Chapter 14

Intelligence

OLIVER WILHELM & ULRICH SCHROEDERS

Ulm University & University of Kassel

Most other chapters in this volume tackle the nature of human thinking from the perspective of cognitive psychology, for example how humans derive deductions. In most of these chapters, human subjects are treated uniformly; that is, an attempt is made to describe and explain the cognitive principles of deductive reasoning that are common to all people. In this chapter, the focus will instead be placed on what makes people different: individual differences between persons. The areas of cognition discussed in most other chapters provide the required background information on what exactly humans engage in while working on an intelligence test. Whereas some measures stress deductive inference, others might provoke complex problem solving behavior. The focus of this chapter is to study why some subjects answer those intelligence items correctly while others get them wrong, and why these differences are meaningful and interesting.

In the first section of this chapter, we will approach the concept of intelligence by briefly summarizing the history of relevant psychometric intelligence models. While a historical overview might seem somewhat inappropriate in an introductory chapter, in the case of intelligence research, this perspective provides us with a set of competing accounts essential for understanding intelligence data and intelligence theories. We will then proceed by describing an established taxonomy of intelligence factors and discuss intelligence as an overarching concept for all measures that provoke maximal cognitive effort. The second section will be very pragmatic, showing how intelligence can be measured, how it can be used for predictive purposes, and whether it can be changed through interventions. We will conclude the chapter by examining important issues for future intelligence research. A very broad and relatively developed field such as intelligence and its assessment cannot be addressed exhaustively in an introductory chapter. We hope that the references provided in this chapter will be helpful for further reading and will enrich one's understanding of contemporary research in the field.

14.1 Understanding Intelligence

Research on the structure of individual differences in intelligence follows an atypical strategy, relative to most other psychological research. Typically, theories and hypotheses are proposed, followed by the development of adequate means for testing and evaluating; intelligence research instead proceeds in reverse. For instance, in the beginning of intelligence research as an independent field, factor analytic methods were invented and refined, with corresponding theories of intelligence developed afterwards. This approach to intelligence research places the focus on competing explanations of individual differences in a broad variety of intelligence tasks. Unfortunately, we must skip some important early contributions: for example, Galton (1883) developed several simple tests of intellectual function-

https://doi.org/10.17885/heiup.470.c6677

ing and made early contributions to the heredity of intelligence. Binet deserves credit for compiling one of the original intelligence tests, although his efforts could hardly be considered a clear theoretical contribution on the structure of intelligence. Moreover, Ebbinghaus (1895) developed several intelligence tests that were reused in other fields before making a much later comeback in intelligence research (Ackerman, Beier, & Bowen, 2000).

14.1.1 The History of Intelligence Models and the Usual Suspects

In the following section, we will present different ways of conceptualizing intelligence (see Figure 14.1). We start with Spearman (1904a, 1904b), who made two seminal methodological contributions in the year he completed his dissertation (on a completely different topic) with Wilhelm Wundt in Leipzig. In one of these contributions, he laid the foundation for what is known today as classical test theory (Lord & Novick, 1968). In his other contribution, he established the groundwork for the general factor theory. This theory is based on two central assumptions. First, a latent factor (*g*) accounts for the correlations between all intelligence tasks. Second, besides this general factor, there are test-specific individual differences. Apart from these two components, there is only unsystematic measurement error in intelligence tasks.

The first assumption is an idea prevalent throughout research on individual differences and applies to traits such as extraversion and achievement motivation. The assumption is that there is a stable disposition within persons to act in specific ways. In the case of the *g*-factor, this disposition is to do well

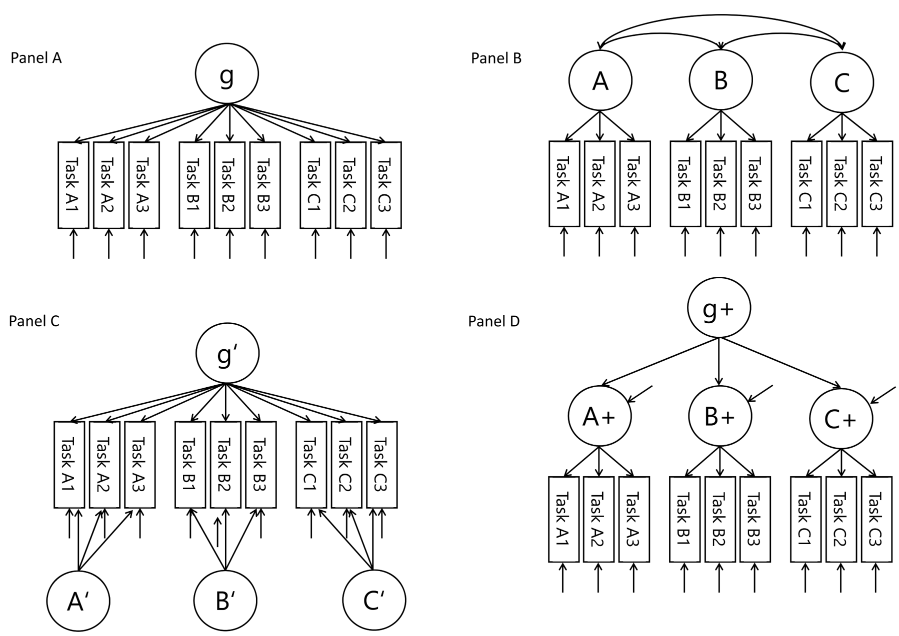

Figure 14.1: Psychometric models of intelligence.

on tasks requiring cognitive effort. This disposition is deemed causal for the correctness or swiftness of responses on each intelligence item. The ideas of a latent trait also apply to most other theories of intelligence structure. Usually, these traits are deemed stable over time, that is, the rank-order of subjects does not change dramatically over time. They are considered broad, in the sense that they do not only apply to a highly-specific test but also to similar examinations. They are expected to be relevant, meaning they predict real-life outcomes that are of individual or societal relevance. In the case of the *g*-factor, only one such latent variable is specified for the field of intelligence. Spearman's theory states that the correlation between any two intelligence tasks is because of the *g*-factor (Figure 14.1, panel A). As a side note, the g-factor theory competed in the early days of intelligence testing with the so-called bond theory (Thomson, 1919). The bond theory stated that the magnitude of the correlation between any two intelligence tasks indicates the proportion of overlapping processes—the higher the correlation, the larger the number of shared processes. Somewhat more cognitively, the componential theory of intelligence proposes that the correlation between two intelligence tasks is a function of shared components and the theory was put to the test for example in the area of analogical problem solving (Sternberg, 1977). This approach has recently gained new attraction (van der Maas et al., 2006). A somewhat related approach pursued with different methods is called facet theory. Here, the overlap of task attributes determines the correlations between intelligence tasks, while the magnitude of correlations is graphically represented by proximity (Guttman & Levy, 1991).

A major competitor of the *g*-factor theory arose with the development of multiple factor analysis (Thurstone, 1931). This procedure allowed for the extraction of more than one factor (Figure 14.1, panel B). Combined with the development of factor rotations, the interpretation of intelligence factors was greatly facilitated (Thurstone, 1934). Thurstone subsequently proposed 7 primary mental abilities (Thurstone, 1938). Thurstone initially proposed reasoning, spatial visualization, verbal comprehension, word fluency, number facility, associative memory, and perceptual speed—and he updated and prolonged this list jointly with his wife three years later (Thurstone & Thurstone, 1941).

