.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53–91). New York, NY: Guilford.

Hornke, L. F., & Habon, M. W. (1986). Rule-based item bank construction and evaluation within the linear logistic framework. *Applied Psychological Measurement, 10*, 369–380.

Hunt, E. (2011). *Human intelligence*. New York, NY: Cambridge University Press.

Irvine, S. H. (2002). The foundations of item generation for mass testing. In S. H. Irvine & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 3–34). Mahwah, NJ: Lawrence Erlbaum.

Jacobs, P. I., & Vandeventer, M. (1972). Evaluating the teaching of intelligence. *Educational and Psychological Measurement, 32*(2), 235–238.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America, 105*(19), 6829–6833. doi:10.1073/pnas.0801268105

Jensen, E. (1998). *The g factor: The science of mental ability (Human evolution, behavior, and intelligence).* Westport, CT: Praeger.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.

Kanfer, R., & Ackerman, P. L. (2004). Aging, adult development, and work motivation. *The Academy of Management Review, 29*(3), 440–458.

Kaufman, J. C. (2012). Self estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences, 22*, 118–122.

Kyllonen, P. C. (1995). CAM: A theoretical framework for cognitive abilities measurement. In D. K. Detterman (Ed.), *Current topics in human intelligence* (Vol. 4, pp. 307–359). New York, NY: Springer.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389–433.

Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics, 3*, 101–128.

Liu, O. L., Bridgeman, B., & Adler, R. (2012). Learning outcomes assessment in higher education: Motivation matters. *Educational Researcher, 41*, 352–362.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology, 38*(1), 115–142.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York, NY: Guilford.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–182). New York, NY: Guilford.

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270–291.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., … Tetlock, P. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science, 10*(3), 267–281.

Melnick, M. D., Harrison, B. R., Park, S., Bennetto, L., & Tadin, D. (2013). A strong interactive link between sensory discriminations and intelligence. *Current Biology, 23*(11), 1013–1017.

Meo, M., Roberts, M. J., & Marucci, F. S. (2007). Element salience as a predictor of item difficulty for Raven's progressive matrices. *Intelligence, 35*(4), 359–368.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12*, 252–284.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., et al. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*(2), 130–159.

Oberauer, K., Schulze, R., Wilhelm, O., & Suess, H.-M. (2005). Working memory and intelligence—Their correlation and relation: Comment on Ackerman, Beier, and Boyle (2005). *Pscyhological Bulletin, 131*, 61–65.

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., et al. (2010). Putting brain training to the test. *Nature, 465*, 775–778.

Pind, J., Gunnarsdóttir, E. K., Jóhannesson, H. S. (2003). Raven's standard progressive matrices: New school age norms and a study of the test's validity. *Personality and Individual Differences, 34*, 375–386.

Postlethwaite, B. E. (2011). *Fluid ability, crystallized ability, and performance across multiple domains: A meta-analysis*. Doctoral dissertation. Retrieved from http://ir.uiowa.edu/etd/1255/

Preckel, F. (2003). *Diagnostik intellektueller Hochbegabung. Testentwicklung zur Erfassung der fluiden Intelligenz [Assessment of intellectual giftedness. Development of a fluid intelligence test]*. Göttingen, Germany: Hogrefe.

Preckel, F., & Thiemann, H. (2003). Online- versus paper-pencil-version of a high potential intelligence test. *Swiss Journal of Psychology, 62*, 131–138.

Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence, 20*, 41–70.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*, 667–706.

Sackett, P. R. (2012). Cognitive tests, constructs, and content validity: A commentary on Schmidt (2012). *International Journal of Selection and Assessment, 20*(1)

Schaie, K. W. (2012). *Developmental influences on adult intelligence: Seattle Longitudinal Study* (2nd ed.). New York, NY: Oxford University Press.

Schalke, D., Brunner, M., Geiser, C., Preckel, F., Keller, U., Spendler, M., et al. (2013). Stability and change in intelligence from age 12 to age 52: Results for the Luxembourg MAGRIP study. *Developmental Psychology, 49*, 1529–1543.

Schank, R. C. (1984). How much intelligence is there in artificial intelligence? *Intelligence, 4*(1), 1–14.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61.

Schmidt, F. L., & Crano, W. D. (1974). A test of the theory of fluid and crystallized intelligence in middle- and low-socioeconomic-status children: A cross-lagged panel analysis. *Journal of Educational Psychology, 66*(2), 255–261.

Schrank, F. A. (2006). *Specification of the cognitive processes involved in performance on the Woodcock-Johnson III* (Assessment Service Bulletin No. 7). Itasca, IL: Riverside.

Sechrest, L. (2005). Validity of measures is no simple matter. *Health Services Research, 40*(5), 1584–1604.

Sells, S. B., Dailey, J. T., & Pickrel, E. W. (Eds.). (1984). *Selection of air traffic controllers*. Report no. FAA-AM-84-2. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2., pp. 47–103). Hillsdale, NJ: Lawrence Erlbaum.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201–292.

Spearman, C. (1924). The nature of "intelligence" and the principles of cognition. *Journal of Philosophy, 21*(11), 294–301.

Spearman, C. (1927). *The abilities of man.* London, England: Macmillan and Co.

Stankov, L., & Schweizer, K. (2007). Raven's progressive matrices, manipulations of complexity and measures of accuracy, speed and confidence. *Psychology Science, 49*, 326–342.

Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and crystallized intelligence: A quasi-experimental study. *Intelligence, 21*, 279–296. doi:10.1016/0160-2896(95)90018-7

Terman, L. M. (1924). The mental test as a psychological method. *Psychological Review*, 31, 93–117.

Thorsen, C., Gustafsson, J.-E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology, 84*(4), 556–570. doi:10.1111/bjep.12041

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric monographs, 1*, ix–121.

Tucker-Drob, E. M., & Salthouse, T. A. (2011). Individual differences in cognitive aging. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell handbook of individual differences* (pp. 242–267). Malden, MA: Wiley-Blackwell.

Undheim, J. O. (1981). On Intelligence III: Examining developmental implications of Catail's broad ability theory and of an alternative neo-Spearman model. *Scandinavian Journal of Psychology, 22*(1), 243–249.

Undheim, J. O., & Gustafsson, J.-E. (1987). The hierarchical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL). *Multivariate Behavioral Research 22*, 149–171.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory. *Cognitive Psychology, 71*, 1–26.

Valentin Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 36*, 422–436.

Wai, J., & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence, 39*(6), 443–455.

Wilhelm, O. (2000). *Psychologie des schlussfolgernden Denkens: Differentialpsychologische Prfung von Strukturuberlegeungen [Psychology of reasoning: Testing structural theories].* Hamburg, Germany: Dr. Kovac.

Wilhelm, O. (2006). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373–392). Los Angeles: Sage. doi:http://dx.doi.org/10.4135/9781452233529.n21

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside.

Zigler, E., & Trickett, P. K. (1978). IQ, social competence, and the evaluation of early childhood intervention programs. *American Psychologist, 33*, 789–798.

Springer
http://www.springer.com/978-3-319-43472-8

Cognitive Abilities and Educational Outcomes
A Festschrift in Honour of Jan-Eric Gustafsson
Rosén, M.; Yang Hansen, K.; Wolff, U. (Eds.)
2017, XXXIV, 327 p. 43 illus., Hardcover
ISBN: 978-3-319-43472-8