Gottfredson, L. S. (2008). Of what value is intelligence? In Prifitera, A., Saklofske, D. H, & Weiss, L. G. (Eds.), *WISC-IV clinical assessment and intervention (2nd ed.)* (pp. 545-563). Amsterdam: Elsevier.

15

Of What Value Is Intelligence?

Linda S. Gottfredson

School of Education, University of Delaware, Newark, DE, USA

First developed a century ago, intelligence tests remain the best known, best researched, and perhaps most frequently used of all psychological assessments. But why? After all, they have been dogged by public controversy since the beginning. The answer turns out to be simple, although hard-won in a century of research and debate among many hundreds of scholars. My aim here is to provide enough of that scientific account to put into broader social perspective what we know about intelligence and how we know it, as well as what intelligence tests reveal about fundamental human differences, how they do so, and why we care about them.

FROM WHAT VANTAGE POINTS – PERSON OR POPULATION – DO WE LOOK AT INTELLIGENCE?

The term "ability," like intelligence, is often used in two ways: first, to refer to a domain of tasks that might be performed well, especially as development proceeds (e.g., a child is better able to reason abstractly at age 15 than age 5), and second, to refer to differences among individuals in such capabilities (e.g., one 15 year old is more adept at abstract reasoning than another 15 year old). The distinction is between "*what* is done well" and "*who* has the edge in doing it well." "What" research focuses on typical trends in development (intraindividual change over time), and seeks to gauge competence against some external criterion. "Who" research focuses on population variation around the trend line at a given age, and it usually takes a norm-referenced approach to abilities, which involves comparing an individual to others in some reference group. In short, one approach studies the common human theme; the other, variations on it. Both are concerned, of course, with the same underlying phenomenon – some particular continuum of competence.

Tests of intelligence, personality, and the like remain mostly in the second tradition referred to as the study of individual differences or differential psychology. So, the key question in research on IQ tests has been "What do they tell us about human variation, and how is that information useful?" Tests can be useful for many purposes without our knowing what they measure. For example, colleges and employers both use ability tests to identify which applicants are most likely to succeed if selected (Campbell & Knapp, 2001). Like my car, a test can be a mysterious black box as long as it gets me where I want to go. Predictive or diagnostic utility was the aim of the first intelligence test, developed by Binet and Simon in 1906, because they wished to identify children who would not succeed in school without special assistance.

For other purposes, clinical, scientific, or political, we also have to understand what phenomenon intelligence tests measure, that is, their *construct validity*. Researchers who seek to understand the nature of intelligence itself must know, at a minimum, what latent constructs the tests measure and how they do so. Evidence about predictive validity is crucial but not sufficient. Knowing what construct a test measures is a long iterative process in which provisional answers to that question are used to generate additional testable predictions about the trait's stability and course of development, which kinds of tasks call it forth most strongly, which circumstances make it rise or fall, how well it predicts different kinds of life success, and so on.

Perhaps in no other applied setting is construct validity more important than for clinicians who are asked to diagnose individuals and intervene in their lives. Such is the case for school psychologists, for example, when they assess individual students who are having difficulties in the classroom in order to design interventions for ameliorating those difficulties. They must create a theory, so to speak, of that particular child based on a broad set of information gathered from tests, teachers, and often parents too. They need such a theory, a close-up ideographic portrait, to understand what is impeding that particular child's learning or adjustment and to develop strategies for eliminating or working around those impediments. Arguably, a battery of cognitive tests is the most important single tool in sketching that portrait. That is why the present book focuses on the richness of information that the WISC-IV index scores can provide about a child's profile of strengths and weaknesses, in addition to overall level of intellectual functioning reflected in the total IQ, in order to gain leverage in diagnosis and treatment.

The aim of the current chapter is to place that ideographic client-centered use of intelligence testing within the broader social context in which IQ tests are used and judged. To do that, I bring to the foreground what has been mostly background in this book's advice for painting complex cognitive portraits of individual children, namely the general intelligence factor, *g*, as gauged by the WISC-IV Full-Scale IQ (FSIQ). So, instead of examining the profile of highs and lows among different abilities within a single person, this chapter turns to examining differences in a single very important ability across many different people. Whatever one's views about the scientific validity of using tests to assess interindividual differences in intelligence, I suspect all would agree that ranking individuals by general intelligence generates the most public controversy (Williams, 2000; Gottfredson, in press).

IS INTELLIGENCE ANYTHING MORE THAN A SCORE ON AN IQ TEST?

Some critics assert that intelligence is no more than what intelligence tests measure, thus encouraging us to doubt that intelligence can be measured, if it exists at all. According to them, IQ tests trap us forever in an endless tautological loop going nowhere because, they suggest, the IQs calculated from a test simply summarize what testers themselves put into it. On the other hand, when testers respond that IQ tests measure something deeper, some phenomenon in its own right, how can we know what that phenomenon is apart from the tests they use to measure it? The fact that different intelligence tests correlate highly among themselves tells us nothing about what any of them measures. The scientific credibility of one test is not enhanced by pointing to others like it, because all could be similarly mistaken. This is the same tautology, just twice removed.

Other critics of intelligence testing suggest that we cannot know whether we have measured intelligence, let alone measured it well, until everyone agrees on a common, carefully specified, *a priori* definition of what it is. This would leave us worse off than before – with no tests of intelligence – as may be the critics' aim. Scholars certainly will never agree on what intelligence is before they have done the research necessary to learn what it is. Empirical phenomena are not defined into existence, but described once known.