Disentangling different contributions to performance on intelligence tasks was also the main purpose of the so-called bifactor approach (Holzinger & Swineford, 1937). Similarly, Schmid and Leiman (1957) proposed rotation techniques to distinguish between independent performance contributions to individual differences in intelligence tasks. Both approaches (Figure 14.1, panel C) are early hierarchical perspectives on intelligence.

Higher-order factor models are another way to conceptualize intelligence because the ubiquitous positive correlation between any two intelligence tasks also leads to correlations between intelligence factors. These factor correlations are the basis for higher-order models of intelligence (Figure 14.1, panel D). In these models, a second-order factor accounts for the correlations between first-order factors, which in turn accounts for the correlations between intelligence tasks (Carroll, 1993).

14.1.2 Accepted Views on the Structure of Intelligence

Among the more contemporary models, Cattell's theory of fluid and crystallized intelligence (Cattell, 1971; see also Brown, 2016) has become a widely accepted and applied model for the description and testing of intelligence. The gf-gc-theory also heavily stimulated theory building, as can be seen in the investment theory (Cattell, 1971) or the PPIK theory (Intelligence-as-Process, Personality, Interests and Intelligence-as-Knowledge, Ackerman, 1996). Furthermore, the integration of the gf-gc-theory into personality research and its validation and use in aging research has contributed to its popularity. In the current version, the gf-gc theory assumes nine primary factors (McGrew, 2009), of which fluid and crystallized intelligence are central (see Table 14.1).

A closely related milestone in intelligence research is the seminal work of Carroll (1993). The comprehensive synopsis and reanalysis of decades of factor-analytic intelligence research and the theory-guided integration of these findings led to a structural model that, in view of the factors postulated, bears much resemblance to the model of

Table 14.1: Overview of the central factors of cognitive ability.

	Label	Description	Example Task
gf	Fluid intelligence	Reason, plan, solve abstract and complex problems; basically the ability to maintain, to mentally manipulate, and to store information; strong link with working memory capacity.	Number series
gc	Crystallized intelligence	Describes the breadth and depth of cultural knowledge that is passed on to the individual through acculturation (e.g., formal learning). Is often measured with (and reduced to) verbal ability indicators, predominantly vocabulary tasks.	Vocabulary
gsm	Short term memory	Retain and maintain a limited amount of information for a short period of time.	Memory span
gv	Visual processing	Perceive, manipulate, store, and retrieve visual images such as shapes, forms, colors, etc., and more complex visual stimuli. This also includes spatial orientation, transformation, and moving visual objects.	Spatial relations
ga	Auditory processing	Analyze, manipulate, understand, and synthesize sound elements, sound groups, and sound patterns. The key feature is the cognitive control in perception of auditory material (i.e., handle the competition between signal and noise).	Speech sound discrimination
glr	Long-term memory and retrieval	Store and consolidate new information in long-term memory. Fluently retrieve stored information (e.g., concepts, ideas, items, names).	Word fluency
gs	Processing speed	Perform over-learned or elementary cognitive tasks under time constraints, high efficiency (i.e., attention and focused concentration) is necessary.	Perceptual speed
gt	Reaction and decision speed	Quickly make elementary responses (i.e., simple reaction time) or several elementary responses (i.e., complex reaction time) when simple stimuli are presented.	Simple reaction task

Note. Labels in the first column are taken from the CHC model.

Cattell and Horn. Carroll (1993) reanalyzed 461 data sets from factor analytic intelligence research including diverse populations, countries, decades, and a full variety of cognitive tasks developed by that time. To this day, Carroll most likely compiled the most comprehensive overview of cognitive ability measures. His analyses led to a structural model distinguishing three levels of generality (see Figure 14.2).

At the middle level of generality, eight broad ability factors are distinguished (see Table 14.1). Once again, any two intelligence tasks will always show a positive correlation and these eight factors will therefore show positive manifold. This positive manifold is captured with an overarching general intelligence factor at the apex of the higher-order model of intelligence. Such models have become more prevalent and popular recently (e.g., Gustafsson 1999), because they a) explicitly address and capture the substantial positive correlations between intelligence tasks and intelligence factors, and b) deliver the best from the two worlds of group factor theories and a general factor theory. In pragmatic terms, the factors from the middle level of generality are not all of equal importance. Whereas fluid and crystallized intelligence are indispensable in intelli-

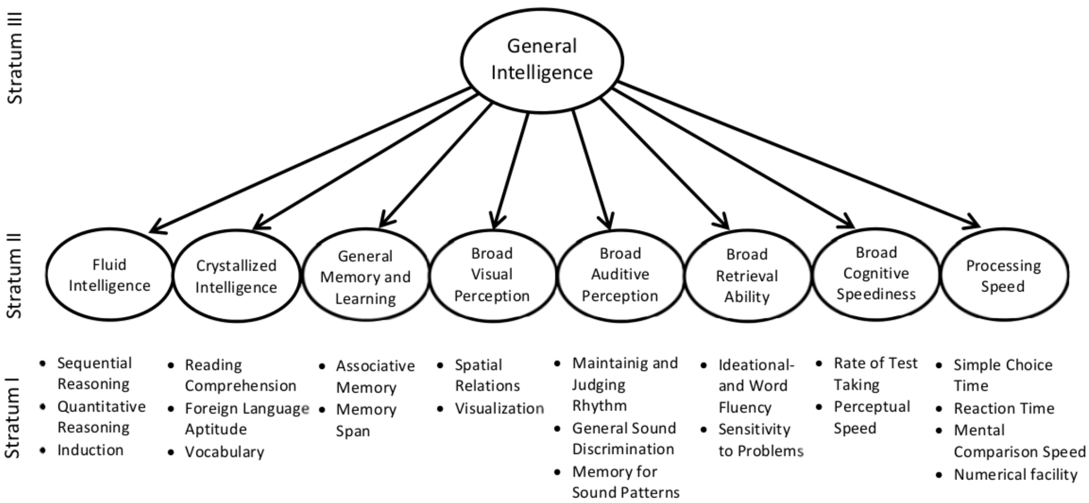

Figure 14.2: A slightly revised version of Carroll's Three-Stratum Theory.

gence tests, other factors are mostly needed to give a comprehensive picture of an individual's cognitive abilities. Unsurprisingly, fluid and crystallized intelligence (and mixtures of both factors) are also most predictive for outcomes such as educational achievement or job performance. Please note that fluid intelligence has been found repeatedly to show the strongest relation with the overarching general factor. Therefore, if only a single task can be used to measure intelligence, your choice should be to pick a fluid intelligence task.

At the lowest level of the hierarchy there are many specific intellectual abilities that serve to underline the breadth of factors at the middle level and to illustrate the exhaustiveness of the model. Taken together, the work of Cattell, Horn, and Carroll by and large converges on the model shown in Figure 14.2. The discussion of research on this model integrates and successively extends the common ground on individual differences in intelligence (McGrew, 2009). In the current version of the model, more specific abilities, such as specialized knowledge in the sense of expertise or reading and writing skills, have been included. Importantly, in the last two decades, popular and frequently used intelligence tests switched to the Cattell-Horn-Carroll (CHC) model—a change that was desperately needed for various Wechsler-tests in particular.