How, then, do we even know that intelligence differences exist as a stable phenomenon to be investigated and measured? Some critics assert that testers find differences in intelligence only because they intend to, specifically, by developing tests that exaggerate minor differences or manufacture new ones (Fischer et al., 1996). They thus posit the ultimate tautology: IQ differences represent nothing but the intent by psychometricians to create the appearance of difference. By the critics' reasoning, there would be no differences absent such intent. This is akin to claiming that heat exists only because scientists have created thermometers to measure it.

Intelligence is, in fact, much like heat. Neither heat nor intelligence can be directly seen, touched, or held. We nonetheless notice differences in both as we go about our daily lives, often experiencing them as immediate and obvious. We might not understand them, but they clearly affect us regardless of whether we ever measure or define them. We have large vocabularies for each, itself indicating our ongoing concern with them, and we shape our lives somewhat in response to them. Both continua exist in nature, ready to be measured and scientifically explained. Psychometricians and other scholars of intelligence have steadily advanced against this measurement challenge, decade after decade, for over a century now (Bartholomew, 2004; Roberts, 2007).

WHAT IS INTELLIGENCE, AND HOW DO WE KNOW THAT IQ TESTS MEASURE IT?

The early intelligence tests might be likened to early thermometers – first efforts to measure a distinction long perceived as relevant to our lives, which are guided by our initial intuitions about that distinction. That is how Binet and Simon proceeded in 1906. Just as thermometer readings must not be influenced by humidity, much psychometric work since their time has gone into assuring that intelligence test scores are highly reliable and not influenced by irrelevant factors such as cultural bias (Jensen, 1980). Indeed, the margin of error in FSIQ scores is smaller than for many physical assessments, such as blood pressure readings, and their diagnostic sensitivity and specificity exceed that of many medical assessments. Researchers have tested competing notions about the structure of human cognitive abilities and of its stability and comparability across different ages and demographic groups. I am not aware of any important behavioral or psychological assessment with greater reliability, demonstrated construct validity, or predictive validity in many life arenas than a professionally developed intelligence test battery properly administered.

The century of research has also revealed a lot about what intelligence represents at the everyday behavioral level, as well as providing tantalizing glimpses of its manifestations in the brain. The most dramatic advance, in my view, has been to escape the tautology that "intelligence is what intelligence tests measure." In fact, as demonstrated shortly, psychometricians now have an independent means of determining *how well* different tests – indeed, any test or task – measure general intelligence.

What does the research reveal? Perhaps most importantly, it shows that global intelligence as measured by IQ tests is a highly organized system of interrelated mental abilities, all of which share a common core. Human intelligence is highly structured in this sense, not merely a collection of separate, independent abilities like marbles in a bag, where all that is required for an individual to be smart is to collect a large number of any type. There are many kinds of cognitive abilities, to be sure, but individuals who possess one tend to possess all others in good measure too. This observation is what led Charles Spearman (1904) to hypothesize a general factor of intelligence, *g*, and what has prompted so many decades of psychometric research aimed at charting the patterns of relatedness and overlap among seemingly different dimensions of cognitive variation.

These many abilities are best distinguished by their breadth of application, that is, by how domain-specific versus domain-general they are, and only secondarily by manifest content (verbal, quantitative, etc.). This structure of observed overlap and relatedness among abilities, based on factor analyses of their intercorrelations, is usually referred to as the hierarchical model of cognitive abilities. It is hierarchical because it classifies abilities into tiers according to their generality of application. The most general abilities are represented in the top tiers and the narrowest and most specific in the bottom tier. This model is useful for integrating all cognitive abilities into a single unifying framework where it can be seen, for example, that the narrower abilities are mostly composites of the broader ones. Carroll's (1993) Three-Stratum Model of cognitive abilities, developed from his re-examination of 500 prior studies, is currently the most influential model of the structure of human mental abilities.

More specifically, when batteries of diverse cognitive tests are factor analyzed, they reveal a smaller number of broad dimensions of ability, sometimes called primary abilities. There are positive correlations among all cognitive tests, as noted earlier, but certain subsets clump together as especially highly intercorrelated, as if they possess something else in common that the others do not. Carroll (1993) placed these broad factors in Stratum II of his model. He identified eight at this level of generality, including Fluid Intelligence, Crystallized Intelligence, General Learning and Memory, and Processing Speed. The four index scores of the WISC-IV represent abilities of comparable breadth: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI). The 12 WISC subtests from which they are calculated represent Stratum I abilities in Carroll's model. It is from these sorts of broad Stratum II abilities that school psychologists and vocational counselors construct ability profiles for individuals. A spatial tilt, for example, is often associated with interest in and aptitude for technical work in the physical sciences and the skilled crafts. Everyday intuition tells us that people are not equally intelligent or unintelligent in all respects, and the somewhat uneven profiles of Stratum II abilities in general populations confirm that. Specific learning disabilities such as dyslexia represent highly unusual disparities in abilities that normally move in tandem.

These broad abilities are themselves strongly intercorrelated, indicating that they are not separate, independent abilities but reflect some deeper commonality – that is, some yet more generalizable ability, a common core, that enhances performance across all the content domains they represent, from verbal reasoning to spatial and auditory perception. When the Stratum II ability factors are themselves factor analyzed, they yield a single higher-order factor of mental ability – called *g*, for the *g*eneral mental ability factor. The *g* factor typically accounts for more of the common variance among tests than do all the other derived factors combined. In essence, most tests of specific abilities measure mostly *g* plus one or more specific additives. Carroll, for example, referred to Stratum II ability factors as differently flavored forms of the same *g*. The large core of *g* in all Stratum I and II cognitive abilities helps to explain why dedicated efforts to develop useful tests of them that do not correlate appreciably with IQ have all failed. *g* certainly cannot be said to encapsulate the whole of intelligence as many conceive it, but it does fit well what most experts and laymen alike think of as general intelligence: a general purpose tool for learning and reasoning well, spotting and solving problems, and using abstract ideas.

The *g* factor is not an artifact of factor analysis, but has been independently confirmed by biological, genetic, and other sorts of non-psychometric evidence (Jensen, 1998; Deary, 2000). This thick web of empirical correlations involving highly *g*-loaded tests also clarifies what intelligence is *not*, regardless of whether we use the label exclusively for *g* or to encompass the entire structure of cognitive abilities. Intelligence is not, as sometimes claimed, just a narrow academic ability, test-taking smarts, aptness with paper-and-pencil tests (the best IQ tests rely on neither), or a collection of narrow, independent skills. *g* is certainly not a thing or a place in the brain. It may not even be an ability as such, but a property of the brain – a sort of cerebral wattage, mental horsepower, or overall efficiency that tones up all parts of the brain and all aspects of cognitive functioning. Whatever *g* turns out to be at the physiological level, for most practical purposes the FSIQ is an excellent, albeit imperfect, measure of it at the psychometric level.

One side benefit of the integrative hierarchical model has been to clarify the different senses in which the term intelligence is often used – and confused. Some of us restrict the term to the single factor, *g*, found at the top of the Three-Stratum Model, although prefacing it with the adjective "general." Others apply the term intelligence to the small collection of broad abilities at the Stratum II level. This is presumably where the more cognitively oriented of Gardner's (1983) proposed multiple intelligences would show up were he to measure them (linguistic, logical-mathematical, spatial, and musical). Other scholars extend the term to include the entire hierarchy or, like Gardner, outside the cognitive realm to include a wide variety of non-cognitive skills and traits, ranging from physical coordination (Gardner's bodily kinesthetic intelligence) to motivation and conscientiousness, on the grounds that these human attributes also are culturally valued or adaptive.

g is a far more precise referent for the major construct which IQ tests measure than is the term intelligence. The distinction between *g* and IQ, on the other hand, is not merely semantic. It is a conceptual distinction whose importance cannot be overstated. The IQ is a reading from a measurement device, but *g* is a theoretical construct that transcends the particulars of any test or population (Jensen, 1998). The power to now separate the two – the yardstick from what it measures – is precisely what allows us to study the empirical phenomenon that IQ tests measure independent of the particular devices commonly used to measure it – in other words, to convincingly repudiate the false claim that "intelligence is what intelligence tests measure." All tests and tasks can now be characterized according to their degree of *g* loading, and the resulting patterns of *g* loadings across tests, jobs, subjects, ages, times, places, and settings allow us to test alternative hypotheses about the phenomenon they prod into greater or lesser action.

At the behavioral level, the *g* factor represents a general proficiency at learning, reasoning, thinking abstractly, and otherwise processing information efficiently and accurately almost without regard to type of the information being processed. *g* is a highly general ability because its practical value in enhancing performance is not restricted to any particular content domain: it is generalizable across many. Because cognitive tests of all types tend to measure *g* more than anything else, tests in one content domain generally predict performance in their own domain little better, if at all, than tests meant for other content domains. This fact, which has been known for decades (e.g., Jencks et al., 1979), is referred to as validity generalization in the personnel selection literature.

Some critics nonetheless still point to specific item attributes or item content on IQ tests to argue that they cannot possibly measure what we already know they do, namely, a general learning or reasoning ability, as if intelligence would necessarily mirror the superficialities of the items used to activate it. Based on their reading of these superficialities, critics usually posit some narrow facility at esoteric or highly structured tasks with no relevance to real life. Perhaps ironically, the most striking observation about intelligence tests such as the WISC-IV and WAIS-III and -IV is that they include such a great variety of items and those items are not similar in any obvious way. The items may use words, numbers, shapes, pictures, or blocks; require specific cultural knowledge or only universal concepts such as up/down and large/small; use language, paper, or pencils – or not.

This brings us to the key question: what, in fact, do IQ items share in common that allows them to activate so effectively the same general ability? Put another way, what is the chief active ingredient in IQ tests that allows them to call *g* forth on demand to be observed and measured? The answer: the complexity of the information to be processed. The more complex the test item or task, the more highly the quality of its performance correlates with individual differences in IQ (the more *g loaded* it is). Any number of elements can increase a task's complexity: amount of information to be integrated, presence of irrelevant information, abstractness, uncertainty and ambiguity, inferences required, and the like. Task complexity is the active ingredient in IQ test items, the source of differential item difficulty on adult literacy tests, the chief barrier to comprehension of health education materials, and the reason why cognitive tests predict job performance successively better in successively more complex jobs (Gottfredson, 1997, 2004). As discussed later, this knowledge also provides leverage to school psychologists, health workers, and others in their clinical work with individuals with cognitive deficits.

HOW ARE INDIVIDUALS DISTRIBUTED ALONG THE IQ CONTINUUM?

To the uninitiated, any distribution might seem plausible. The distribution of IQs within a population might conceivably resemble its distribution of income.

In most societies, mean income is much higher than median income because a small proportion of individuals make many multiples the typical income. Alternatively, the distribution might be rectangular, that is, evenly populated from low to high across the so-called normal range of IQ 70–130. On the other hand, if general intelligence is labile, then perhaps the distribution of IQs is unpredictable and fluctuates over time. In actuality, mental test performances appear to be fairly normally distributed at any given age (even before being statistically "normalized"), as are many biological traits. This alone is an interesting fact about general intelligence.

All populations manifest a *g* continuum of cognitive differences, and, as we can best discern, the members within each are normally distributed in IQ. The continuum itself seems isomorphic across all populations examined thus far, but they commonly differ in where along the IQ continuum they are centered or how spread out they are along it, that is, their means and standard deviations tend to differ somewhat. Small to moderate differences by both racial and ethnic heritage appear to be the rule, not the exception, both within and outside the United States and Europe (Frisby & Reynolds, 2005).