Despite its unifying character, the CHC model must not be misunderstood as a final model of intelligence structure. There are many open questions, some of which we will discuss in later sections of this chapter. In addition, our presentation of intelligence relies on psychometric, mainly factor-analytical approaches for studying individual differences in cognitive abilities. However, we want to mention that there are several theories of intelligence that cannot be given full consideration in the course of an introductory chapter. A theory that is popular, especially among educators and teachers, is the theory of "Multiple Intelligences" by Gardner (1983, 1991) who advocated against g, proposed distinct forms of intelligence and claimed that students can be categorized in eight different types of learners (i.e., visual-spatial, bodily-kinesthetic, musical-rhythmic, interpersonal, intrapersonal, verbal-linguistic, logical-mathematic, naturalistic). However, multiple intelligences appear to be a blend of g, broad ability factors below g, and other non-cognitive factors (Visser, Ashton, & Vernon, 2006) and there is no adequate empirical evidence to justify incorporating learning-styles into education (Pashler, McDaniel, Rohrer, & Bjork, 2008).

The concept of emotional intelligence has also gained considerable attention (Salovey, Mayer, & Caruso, 2004) and received substantial criticism (Davies, Stankov, & Roberts, 1998). It is argued to comprise the abilities to perceive emotions, the abilities to access, generate, and use emotions, the abilities to understand and regulate emotions and finally to enclose knowledge about emotions (Salovey et al., 2004). For most of these abilities it is difficult to come up with an unequivocal response standard, i.e. what might work to regulate Persons A's emotions might be counterproductive for person B. Nevertheless, recent efforts to include some aspects of emotional intelligence into a higher-order model of intelligence were successful (MacCann, Joseph, Newman, & Roberts, 2014) and future research in this area might be promising.

14.1.3 Intelligence as Overarching Concept of Maximal Cognitive Effort

Our discussion of intelligence has yet to include an actual, clear definition of intelligence. Indeed, prior attempts of specifying what intelligence is and what it is not were of limited success. The infamous definition that intelligence is what the test measures (Boring, 1923) begs the question of which tasks or factors of intelligence are indispensable and what should not be part of the concept "intelligence". In response to public controversy over the term intelligence, Gottfredson and 52 other researchers (1997, p.13) gave a very broad definition of intelligence: "A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience." Similarly, Neisser et al. (1996, p. 77) defined intelligence as individual differences between persons "[...] in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought." The essential components of these definitions center on aspects of fluid intelligence and neglect other factors we described above. In addition, both definitions are opaque with respect to concepts such as ability, achievement, aptitude, competence, proficiency, talent, skill, and so on (Schroeders, 2018). Some of these terms are used in specific research traditions or serve to set a specific focus. For example, *competency* or *proficiency* are preferred in an educational setting because within the spectrum of abilities under consideration, the abilities trained in formal learning (e.g., schooling) are perceived as malleable and acquired. In contrast, *talent* often labels an inherited or an exceptional ability (e.g., musical or artistic talent). The subtle nuances between these concepts, which are all related to effortful cognitive processing, are best seen in the context of the research tradition from which they originate. If you were asked to classify existing measures of intelligence, competence, aptitude, skill, etc., you would hardly be able to come up with a dependable classification of tasks. Therefore, these terms should be characterized as "fuzzy" and insufficient when it comes to explaining relations between tasks or to assigning tasks to factors.

In order to derive a dependable and inclusive understanding of what constitutes an intelligence task, we recommend using intelligence as an overarching concept of maximal cognitive effort. The distinction between typical behavior and maximum cognitive performance dates back to Cronbach (1949): typical behavior refers to the ways individuals usually behave and what they like or dislike. It is usually captured through self-reports on behaviors, preferences, and valences. For example, the question "Do you like solving math puzzles?" arguably describes an individual's preference for engaging in mathematical problem solving. Responses to such questions presuppose the willingness and ability of subjects to introspect. As well, these responses are very vulnerable to subjective judgments and biases (e.g., social desirability). In contrast, maximal cognitive performance refers to the measurement of abilities, achievements, skills, talents, etc. An item such as, "What is the solution to $f(x) = 3x^2 + 12$?" differs fundamentally from the assessment of typical behavior in several ways. Items of maximal behavior will only be used in contexts in which a) the person being examined is aware of the performance appraisal, b) the person is willing and able to show maximal cognitive effort, and c) the standards for evaluating the response behavior are adequate for the purpose of making a diagnostic judgment (Sackett, Zedeck, & Fogli, 1988). Preferably, objectively correct solutions are used as a benchmark for actual response behavior. In some domains, providing a veridical response standard is not feasible. For example, it is very difficult to provide such a standard for written essays and tasks designed to tap into interpersonal intelligence factors such as understanding emotions (for a recent though incomplete summary concerning intrinsically personal tests, see Mayer, 2018). Rather, these tasks often rely on situational judgment methodology (Oostrom, DeSoete, & Lievens, 2015).

One important aspect that we want to stress is the unfortunate division between psychological and educational testing of maximal effort concepts. More than a century ago, Binet (1904) distinguished between medical, pedagogical, and psychological methods in intelligence testing. The medical method aims "to appreciate the anatomical, physiological, and pathological signs of inferior intelligence" (Binet, 1904, p. 194). Thus, this method will receive no further consideration in this chapter. The psychological method "makes direct observations and measurements of the degree of intelligence" (Binet, 1904, p. 194) and focuses on reasoning and memory-related abilities. The pedagogical method "aims to judge intelligence according to the sum of acquired knowledge" (Binet, 1904, p. 194). It is clear in our earlier presentation of essential intelligence factors that the psychological and the pedagogical method roughly correspond to fluid and crystallized intelligence respectively. This early distinction by Binet, unfortunately, led to a subsequent separation of efforts related to his two methods. Consequently, fluid intelligence or equivalent concepts such as decontextualized thinking, academic intelligence, etc., are hardly accepted determinants of educational outcomes and have often been considered taboo in an educational context. Conversely, elaborating on crystallized intelligence or related concepts such as expertise and how they could enrich cognitive ability testing has yet to become popular in psychometric research contexts. Unfortunately, the separation between these two fields has yet to be overcome. As a remedy, we propose that the term intelligence be used as an overarching concept that encompasses mechanical abilities such as fluid intelligence, memory, and processing speed, as well as knowledge-driven aspects, such as crystallized intelligence with its myriad of facets.

Next, we want to relate intelligence assessment with educational assessment to illustrate the overarching/unifying aspect of intelligence. The debate regarding the extent intelligence tests and educational achievement tests measure the same underlying abilities has a long history (Baumert, Lüdtke, Trautwein, & Brunner, 2009). We propose that the problem of distinguishing between intelligence tests and other measures for assessing cognitive abilities (e.g., educational achievement tests) is not whether a person's scores on both methods are perfectly correlated (Bridgeman, 2005). To understand differences between both fields, it is more instrumental to study attributes in which such measures differ: for example, where they are located on the continuum "decontextualized" vs. "contextualized" and which predictions the contextualization of measures affords (Brunner, 2008). This approach clearly places the competencies studied in educational psychology below crystallized intelligence. For example, Baumert and colleagues (2009) suggested that international education studies, such as PISA (Program for International Student Assessment), primarily capture the cumulative outcomes of a knowledge acquisition process. This understanding of competence is broadly identical to Cattell's definition of crystallized intelligence (1971), according to which crystallized intelligence encompasses the totality of knowledge that people acquire and use to solve problems throughout their lives. Whereas the nature and content of educational tests are usually carefully studied, many traditional tests of crystallized intelligence neglect **content validity**—a lesson that can and should be learned from educational testing.