There are other striking regularities in the distribution of IQ scores within populations and hence, presumably, of *g* itself. At least in developed nations, IQs are fairly stable from late childhood to late adulthood. For example, IQ at age 11 correlated about 0.7 with IQ near age 80 in a Scottish cohort born in 1921 (Deary et al., 2004). Stability refers to maintaining similar rank within one's age group, because IQs are calculated as deviations from the average of one's age cohort. (The average IQ is set to 100 at every age.) The stability of IQ scores thus does not by any means suggest that intelligence level is "fixed," but only that individuals within an age cohort tend to retain the same rank in IQ as their age cohort rises or declines in average ability in absolute terms. If IQs were not stable in this sense, cognitive tests would not predict later outcomes nearly as well as they do.

Lacking ratio-level measurement, we cannot yet measure an individual's change in absolute ability level from one age to the next, but there is sufficient evidence to gauge general trends by age. General mental horsepower, as indexed by fluid *g*, increases rapidly in childhood, peaks in early adulthood, and then gradually declines into old age, though it may fall precipitously shortly before death. Very general language skills and cultural knowledge, indexed by crystallized *g*, are more robust and tend not to decline substantially until later in life. There is variation around these trend lines, of course, but they seem quite regular in the populations studied thus far.

Overall cognitive proficiency clearly changes over the lifetime, first rising and then declining. Whether it is malleable, that is, amenable to intentional change, is a different question. Adoption studies and interventions to raise low IQs have both been somewhat disappointing on this score, though many scholars would say the jury still out (Neisser, 1998; Williams, 2000; Flynn, 2007). It is fair to say, however, that there is as yet no effective and feasible means of substantially enhancing general intelligence in any significant proportion of the population, except perhaps in the most malnourished regions of the world. There is some evidence that it may be possible, however, to forestall some of the cognitive declines with advancing age. But saying that intelligence may be improvable or recoverable does not mean that it is equalizable across a population. Evidence for the former is sometimes mistaken as evidence for the latter.

The stability of IQ is commonly regarded as a problem to be solved. It is most certainly a problem for low-*g* individuals and a constraint on the ease and effectiveness with which families and organizations can intervene to improve performance in schools, work, health self-care, and other important life arenas. Yet, from some perspectives, the stability of general intelligence might be considered a blessing. It is hard to imagine how human intelligence could ever have evolved had it not been robust in the face of adverse circumstances – which must have been the lot of humankind throughout evolution.

In summary, intelligence tests reveal that phenotypic (observed) differences in general intelligence (*g*) are systematically patterned and probably a highly regular and predictable feature of human populations the world over.

WHAT IS THE PERSONAL AND SOCIAL IMPORT OF DIFFERENCES IN GENERAL INTELLIGENCE (g)?

Diverse sorts of research have confirmed the common intuition that differences in general intelligence have considerable practical importance in our lives, individual and collective (Jencks et al., 1979; Herrnstein & Murray, 1994; Ceci, 1996; Gottfredson, 1997a; Sternberg & Grigorenko, 2002; Lubinski, 2004). Most studies to date have focused on how higher intelligence levels improve the life chances of individuals, but the controversy over intelligence testing stems primarily from the societal-level waves generated by the resulting socioeconomic inequalities. The depth, strength, and patterning of those waves depend, in turn, primarily on the distribution of *g* because it overwhelms all other cognitive abilities at the population level.

I focus below on the real-world consequences for individuals of variation in *g*, whether or not it is ever measured. Using test scores to make decisions about individuals and organizations obviously can affect their well-being, but that is not the issue here. Making decisions based on test results constitutes social intervention rather than an empirical finding about the attribute being measured.

First, IQ/*g* is the best single predictor of a wide variety of outcomes in modern life, from performance on jobs and standardized achievement tests, to level of income and occupation, to law-abidingness, bearing children within marriage, and health self-care. IQ typically outperforms standard measures of socioeconomic advantage. Specific abilities add little incremental validity, as evidenced the fact that a whole battery of cognitive tests does little better than *g* alone in predicting differences in such outcomes.

The prognostic value of *g*/IQ differs greatly, however, by kind of outcome. That is, *g* affects a person's odds of success far more in some arenas than others. Moreover, these gradients of effect reflect the psychometric properties of everyday life (Gordon, 1997). For example, when everyday life is conceptualized as an IQ test writ large, its separate demands can be characterized by how *g* loaded they are, the extent to which outcomes depend exclusively on the person's own independent efforts, and how reliably and objectively the outcomes are evaluated. Because demands differ predictably in these regards, so do *g*'s gradients of effect.

Correlations with IQ (corrected for range restriction) extend, roughly, from about 0 for happiness, to −0.2 for delinquency, 0.3–0.4 for income (correlations rise by mid-career), 0.6 for education and occupation level attained, and 0.8 for standardized academic achievement. The odds of unfavorable outcomes on dichotomous measures also differ greatly by kind of outcome. Consider, for instance, the steadily greater odds for unfavorable outcomes of the following sort when comparing adults whose IQs are somewhat below average (IQs 75–90) to those who are somewhat above average (IQs 110–125): unemployed or getting divorced within 5 years (3:2), women bearing an illegitimate child (4:1), living in poverty (5:1), and mothers going on welfare after birth of first child (5:1) or being a chronic welfare recipient (8:1; Gottfredson, 1997, Table 10). Differences in *g* are not solely responsible for such differences, but they account for many of the upstream problems, such as school failure, contributing to them. The new fields of cognitive epidemiology and health literacy are also documenting pervasive differences in morbidity and mortality by level of IQ or aptness in learning, reasoning, and problem solving (Deary et al., 2004; Gottfredson, 2004), regardless of income or access to health care. Evidently, material resources do not substitute for mental resources in health either.