We advise against relying on a test's purpose to understand what the test measures. College admission tests do not measure the ability to study. Such tests usually include measures of fluid intelligence along with domain-specific crystallized intelligence tasks. School readiness tests do not capture the ability to attend school—instead, they are best seen as a composite of gc tasks and social skills. If you want to understand what a measure of maximal cognitive performance captures, it is not wise to focus on the purpose of testing. Instead, it will be more useful to classify a measure according to the intelligence factors described here.

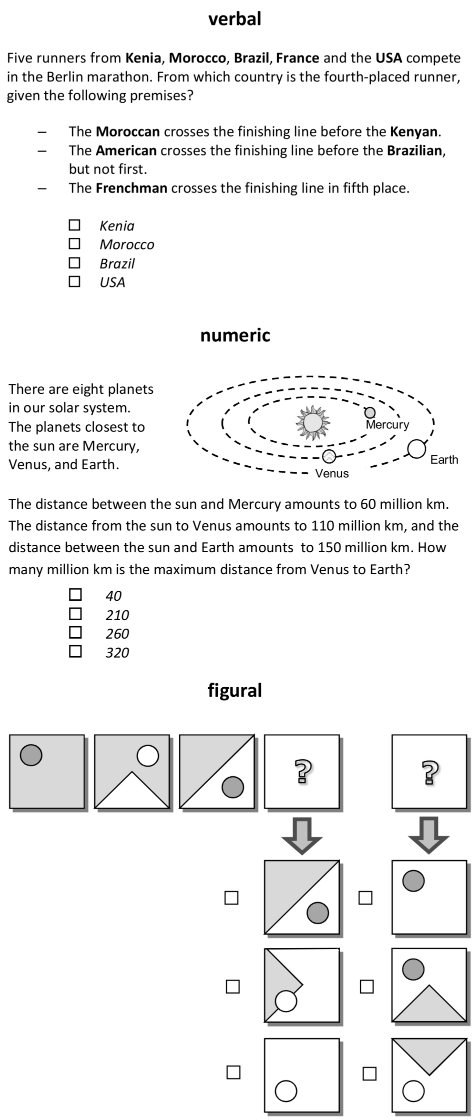

Figure 14.3: Example item for fluid intelligence: verbal, numeric, and figural.

14.2 Measuring and Using Intelligence

14.2.1 Tasks for Measuring Intelligence

In this section, we introduce selected intelligence tasks designed for use with adults and discuss the cognitive demands of these tasks. We focus on the two factors of fluid and crystallized intelligence because these factors are the most decisive and important predictors in most applied settings, such as college admission or personnel selection.

14.2.1.1 Tasks for Measuring Fluid Intelligence

Earlier in this chapter, we argued that a **fluid intelligence** task should be chosen when only a single task can be used to measure intelligence. Such a fluid intelligence task would then serve as a marker task for intelligence. Below the fluid intelligence factor, Carroll (1993) distinguished three reasoning factors:

- *Sequential reasoning* tasks require drawing conclusions from certain premises. The cognitive operations that are required to solve such tasks can be many, particularly when comparing stimuli with regards to their properties or determining class membership of elements. The focus of sequential reasoning tasks is not on finding or discovering regularities (induction), but on successfully applying rules. For prototypical tasks of sequential reasoning, Carroll (1993) listed deductive reasoning (identifying logical inferences from a verbal problem), syllogisms (evaluating the correctness of a conclusion based on two premises), and verbal reasoning (see Figure 14.3 a) for an illustration).

- *Inductive reasoning* tasks require finding similarities or differences in features or relations between stimuli. In contrast to the deductive reasoning tasks, the difficulty lies in identifying the underlying rules rather than applying them. These rules may represent a concept, a class membership, or a causal relationship. Typical tasks include classification (identifying a figure that does not belong in an ensemble of similar figures) or working with matrices (identifying a figure that replaces a placeholder within a matrix so that the pattern found in rows and columns persist) (see Figure 14.3 c) for an illustration). Formally, all inductive intelligence tasks are essentially enthymemes, that is, deductive inferences in which one or more premises are implicitly "added" rather than explicitly formulated (Wilhelm, 2004).

- *Quantitative reasoning* tasks assess quantitative-numerical components of reasoning. These demands may be deductive, inductive, or a combination of both. Typical examples of quantitative reasoning are mathematical word problems or number series (see Figure 14.3 b) for an illustration). In general, the difficulty lies in mathematical modeling, the numerical-formalization of a problem, rather than in the actual calculation (Carroll, 1993).

A closer examination of the tasks subsumed below Carroll's three reasoning factors suggests that the sequential reasoning factor is predominantly a verbal reasoning factor, the inductive factor is mostly covered by tasks with figural content, and quantitative reasoning relies on numeric content. This interpretation is also supported by Carroll's observations (see Table 6.2 in Carroll, 1993, pp. 217) and his interpretation of the factors of the individual studies. Wilhelm (2004) used confirmatory factor analysis to examine this relationship between 12 different fluid intelligence tasks more closely. Among these tasks, prototypical indicators of deductive reasoning (e.g., propositions and syllogisms) and inductive reasoning tasks (e.g., series and matrices) were selected. The comparison of competing measurement models revealed that a model in which the correlation between inductive and deductive thinking was freely estimated described the data as well as a model in which inductive and deductive thinking were modeled as a common factor. Thus, a distinction between inductive and deductive thinking is artificial and unnecessary from the perspective of differential psychology. Another important finding was that a model with three correlated content factors, covering verbal, numeric, and figural stimulus material, described the data much better than a model with a single reasoning factor (Wilhelm, 2004). The model with three correlated content factors (with no other covariates) is statistically equivalent with a higher-order model in which the content factors load on a higher-order fluid intelligence factor. In line with previous research (e.g., Marshalek, Lohman, & Snow, 1983), the figural reasoning task showed the strongest relation with the overarching fluid intelligence factor, which suggests that the figural content is the best single indicator of fluid intelligence. In summary, the classification of fluid intelligence tasks based on its content is both theoretically and empirically well supported (see Figure 14.3 for example items). Please note that broad visual perception includes spatial ability parts that are close to the reasoning factors discussed here (Lohman, 1996).

Developing a sound and efficient fluid intelligence task is more of an art than a science (Kyllonen & Christal, 1990). This position is predominantly due to a theoretical deficit: most available intelligence tasks suffer from a lack of well-founded theoretical assumptions about the cognitive processes required to successfully complete the tasks in question. Such an underlying theory could be used to derive procedures that generate items automatically, and it could provide *a priori* estimates of item difficulty. For example, for the figural aspect of fluid intelligence, numerous taxonomies for constructing matrix items have been proposed (Carpenter, Just, & Shell, 1990). In his review, Primi (2001) reduced the complexity of influencing factors on item difficulty to four main attributes: (1) the number of elements, (2) the number of transformations or rules, (3) the type of rules, and (4) the perceptual organization. The success of this proposal and similar efforts is mixed, and moreover, most efforts are limited to specific types of tasks.

A promising approach to circumvent these problems and to gain a more profound understanding of reasoning is to instead rely on the concept of **working memory capacity** (WMC). With respect to going beyond task-specific models of what changes the difficulty and nature of a task, WMC can be applied to many working memory items by specifying the storage and processing demands of a task. In a memory updating task, for example, subjects might be shown digits presented in four different locations. These digits disappear, and subjects briefly receive instructions for simple computations at the location of individual digits, one after another. After several such computations, subjects are asked to provide the final results for each of the locations. Such tasks can easily be generated by computers, and their difficulty can be predicted very well with just a few task attributes. WMC tasks might not only prevent some of the problems prevalent with reasoning measures but they are also the key to understanding fluid intelligence and intelligence in general (Engle, 2018; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016).

The relation between WMC and fluid intelligence has received considerable attention (Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004; Oberauer, Schulze, Wilhelm, & Süß, 2005) and there is a broad consensus that this relation is very strong, though not perfect (Kyllonen & Christal, 1990). The main reasons for the very strong, but not perfect, relation might be twofold. First, despite being unwanted, many reasoning tasks do have knowledge requirements that might bias the relation with WMC in a downwards fashion. Second, many WMC tasks have an intrinsic speed requirement by limiting stimulus exposition or time windows for responding. If these biases were adjusted for, the relations between fluid intelligence and WMC might be perfect (Wilhelm, Hildebrandt, & Oberauer, 2013).

14.2.1.2 Tasks for Measuring Crystallized Intelligence

According to Cattell (1971), **crystallized intelligence** should be seen as the result of the investment of fluid intelligence in learning situations, but also depends on additional sources such as investment traits (Ackerman, 1996) and interests (Su, Rounds, & Armstrong, 2009). Thus, gc reflects the impact of education, learning, and acculturation on knowledge-related intelligence tasks. During school years, the item universe for gc measurement is at least partly predetermined through the canon of formal education and through cultural standards that roughly prescribe what children and adolescents are expected to learn and know (Cattell, 1971). This notion suggests an assessment of gc via factual knowledge tests that captures both school and extracurricular content (for an example item see Figure 14.4). As learning opportunities become more and more diverse across one's lifespan and after regular schooling, the assessment of gc becomes increasingly difficult. An ideal measurement of gc must include the whole variety of knowledge that people *can* acquire during their lives (and that are somewhat valued culturally). Consequently, it would require as many different tasks as there are occupations, recreational activities, and other differential learning opportunities. The central role of knowledge in the concept of crystallized intelligence is also emphasized by Ackerman (1996), who stated that gc measures should not be an in-depth assessment of knowledge within a specific domain or a few selected domains; rather, gc measures should be conceptually broad.

Who benefits most from unexpected inflation?

- ☐ *people with debts*
- ☐ *savers*
- ☐ *people with life insurance*
- ☐ *share holders*

Figure 14.4: Example item for crystallized intelligence.

In reality, gc is predominantly assessed via verbal indicators such as vocabulary and verbal fluency tasks. There is no doubt that language skills are important and a result of formal education and, thus, culturally-shared knowledge. This idea is also consistent with the description of gc in Carroll's Three-Stratum Theory (1993). However, the factor-analytic results could instead be an artifact of current assessment practices which have an overrepresentation of verbal ability measures.

But it is also apparent that language command describes only a section of culturally-shared knowledge. In fact, in a large-scale educational assessment study, Schipolowski, Wilhelm, and Schroeders (2014) administered various language tasks, including reading comprehension, listening comprehension, language use and writing, together with a broadly sampling knowledge test covering 16 content domains (e.g., physics, art, law) to an unselected sample of 6,071 adolescents. The correlation between latent variables representing language command and knowledge was very high (ρ = .91), but significantly different from unity. About 17% of the variance in the knowledge factor was independent of individual differences in language command and fluid intelligence (and vice versa). Thus, a restriction to purely language-related content must be regarded as deficient in light of the abovementioned definition of gc because it equates a part of gc with the overarching gc factor (Amthauer, Brocke, Liepmann, & Beauducel, 2001). Please note that command of language may or may not be different from a concept-labeled verbal ability by some researchers.

Cattell (1971) also drew attention to the fact that verbal ability tasks do not necessarily cover gc adequately, especially if the verbal content is strongly over-trained knowledge or decontextualized. Furthermore, knowledge tests also have the greatest potential to minimize the risk of being confounded with fluid intelligence. The maximum separation of gc and gf should be an overriding principle in constructing efficient and distinct measures of cognitive ability (Carroll, 1993). Language skills and reasoning abilities are minimal requirements for knowledge tests, as they are necessary to understand the question and answer options at a basic level. Taken together, we conclude that declarative knowledge tests should take into account as many areas of knowledge as possible to be used as marker variables of gc, as they include a variety of learning experiences that go beyond language skills and competencies.

14.2.2 Validity of Intelligence Tests

Up until this point, we have presented different conceptualizations of intelligence and ways to measure it. We can also take a very pragmatic position while discussing the strengths and benefits of intelligence testing. Intelligence tests are used in psychology, educational research, and other behavioral sciences for a wide range of purposes because intelligence is one of the best predictors of educational, vocational, academic success, and job performance (e.g., Schmidt & Hunter, 1998; Schmidt, 2002). Intelligence in this context mostly refers to the ability to reason (gf) and domain-related knowledge (gc). The **predictive validity** of both components seems to vary during the course of life. In a comprehensive review, Baumert and colleagues (2009) compared the results of various educational-psychological studies and showed that the predictive power of domain-specific knowledge in comparison to reasoning becomes more important the older students are. Obviously, the contributions of gf and gc are hard to distinguish because they are strongly correlated. The relevance of knowledge on significant outcomes and its underrepresentation in contemporary intelligence assessment led Ackerman (2000) to the conclusion that domain-specific knowledge is the “dark matter” of adult intelligence. His PPIK theory (intelligence-as-process, personality, interests, and intelligence-as-knowledge; Ackerman, 1996), builds on Cattell’s gf-gc-theory. It distinguishes several types of knowledge (e.g., occupational knowledge) to give domain-specific knowledge the space it deserves.

Much research was conducted to shed light on the developmental interplay between gf and gc. In the investment theory, Cattell (1971) proposed that crystallized knowledge develops through the investment of fluid ability. However, empirical evidence for this assumption is sparse. For example, Ferrer and McArdle (2004) used linear dynamic models to study the trajectories of gf and gc from childhood to early adulthood. The results showed no coupling between gf and gc within the studied age range, which clearly contradicts the investment theory. When reviewing available empirical evidence and methodological approaches on the development of gf and gc, it becomes evident that there is no direct or simple explanation to account for the development and mutual relation between cognitive abilities in general, and gf and gc in particular. To overcome this issue, Savi, Marsman, van der Maas and Maris (2018), for example, proposed to abandon factor analytic methods in intelligence research and instead conceptualize intelligence as evolving networks in which new knowledge and processes are wired together during development. This approach might also bridge the gap between the study of individual differences in intelligence and phenomena primarily studied in cognitive psychology, such as forgetting.

The great importance of intelligence is evident not only in school or university education (Kuncel & Hezlett, 2007; Schmidt & Hunter, 1998), but also in professional training (Ziegler, Dietl, Danay, Vogel, & Bühner, 2011). As a cautionary note, even though intelligence is the most influential single predictor of academic achievement, it still accounts for only about a quarter of variation in the outcome. Accordingly, successful learning at school and the university depends on a plethora of individual characteristics—such as the personality trait conscientiousness (Barrick & Mount, 1991) or interests (Holland, 1997)— in addition to intelligence.

A last aspect of predictive validity we would like to touch upon has to do with death. Initially labeled "ultimate validity" (O'Toole & Stankov, 1992), the relevance of intelligence for longevity becomes increasingly clear. It turns out intelligence might be an essential contributor to epidemiological outcomes in that premorbid intelligence predicts all sorts of health related behaviors and diseases which in turn are related with mortality (Batty, Deary, & Gottfredson, 2007).