The differences in tightness of linkage of IQ with socio-economic outcomes arise in part because institutional practices often limit the control that individuals have over certain kinds of performance or whether their performance counts toward advancement (e.g., pay scales and promotions based on seniority), and partly because other personal traits and circumstances can make a big difference in some outcomes (e.g., conscientiousness for job performance, and family background for educational opportunities). In general, *g* predicts best when tasks are instrumental rather than socio-emotional and when individuals perform them independently of co-workers. This includes most work in school, training, jobs, health – and cognitive tests.

When individuals have reasonable control over their own performance, *g* predicts performance best when the tasks to be performed are more complex, as was noted earlier. Task complexity best explains, for example, why the correlations between mental test scores and job performance (corrected for range restriction) increase from about 0.2 in the simplest, most routine, and highly supervised factory and service jobs, to 0.5 in much clerical, crafts, and protective services work, to nearly 0.8 in highly demanding executive-level and professional jobs. Job analyses chart the increasing variety, fluidity, cognitive complexity, and learning demands of the task constellations comprising jobs higher up the occupational ladder. Higher-level jobs are like more difficult intelligence tests. Both civilian and military studies show that higher *g* typically overtakes the advantages of longer experience within several months to several years on the job.

Conversely, the distribution of task complexities in a setting allows us to forecast when and where differences in worker performance will be largest and smallest. Schools, for example, pose an increasingly complex set of cognitive tasks for students to master as they ascend in grade level, which gives an increasing edge in performance to brighter pupils and creates an increasingly grueling if not punishing experience for low-*g* individuals. And as pointed out in Chapter 1 (Weiss, Beal, Saklofske, Pickiam-Alloway, & Prifitera, 2008), learning also becomes more autonomous, instructions lengthier, and visual aids fewer in number as lessons become more complex – all of which can serve to widen the gap between children of average abilities and those with impairments. There is no more prolonged, relentless, highly public de facto test of general intelligence than that which the public schools administer on a daily basis in the normal course of instruction. As psychometricians know, the longer a test is, the better it will discriminate between individuals of low and high cognitive ability. In this case, the increasing discrimination by ability level manifests itself in widening gaps in academic achievement (except, of course, where achievement levels are artificially capped by limiting opportunities for brighter students to advance in the curriculum).

We can be grateful that *g* level does not correlate with happiness (subjective well-being), but higher levels of *g* do confer greater resilience in the face of adversity. Brighter children are better able to weather destructive rearing circumstances, and brighter soldiers are less susceptible to post-traumatic stress syndrome. As described in the accident literature, spotting hazards and avoiding accidental injury are quintessentially cognitive tasks, and research is beginning to confirm that a lower IQ increases risk of preventable injury and death. Demands for learning and reasoning are not limited to the classroom or workplace.

Finally, although it may be the single most useful mental tool in the larger human toolkit, *g*, like any tool, yields good results only when used diligently and wisely. High *g* allows for high levels of achievement but never guarantees them. General intelligence is not an achievement, but the mental capacity for it. Certain other personal traits, such as conscientiousness, also have general utility. They cannot compensate for low *g* when the situation requires apt learning and reasoning, but they can somewhat buffer the consequences of low *g*. For example, sociability and agreeableness allow individuals who are mildly mentally retarded to elicit more assistance, sustain a network of friends, and sometimes pass as "normal."

A corollary is that we, as professionals, ought to use wisely our knowledge about *g* when seeking to assist these and other individuals at risk. We know that intelligence is a general facility for learning and reasoning which becomes increasingly advantageous as learning and reasoning tasks become more complex. This means, conversely, that complex tasks pose bigger cognitive barriers to individuals who learn and reason poorly. We simply cannot expect all individuals to master complex learning tasks equally well when given the same amount of time and support to do so. Tasks that seem mindlessly simple to many people may be prohibitively difficult for some. That is why health educators advocate limiting the complexity of informational materials to the equivalent of a Grade 5 reading level – to provide low-literacy patients cognitive access to critical information.

Neither employers nor schools can reduce instruction to the lowest common dominator (nor should the health system, for that matter). Studies in both settings indicate, however, that learning is enhanced for all when the complexity of instruction is tailored to meet learning needs at different levels of the IQ continuum (Cronbach & Snow, 1977; Snow, 1994). To illustrate, between IQs 70 and 80, individuals tend to profit most from instruction that is highly concrete, step by step, involves much repetition and feedback, is tightly focused on the bare essentials, and is delivered one on one (Gottfredson, 1997). In the neighborhood of IQs 90–105, individuals typically profit most from programmed or mastery learning approaches that provide ample time and hands-on training experiences. This contrasts with individuals at the high end of the IQ distribution, say, above IQ 115, who in many ways are self-instructing. They are able to gather and synthesize information on their own, as well as independently analyze and draw inferences from the often confusing, swirling mass of information enveloping us in daily life and work. High IQ individuals profit from instruction in theory, but low-IQ individuals do not. Low-IQ individuals require massive scaffolding of instruction, but that interferes with learning at high IQ levels.