14.2.3 Training of Intelligence

"How Much Can We Boost IQ and Scholastic Achievement" was the title of an influential and very controversial paper published in the late sixties (Jensen, 1969). In this paper, Jensen drew a somewhat pessimistic conclusion concerning interventions intended to improve IQ or scholastic achievement. In their notorious book, "The Bell Curve: Intelligence and Class Structure in American Life", Herrnstein and Murray (1994) also concluded with negative inferences concerning the improvement of IQ and scholastic achievement. The contributions by Gottfredson (1997) and Neisser et al. (1996) for defining intelligence as a concept (discussed earlier in this chapter) were, in fact, both reactions to the controversy triggered by the Hernstein and Murray book. Importantly, both publications suggested relatively explicitly that many of the observed group differences in IQ and scholastic achievement are determined genetically. Obviously, today's scientists working in the fields of behavior or molecular genetics of traits have gained a more profound understanding of heritability and use more advanced statistical methods and designs to study the relevance of nature and nurture.

For example, Plomin and von Stumm (2018) summarized recent findings on genome-wide association studies, identifying genome sequence differences that account for 20% of the 50% heritability of intelligence. Such reports on the genetic transmission of intelligence seem to be contradicted by the fact that schooling affects both scholastic achievement (for a comprehensive account, see the classes of evidence described by Ceci, 1991) and intelligence (Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012; Cliffordson & Gustafsson, 2008). However, there is nothing contradictory about these findings once genetic effects are interpreted correctly (Johnson, Turkheimer, Gottesman, & Bouchard, 2010). Also, in a recent meta-analysis of quasi-experimental studies with strong designs (i.e., those that allow statements about increases in intelligence as a function of schooling), Ritchie and Tucker-Drob (2018) summarized overwhelming evidence for education being the most consistent, robust, and durable method for raising intelligence. They found an increase between 1 and 5 IQ points for every additional year of schooling.

Somewhat related, it can be shown that non detrimental or supporting environments have a positive effect on intelligence over a broader time period (Flynn, 1984). Despite contradicting results, the so-called Flynn-effect might in fact not have leveled off in the past two decades (Trahan, Stuebing, Fletcher, & Hiscock, 2014). Beside the aforementioned changes in the educational system, different factors have been discussed for being responsible for the IQ gains. In particular, education and health-related factors such as better nutrition and reduced pathogen stress appear to be related to IQ gains (Pietschnig & Voracek, 2015).

The evidence presented so far is correlative and at a macroscopic level. If we want to answer the question laid out at the beginning of this section, we should take a closer look at the experimental evidence. Prior to evaluating such evidence, the benchmark for such an evaluation should be clear.

Training effects on intelligence should a) persist after the training ended (effect duration), b) be present in non-trained tasks (effect transfer), c) be specific to the targeted intelligence so that not everything is improving but only trained aspects, d) be stronger in trained than in non-trained subjects (who should be engaged in other training instead of simply waiting), and e) be rational and sensible in the way that the intervention is tailored to what it should accomplish and it provides a non-trivial gain.

Moreover, bearing in mind the current replication crisis (Open Science Collaboration, 2015), training studies should fulfill the requirement of experimental studies concerning sample size, sound measurement instruments, *a priori* specified outcome variable, etc.. Unfortunately, many popular studies that received extensive mass media coverage do not adhere to these requirements (Melby-Lervag, Redick, & Hulme, 2016). Accordingly, many of the bold claims about successfully training intelligence or its most important facets (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) can be attributed to methodological flaws and are not due to some miraculous interventions (Melby-Lervag et al., 2016).

Reviewing most interventions shows that they were designed with the hope that a few hours of training would bring about long-lasting, transferable, and relevant improvements in highly general intellectual abilities. This claim is not only bold; it is completely unrealistic. Even if we adhere to a lifestyle that spares us intellectual effort, we can hardly be functioning members of society if we do not regularly engage in effortful, intellectual, and challenging thinking. In other words, our everyday lives provide daily intellectual exercises, no matter how trivial and dull they feel from time to time. Whether we like it or not, we use our intellectual capacity constantly. Training must provide a sufficiently large additional dosage to make a real difference. Moreover, the mechanisms being stressed by intelligence training should also be suited to bring about the desired change. Alas, most training—in a sense of over-learning rather simple tasks—just have people repeatedly completing different variations of the same type of question. Simply adjusting the difficulty of questions to say 50% is not an impressive improvement of the retesting-ad nauseam-approach. Studies with a more substantial dosage provide a much better read and a more realistic picture (Schmiedek, Lövdén, & Lindenberger, 2010).

Another field in which fostering intellectual functioning was studied is cognitive ageing. The use-it-or-lose-it hypothesis (Hultsch, Hertzog, Small, & Dixon, 1999) suggests that being intellectually active prevents an age-associated cognitive decline. Obviously, it is difficult to collect strong data on cognitively-active lifestyles over decades and, thus unsurprisingly, there still seems to be no conclusive evidence (Salthouse, 2006). Given that intellectually engaging activities will hardly have adverse effects, living a mentally active life is not a bad choice. However, if you are hoping to maintain or improve your intelligence by skipping physical activity in exchange for intellectual activity—this is probably a bad idea in the long run, as physical exercise has been shown to be beneficial for intellectual functioning (Kramer & Colcombe, 2018).

14.3 Conclusions

We want to use this section to point out a few pervasive problems in intelligence research, raise open questions, and hint to potential solutions for such problems. We began this chapter by highlighting that intelligence research is about individual differences and covariation, whereas most other chapters in this book are about the general psychology of cognition and experimental effects. There is some lamentation about the unfortunate nature of the barriers between these two disciplines (Cronbach, 1957). Indeed, the intelligence model we introduced as widely accepted has a substantial lack of cognitive sophistication. For example, despite its essential role in intelligence research, our understanding of most reasoning tasks is severely limited. Popular definitions of the construct often stress the novelty of reasoning tasks as an essential feature, yet we have no clear idea of what novelty actually means. Usually, these discussions move on by pointing to induction, deduction, and sometimes abduction—but rarely is there ever a connection between reasoning tasks used in intelligence research and the same tasks being used in experimental settings to study competing theories about inductive thinking, for example. Taken together, the lamentation about these two disciplines of psychology remains justified.

In the end, gc can be considered as a collection of all sorts of pragmatic and knowledge-driven thinking. We have merely begun to understand the breadth of all the aspects we are subsuming here: wisdom, command of a language, foreign-language aptitude, declarative and procedural knowledge of all sorts etc.. Crystallized intelligence needs a lot more attention. And research on gc demands specific methods due to its intrinsic orientation towards change and its idiosyncrasy that grows over the course of one's life.

A closer look at the general learning and recognition factor provokes a few questions, too. The factors below glr mostly refer to specific methods for measuring memory. Of course, no one can claim that associative memory is a different memory store than free recall memory, for example, even though the factor labels suggest so. Additionally, researchers are at a loss when it comes to choosing a glr test because the method selected heavily affects outcomes. A much stronger connection with experimental approaches is essential to further our understanding of this factor.

The discussion of potential shortcomings of the taxonomy we actually endorse seems endless. Should originality and creativity really be located below learning and retrieval? What about interpersonal abilities, such as emotional competence? Clearly, there is no shortage of questions and problems. It is therefore important to understand this taxonomy as a starting point rather than as an end result. There is much to be improved, but intelligence testing in all its varieties is also a major success story from an applied perspective. It is a strong predictor for several desirable outcomes and it is no doubt essential for determining how cognitively rich our lives are.

Summary

1. We began this chapter by briefly reviewing milestones of intelligence research and juxtaposed competing models of individual differences in intelligence.
2. We described contemporary accepted models on the structure of intelligence in which fluid and crystallized intelligence are the most important factors.
3. We argued for an extension of the use of the term intelligence to all tasks essentially reflecting individual differences in maximal cognitive effort.
4. In the second section, we presented prototypical tasks for fluid and crystallized intelligence.
5. We discussed a number of weaknesses in contemporary intelligence models and argued that - from a pragmatic viewpoint - the measurement of intelligence is still a success story.
6. We discussed several efforts to improve intelligence, a class of interventions that should interest not only individuals but also society as a whole.

Review Questions

1. Think about the models in Figure 14.1. Which one do you think best captures individual differences in human intelligence? Why?
2. Can you think of maximal cognitive effort concepts that are not part of the CHC-model?
3. Where in Figure 14.2 would you locate emotion perception? What skills and abilities make a good insurance broker?
4. Can you come up with example items for a fluid intelligence task? Are these items deductive or inductive; are they verbal, numeric, or figural?
5. Why is the enthusiasm concerning interventions for working memory premature?

Hot Topic: Sex Differences in Crystallized Intelligence?

Oliver Wilhelm

Few topics in ability research are regarded as controversial as sex/gender differences in cognitive abilities. According to the Gender Similarity Hypothesis (Hyde, 2005), sex differences in cognitive abilities are mainly small and unsystematic. This general conclusion is empirically supported for fluid intelligence but is challenged for crystallized intelligence, when measured with knowledge tests. Most studies show that males outperform females in general knowledge, with an average overall effect of $d = .26$ (Schroeders, Wilhelm, & Olaru, 2016). A closer examination on the level of domains reveals a more complex pattern: for example, females clearly outperform men in health-related domains, such as aging and nutrition, but large differences in favor of males were found for technology and the natural sciences (Ackerman, Bowen, Beier, & Kanfer, 2001). It is striking that such stereotypic sex-related differences in knowledge domains seem to match the sex differences in interest as reported in the famous "Men and Things, Women and People" meta-analysis by Su and colleagues (2009). On the other hand, we should avoid overgeneralizing such differences. For example, the magnitude and direction of sex or gender differences in mathematical competencies varies dramatically across countries (Stoet & Geary, 2013).

Ulrich Schroeders (Photo: Sonja Rode/ Lichtfang.net)

An aspect that is often neglected in studies on group differences in cognitive abilities is the aspect of item sampling. The same way participants of a study are selected from a population (person sampling), items can be thought of as being drawn from a population of items (item sampling). In the construction and validation of psychological measures, we usually assume that we draw items from a theoretically infinite item universe. In a recent study, we put this idealistic assumption to the test (Schroeders et al., 2016). We used metaheuristic sampling procedures (i.e., ant-colony-optimization algorithms) to compile psychometrically sound short forms of a knowledge test. The algorithm was set for two criteria, a) to select items from an initial set that adhere to strict psychometric criteria concerning fit of the data to a model, and

b) to deliberately tilt sex differences to either favor males or females. The results show that sex differences vary considerably depending on the indicators drawn from the item pool. In other words, we could compile knowledge tests for sciences and technology in which females outperformed males. They also could compile health tests in which males outperformed females. This result questions the generalizability of previously reported findings on sex differences in crystallized intelligence. On a more general stance, the results corroborate the notion of Loevinger (1965, p. 147) that the random sampling assumption of items (and tests) is unrealistic because test development is "almost invariably expert selection rather than sampling". Unfortunately, many studies concerning group differences in cognitive abilities fail to acknowledge item selection effects.

References

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*, 797–825. doi:10.1037/0022-0663.93.4.797

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592. doi:10.1037/0003-066x.60.6.581

Loevinger, J. (1965). Person and population as psychometric concepts. *Psychological Review*, *72*, 143–155. doi:10.1037/h0021704

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). The influence of item sampling on sex differences in knowledge tests. *Intelligence*, *58*, 22–32. doi:10.1016/j.intell.2016.06.003

Stoet, G., Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data. *PLoS ONE*, *8*, e57988. doi:10.1371/journal.pone.0057988

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*, 859–884. doi:10.1037/a0017364

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257. doi:10.1016/s0160-2896(96)90016-1

Ackerman, P. L. (2000). Domain-specific knowledge as the "dark matter" of adult intelligence: Gf/Gc, personality and interest correlates. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *55*, 69–84. doi:10.1093/geronb/55.2.p69

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2000). Explorations of crystallized intelligence: Completion tests, cloze tests, and knowledge. *Learning and Individual Differences*, *12*, 105-121. doi:10.1016/s1041-6080(00)00034-0

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, *44*, 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x

Batty, G. D., Deary, I. J. & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: systematic review. *Annals of Epidemiology*, *17*, 278–288. doi:10.1016/j.annepidem.2006.07.010

Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review*, *4*, 165–176. doi:10.1016/j.edurev.2009.04.002

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

Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année Psychologique*, *11*, 191–244. doi:10.3406/psy.1904.3675

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

Bridgeman, B. (2005). Unbelievable results when predicting IQ from SAT scores: A comment on Frey and Detterman (2004). *Psychological Science*, *16*, 745–746. doi:10.1111/j.1467-9280.2005.01606.x

Brown, R. E. (2016). Hebb and Cattell: The genesis of the theory of fluid and crystallized intelligence. *Frontiers in Human Neuroscience*, *10*, 1–11. doi:10.3389/fnhum.2016.00606

Brunner, M. (2008). No g in education? *Learning and Individual Differences*, *18*, 152–165. doi:10.1016/j.lindif.2007.08.005

Carpenter, P. A., Just, A. J. & 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. doi:10.21236/ada221448

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. New York: Cambridge University Press. doi:10.1017/cbo9780511571312

Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action*. Boston: Houghton Mifflin Company.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*, 703–722. doi:10.1037//0012-1649.27.5.703

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. doi:10.1016/j.intell.2007.03.006

Cronbach, L. J. (1949). Essentials of psychological testing. New York: Harper & Brothers. Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684. doi:10.1037/14156-015

Davies, M., Stankov, L., & Roberts, R. (1998). Emotional intelligence in search of an elusive construct. *Journal of Personality and Social Psychology*, *75*, 989-1015. doi:10.1037/0022-3514.75.4.989

Ebbinghaus, H. (1895). Über eine neue Methode zur Prüfung geistiger Fähigkeiten und ihre Anwendung bei Schulkindern. *Zeitschrift für die Psychologie*, *13*, 401–459. doi:10.1037/h0064465

Engle, R.W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*, 190–193. doi:10.1177/1745691617720478

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, *40*, 935–952. doi:10.1037/0012-1649.40.6.935

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin*, *95*,29–51. doi:10.1037/0033-2909.95.1.29

Galton, F. (1883). *Inquiries into Human Faculty and its Development*. London: The Eugenics Society with Dent & Sons. doi:10.1037/14178-000

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, New York, NY.