A focus on task complexity also reveals much needless complexity in instructional settings, whose elimination can make the difference between failure and success for children with cognitive impairments, specific or general. Consider, for example, the suggestions for intervention in Part II of Chapter 1. Many focus on paring away non-essential complexity, often mirroring the strategies developed independently by health educators and work literacy researchers seeking to enhance learning among low ability patients (Doak et al., 1996) and military recruits (Sticht, 1975): eliminate distractions, reduce visual clutter, include more white space on written materials, use simple words and short sentences, break tasks into a clear sequence of steps, pace them to avoid overwhelming the student, use lots of repetition and feedback, allow extra time, use physical manipulatives and visual aids where helpful, limit the need for drawing inferences, and limit the task to its most critical elements (e.g., allow oral rather than written responses). Others constitute a "metacurriculum" (Snow, 1996) in the sorts of mental self-regulation and reflection that allow learners to deploy those resources more efficiently, such as by using memory aids, outlines, advance organizers, and strategies for parsing tasks and self-pacing. Techniques for reducing cognitive load are especially important for children with generalized intellectual deficits (low *g*). But they also help children with specific impairments, say, ones involving visualization or memory, by freeing up mental resources for deployment in their area of impairment.

Better tailoring the complexity of learning tasks to the ability levels of learners clearly does not result in equal amounts learned at all ability levels, but it does result in *more* learning for everyone. This happy pedagogical result comes with a political downside, however, which may help explain schools' reluctance to pursue it. Namely, interventions that succeed in raising average levels of performance generally increase the spread (standard deviation) of performance levels (Ceci & Papierno, 2005; see also Fuchs & Young, 2006, on responsiveness to instruction). In common parlance, variance around the mean is described as inequality, and inequality is perceived as a bad thing to be eradicated.

Together, the foregoing generalizations allow us to predict with clockwork-like accuracy the ramifications of intelligence differences across social and political life. For instance, employee selection psychologists can predict the amount of disparate impact (racial disparities in test results) they will observe in different sorts of applicant pools if they administer selection batteries with certain combinations of cognitive and non-cognitive assessments rather than others (Schmitt et al., 1997). They want this information to determine which batteries of valid unbiased tests would, if implemented, put companies at greatest risk of employment discrimination lawsuits. These generalizations could also have been used to foresee the impact of the US Congress's 2001 *No Child Left Behind Act*, which mandates that the public schools eliminate achievement gaps between social groups by raising all children to a high level of academic proficiency by the year 2014. They forecast not just the Act's certain failure but also its side-effects, including the distorted testing and teaching strategies to which schools are turning to appear to satisfy it without actually doing so.

WHERE DO INTELLIGENCE DIFFERENCES ORIGINATE AND RESIDE?

Research on the physiology and genetics of intelligence helps to explain the regularity and stability of phenotypic differences in cognitive abilities. Perhaps ironically, it also allays common fears about what a sturdy biological basis for many individual differences would mean.

Individual differences in IQ are firmly but not entirely rooted in the genetic differences among individuals (Sternberg & Grigorenko, 1997; Plomin et al., 2001). This fact has been known far longer than it has been widely accepted. In fact, some behavior geneticists are concluding that variation in *all* human traits – abilities, personality, attitudes, interests, sexual preference, and so on – must now be presumed to be at least somewhat heritable until proved otherwise. The broad Stratum II abilities are somewhat less heritable than *g*, but their phenotypic correlations with *g* seem to be entirely genetic. That is, to the extent that they correlate with *g*, they arise from the same genetic sources. Neither shared nor non-shared environments contribute to their overlap with *g*.

Most of the broad abilities that behavioral geneticists have analyzed so far – but especially memory – are also influenced, though to a far lesser degree, by genes specific to that ability and not shared by *g*. This is not surprising because, unlike *g*, primary abilities such as verbal and spatial ability can be associated with particular regions of the brain. Differences in *g*, on the other hand, correlate with cross-brain attributes such as joint mobilization of particular Brodmann's Areas across the brain, volume of white and gray matter, dendritic branching, and such. The pre-frontal lobes may be especially important to *g*, but they are not the seat of it. A useful analogy is that *g* is like a giant turbine which powers an array of small motors, each devoted to different specialized activities. Consistent with findings in neuropsychology, the smaller motors do not necessarily work with equal efficiency and any one of them could be knocked off-line without materially affecting the turbine powering them all.

More surprising are several findings that even the behavior geneticists did not expect, and which help to explain the empirical regularities involving phenotypic *g*. One is that the heritability of IQ increases, not decreases, with age: from perhaps 0.2 in infancy (to the extent *g* can be measured at that age), to 0.6 by adolescence, to 0.8 by middle to late adulthood in typical Western research populations. That is, our phenotypic differences in intelligence line up more closely with our genotypic differences as we become more independent and freer to select and shape our own learning opportunities.

A second surprise is that the only non-genetic influences on general intelligence that last into adolescence originate in our non-shared environments. These are the influences that affect siblings one by one, not uniformly as members of the same household – say, an illness rather than parental education. Both unexpected findings mean that our IQs actually become less similar to those of parents and siblings as we age, because the shared environmental influences that boosted our similarities in childhood fade away as we become independent adults. The IQs of biological relatives remain moderately similar for genetic reasons, however, because we share exactly half of our genes with each parent and half, on average, with biological siblings. An important societal-level consequence of the 50% genetic *dissimilarity* among biological parents and children, combined with the high heritability of IQ, is that it generates considerable intergenerational social mobility (up and down) in societies allowing individuals to rise on the basis of their own merits rather than family advantage.

Also a surprise to many, but not to behavior geneticists, is the finding that personal environments and life events are also somewhat heritable, especially those over which we exercise obvious control such as having serious marital or financial troubles. Not surprisingly, then, the phenotypic correlations between IQ, academic achievement, and years of education completed turn out to be mostly genetic in nature, but those for occupation and income successively less so (Rowe et al., 1998). The environment is not just "out there" but is perceived, shaped, and exploited differently depending on our own genetically conditioned proclivities and capabilities (Dawkins, 1999). For instance, the same classroom with the same teacher does not constitute the same environment or exert the same influence on genetically diverse students. Some students will be more willing or able to exploit the learning opportunities in that classroom. Some will build warm relations with their teachers, but others will consistently behave in a manner that engenders hostility among peers and teacher alike. Students are not identical lumps of clay which teachers can mold and remold in ways that we or they might wish. Children are born neither fungible nor passive, but ineradicably individual, self-constructing, and resistant to fundamental redesign according to the dictates of others. That seems mostly a good thing, even if sometimes frustrating to parents and teachers.