Gardner, H. (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*, Basic Books, New York, NY.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*, 79–132. doi:10.1016/s0160-2896(97)90014-3

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. 275–292). Washington, DC: APA Books. doi:10.1037/10315-012

Guttman, L. & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, *15*, 79–103. doi:10.1016/0160-2896(91)90023-7

Hernstein, R. J., & Murray, C. (1994). *The Bell curve: The Reshaping of American Life by Difference in Intelligence*. New York: The Free Press.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Odessa, FL: Psychological Assessment Resources.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*, 41–54. doi: 10.1007/BF02287965

Hultsch, D. F., Hertzog, C., Small, B. J., & Dixon, R. A. (1999). Use it or lose it: engaged lifestyle as a buffer of cognitive decline in aging? *Psychology and Aging*, *14*, 245–263. doi:10.1037/0882-7974.14.2.245

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*, *105*, 6829–6833. doi:10.1073/pnas.0801268105

Jensen, A. (1969). *How much can we boost IQ and scholastic achievement*. Harvard Educational Review, 39, 1–123. doi:10.17763/haer.39.1.l3u15956627424k7

Johnson, W., Turkheimer, E., Gottesman, I. I., & Bouchard, T. J. (2010). Beyond heritability: twin studies in behavioral research. *Current Directions in Psychological Science*, *18*, 217–220. doi:10.1111/j.1467-8721.2009.01639.x

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71. doi:10.1037/0033-2909.131.1.66

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. doi:10.1037/0096-3445.133.2.189

Kramer, A. F., & Colcombe, S. (2018). Fitness effects on the cognitive function of older adults: A meta-analytic study – revisited. *Perspectives on Psychological Science*, *13*, 213–217. doi:10.1177/1745691617707316

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080–1081. doi:10.1126/science.1136618

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433. doi:10.1016/s0160-2896(05)80012-1

Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *Journal of Organizational Behavior*, *26*, 411–424. doi:10.1002/job.317

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Lohman, D. F. (1996). Spatial ability and g. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97-116). Mahwah, NJ: Erlbaum. doi:10.4324/9780203774007

MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models. *Emotion*, *14*, 358–374. doi:10.1037/a0034755

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107–127. doi:10.1016/0160-2896(83)90023-5

Mayer, J. D. (2018). Intelligences about things and intelligences about people. In R. J. Sternberg (Ed.). *The nature of human intelligence* (pp. 270-286). Cambridge, England: Cambridge University Press. doi:10.1017/9781316817049.018

Mayer, J. D., Salovey, P., & Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry*, *15*, 197-215. doi:10.1207/s15327965pli1503_02

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. doi:10.1016/j.intell.2008.08.004

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer" evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*, 512–534. doi:10.1177/1745691616635612

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101. doi:10.1037/0003-066X.51.2.77

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*, 758–799. doi:10.1037/bul0000046

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence – their correlation and their relation: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61–65. doi:10.1037/0033-2909.131.1.61

Oostrom, J. K., De Soete, B., & Lievens, F. (2015). Situational judgment testing: A review and some new developments. In I. Nikolaou, & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment: Contemporary issues for theory and practice* (pp. 172-189). Sussex, UK: Psychology Press. doi:10.4324/9781315742175

O'Toole, B. I., & Stankov, L. (1992). Ultimate validity of psychological tests. *Personality and Individual Differences*, *13*, 699–716. doi:10.1016/0191-8869(92)90241-G

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles concepts and evidence. *Psychological Science in the Public Interest*, *9*, 105–119. doi:10.1111/j.1539-6053.2009.01038.x

Pietschnig J, Voracek M. (2015). One century of global IQ gains: a formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science*, *10*, 282–306. doi:10.1177/1745691615577701

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*, 148–159. doi:10.1038/nrg.2017.104

Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, *30*, 41–70. doi:10.1016/s0160-2896(01)00067-8

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*, 1358–1369. doi:10.31234/osf.io/kymhp

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion*, *1*, 196–231. doi:10.1037/1528-3542.1.3.196

Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, *73*, 482–486. doi:10.1037//0021-9010.73.3.482

Salovey, P., & Mayer, J., & Caruso, D. (2004), Emotional Intelligence: Theory, Findings, and Implications, *Psychological Inquiry*, *15*, 197–215. doi:10.1207/s15327965pli1503_02

Salthouse, T. A. (2006). Mental exercise and mental aging: Evaluating the validity of the "use it or lose it" hypothesis. *Perspectives on Psychological Science*, *1*, 68–87. doi:10.1111/j.1745-6916.2006.00005.x

Savi, A. O., Marsman, M., van der Maas, H. L. J., & Maris, G. (2018). *The wiring of intelligence*. doi:10.31234/osf.io/32wr8

Schipolowski, S., Wilhelm, O., & Schroeders, U. (2014). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence*, *46*, 156–168. doi:10.1016/j.intell.2014.05.014

Schmid, J. & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53–61. doi:10.1007/bf02289209

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187-210. doi:10.1080/08959285.2002.9668091

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274. doi:10.1037//0033-2909.124.2.262

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, 1–10. doi:10.3389/fnagi.2010.00027

Schulte, M. J., Ree, M. J., & Carretta, T. R. (2004). Emotional intelligence: not much more than g and personality. *Personality and Individual Differences*, *37*, 1059–1068. doi:10.1016/j.paid.2003.11.014

Schroeders, U. (2018). Ability. In M. H. Bornstein (Ed.), *The SAGE Encyclopedia of Lifespan Human Development* (pp. 1–5). Thousand Oaks, CA: SAGE Publications. doi:10.4135/9781506307633.n8

Spearman, C. (1904a). "General Intelligence" objectively determined and measured. *American Journal of Psychology*, *15*, 201–293. doi:10.2307/1412107

Spearman, C. (1904b). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*, 72–101. doi:10.2307/1412159

Sternberg, R. (1977). Component processes in analogical reasoning. *Psychological Review*, *84*, 353-378. doi:10.1037/0033-295X.84.4.353

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*, 859–884. doi:10.1037/a0017364

Thomson, G. H. (1919). On the cause of hierarchical order among correlation coefficients. *Proceedings of the Royal Society A*, *95*, 400–408. doi:10.2307/2331895

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, *38*, 406–427. doi:10.1037/h0069792

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, *41*, 1–32. doi:10.1037/h0075959

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn Effect: A Meta-Analysis. *Psychological Bulletin*, *140*, 1332–1360. doi:10.1037/a0037173

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M. & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861. doi:10.1037/0033-295x.113.4.842

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). g and the measurement of Multiple Intelligences: A response to Gardner. *Intelligence*, *34*, 507–510. doi:10.1016/j.intell.2006.04.006

Wilhelm, O. (2004). Measuring reasoning ability. In Wilhelm, O., & Engle, R. W. (Eds.). *Handbook of Understanding and Measuring Intelligence* (pp. 373–392). Sage Publications. doi:10.4135/9781452233529.n21

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *433*. doi:10.3389/fpsyg.2013.00433

Ziegler, M., Dietl, E., Danay, E., Vogel, M., & Bühner, M. (2011). Predicting training success with general mental ability, specific ability tests, and (un)structured interviews: A meta-analysis with unique samples. *International Journal of Selection and Assessment*, *19*, 170–182. doi:10.1111/j.1468-2389.2011.00544.x

Glossary

content validity A form of validity that addresses whether or not the items of a test or the tests of a battery represent a predefined or circumscribed item universe. 261

crystallized intelligence Breadth and depth of cultural knowledge often measured with declarative knowledge tests or tests of language proficiency. 264

fluid intelligence Reason, plan, solve abstract and complex problems that can not be solved without effortful thinking. 262

predictive validity Extent to which measures such as gf-tests predict relevant outcomes such as college grade point average. 265

working memory capacity A person's capacity to simultaneously store and process information. 263