None of this should be taken to mean, as is sometimes mistakenly suggested, that school achievement is fixed by intelligence level or that we should give up on low-*g* individuals. Far from it. As already noted, intelligence is not achievement itself, but a mental capacity that facilitates many forms of achievement. Everyone can learn more than they currently do and apply their knowledge more effectively. That is, we can always improve achievement to some degree without changing *g* at all. Behavior genetic research has already revealed that academic achievement, unlike *g* itself, is permanently influenced to some extent by shared environmental factors. In fact, such research may provide the most efficient and effective methodology for ferreting out which specific environments and experiences account for the *non-genetic* influences on academic achievement.

Perhaps the most surprising but illuminating genetic finding is this. Despite high hopes and a decade of large-scale studies, no gene (to my knowledge) has yet been unambiguously identified that influences intelligence level within the normal range. Single gene mutations and chromosomal anomalies tend to send individuals straight to the low end of the IQ continuum, so they are easier to identify. The failure to locate genes for normal variation along the IQ continuum suggests that it originates in the actions of many genes with only small individual effects on the phenotype. Evincing this expectation, geneticists now pin their hopes on finding genes that account for at most 0.5% of the variance in IQ – and they do not seem terribly optimistic about that either.

Their challenge seems consonant with the growing body of results from research on the electrical, chemical, metabolic, structural, and functional properties of the human brain. They suggest that all elements of the brain are correlated with IQ differences to some extent: rate of glucose metabolism (negatively), complexity of EEG waves, size of various regions of white and gray matter, and so on (Vernon, 1993; Bock et al., 2000; Deary, 2000; Jung & Haier, 2007). Moreover, these phenotypic correlations turn out to be mostly genetic when the necessary data are available to investigate the matter. There are probably no genes for intelligence as such, but rather for many particulars in the brain, such as degree of nerve myelination and amount of dendrite branching at the end of nerve axons. Myelination shows the same developmental course as does IQ (it increases during childhood and decreases in adulthood), and dendritic branching correlates with differences in IQ.

The brain is also responsive to non-genetic influences, partly for genetic reasons. The genome operates more like a playbook than a blueprint, with contingencies for all manner of environmental conditions. Biologic factors such as drugs and lack of sleep can cause large day-to-day fluctuations in mental efficiency. Other posited non-genetic influences may have longer term although not necessarily permanent effects, such as intensive training and practice, malnutrition, or serious illness (Sternberg & Grigorenko, 2001). Socio-educational interventions to raise low IQs have not borne much fruit, but perhaps we have been looking in the wrong places for leverage in toning up intelligence and exploiting people's unused potential.

Some scholars wonder whether biologic factors such as these might explain the secular increase in IQ scores observed during the 20th century and throughout much of Europe and North America (Neisser, 1998). This steady increase of three IQ points per decade, dubbed the Flynn effect, is usually attributed to entirely socio-cultural influences, but this seems implausible in view of the fact that average heights have risen in tandem with IQs. I see no evidence that the rise in IQs is actually a rise in *g* itself, because the increases seem almost random with regard to the *g* loadings of the subtests on which scores have risen the most and least (Matrices and Similarities way up, but Vocabulary, Information, and Arithmetic not). Moreover, the heritabilities for IQ have not, to my knowledge, fallen as would be expected with the hypothesized injection of new environmental influences. This peculiar pattern of change and lack of change might be consistent, however, with potential biologic effects on the brain regions or specific abilities contributing variance (beyond *g*) to the subtests most responsible for boosting overall IQ. These localized effects could also be genetic, the effect of heterosis (hybrid vigor) resulting from individuals increasingly marrying outside their villages and small towns (Mingroni, 2007).

The puzzle of the Flynn effect remains unsolved in large part because intelligence tests lack ratio-level measurement, that is, a scale that starts at zero quantity and counts in equal-size units from there. Without that capability, we cannot know whether the observed secular rise in average IQ reflects an absolute change in the level of any of the constructs captured by IQ tests, either in *g* itself or some normally inconsequential source of variance in the composite IQ score. Critics often implicitly assume that IQ scores can register changes in absolute ability level when they criticize intelligence tests. A currently popular critique in this vein is that the Flynn effect disproves what had come to seem incontrovertibly true, namely, that IQ tests measure a stable, heritable, psychometrically unitary general intelligence. Their argument is that *g* cannot at once be heritable, unitary, and yet some subtest scores increase dramatically over

time. Flynn (2007), for example, argues that the rise has "shattered *g*." It obviously has not. The problem is not with our conception or measurement of *g*, but with our measurement technology not yet allowing us to measure absolute rather than just relative positions along any ability continuum. This is the severest limitation of current IQ tests, yet to be overcome (Jensen, 2006), but it does nothing to invalidate the construct of *g* or the utility of current IQ tests for many diagnostic, selection, placement, training, and treatment purposes.

OF WHAT VALUE IS TESTING FOR INTELLIGENCE?

Intelligence testing is a highly sophisticated technical enterprise for measuring human cognitive diversity, itself perhaps the most socially consequential, genetically conditioned individual difference in modern life. Statistically and theoretically, testing has advanced tremendously since its Model-T days generations ago. Used with skill and responsibility, intelligence tests yield information of great value and precision not otherwise available to administrators, clinicians, and scientists. They can be misused, and they are often misunderstood or wrongly maligned (Gottfredson, in press). They are imperfect yardsticks, yet far superior to most others in the social sciences. If they and their results sometimes rouse controversy, it is usually because they have done their job all too well – illuminated human differences that are important to us. As this book shows, however, intelligence tests can be used to fashion more humane and constructive responses to those differences.

REFERENCES

Bartholomew, D. J. (2004). *Measuring Intelligence: Facts and Fallacies*. Cambridge: Cambridge University Press.

Bock, G. R., Goode, J. A., & Webb, K. (Eds.) (2000). *The Nature of Intelligence*. Novartis Foundation Symposium 233. Chichester, UK: Wiley.

Campbell, J. P., & Knapp, D. J. (Eds.) (2001). *Exploring the Limits of Personnel Selection and Classification*. Mahwah, NJ: Erlbaum.

Carroll, J. B. (1993). *Human Cognitive Abilities*. Cambridge: Cambridge University Press.

Ceci, S. J. (Ed.) (1996). Special theme: IQ in society. *Psychology, Public Policy, and Law*, 2(3/4), 403–645.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing – When the "Have-Nots" gain but the "Haves" gain even more. *American Psychologist*, 60(2), 149–160.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. New York: Irvington.

Dawkins, R. (1999). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.

Deary, I. J. (2000). *Looking Down on Human Intelligence: From Psychometrics to the Rain*. Oxford: Oxford University Press.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, 86, 130–147.

Doak, C. C., Doak, L. G., & Root, J. H. (1996). *Teaching Patients with Low Literacy Skills* (2nd ed.). Philadelphia, PA: Lippincott.

Fischer, C. S., Hout, M., Jankowski, M. S., Lucas, S. R., Swidler, A., & Voss, K. (1996). *Inequality by Design: Cracking the Bell Curve Myth*. Princeton, NJ: Princeton University Press.

Flynn, J. R. (2007). *What is Intelligence? Beyond the Flynn Effect*. Cambridge: Cambridge University Press.

Frisby, C. L., & Reynolds, C. R. (Eds.) (2005). *Comprehensive Handbook of Multicultural School Psychology*. New York: Wiley.

Fuchs, D., & Young, C. L. (2006). On the irrelevance of intelligence in predicting responsiveness to reading instruction. *Exceptional Children*, 73(1), 8–30.

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

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, 24(1), 203–320.

Gottfredson, L. S. (Ed.) (1997a). Intelligence and social policy. *Intelligence*, 24(1), 1–320 (special issue).

Gottfredson, L. S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence*, 24(1), 79–132.

Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, 86, 174–199.

Gottfredson, L. S. (in press). Logical fallacies used to dismiss the evidence on intelligence testing. In R. Phelps (Ed.), *The True Measure of Educational and Psychological Tests: Correcting Fallacies About the Science of Testing*. Washington, DC: American Psychological Association.

Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Jencks, C., et al. (1979). *Who Gets Ahead? The Determinants of Economic Success in America*. New York: Basic Books.

Jensen, A. R. (1980). *Bias in Mental Testing*. New York: Free Press.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Amsterdam: Elsevier.

Jung, R. E., & Haier, R. J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavior and Brain Sciences*, 30, 135–187.

Lubinski, D. (Ed.) (2004). Cognitive abilities: 100 years after Spearman's (1904) "General Intelligence, Objectively Determined and Measured" (Special section). *Journal of Personality and Social Psychology*, 86, 96–199.

Mingroni, M. A. (2007). Resolving the IQ paradox: Heterosis as a cause of the Flynn effect and other trends. *Psychological Review*, 114(3), 806–829.

Neisser, U. (Ed.) (1998). The Rising Curve: Long-term Gains in IQ and Related Measures. Washington, DC: American Psychological Association.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics* (4th ed.). New York: Worth.

Roberts, M. J. (Ed.) (2007). *Integrating the Mind: Domain General Versus Domain Specific Processes in Higher Cognition*. New York: Psychology Press.

Rowe, D. C., Vesterdal, W. J., & Rodgers, J. L. (1998). Herrnstein's syllogism: Genetic and shared environmental influences on IQ, education, and income. *Intelligence*, 26, 405–423.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg, & R. K. Wagner (Eds.), *Mind in Context: Interactionist Perspectives on Human Intelligence* (pp. 3–37). Cambridge: Cambridge University Press.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, 2(3/4), 536–560.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201–293.

Sternberg, R. J., & Grigorenko, E. L. (Eds.) (1997). *Intelligence, Heredity, and Environment*. Cambridge: Cambridge University Press.

Sternberg, R. J., & Grigorenko, E. L. (Eds.) (2001). *Environmental Effects on Cognitive Abilities*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Sternberg, R. L., & Grigorenko, E. L. (Eds.) (2002). *The General Intelligence Factor: How General Is It?* Mahwah, NJ: Erlbaum.

Sticht, T. G. (Ed.) (1975). *Reading for Working: A Functional Literacy Anthology*. Alexandria, VA: Human Resources Research Organization (HumRRO).

Vernon, P. A. (Ed.) (1993). *Biological Approaches to the Study of Human Intelligence*. Norwood, NJ: Ablex.

Weiss, L. G., Beal, A. L., Saklofske, D. H., Pickiam-Alloway, T., & Prifitera, A. (2008). Interpretation and intervention with WISC-IV in the clinical assessment context. In A. Prifitera, D. Saklofske, & L. G. Weiss (Eds.), *WISC-IV Applications for Clinical Assessment and Intervention* (pp. 1–66). Amsterdam: Elsevier.

Williams, W. M. (Ed.) (2000). Special themes: Ranking ourselves: Perspectives on intelligence testing, affirmative action, and educational policy (Special issue). *Psychology, Public Policy, and Law*, 6(1), 1–252.