23 Problem-Solving and Intelligence

David Z. Hambrick, Alexander P. Burgoyne, and Erik M. Altmann

The ability to solve complex problems is a defining feature of what most laypeople think of as *intelligence.* This is also a common theme in how intelligence researchers describe intelligence (Sternberg, 1985a; Sternberg et al., 1981). Over a century ago, the German psychologist Wilhelm Stern (1914), who introduced the formula for computing the *intelligence quotient*, defined intelligence as "a general mental adaptability to new problems and conditions of life" (p. 101). And, in his early book on aptitude testing, Bingham (1937) explained that "[w]e shall use the term 'intelligence' to mean the ability of an organism to solve new problems" (p. 36). More recently, fifty-two intelligence researchers published a letter in the *Wall Street Journal* in the wake of the *Bell Curve* controversy, defining intelligence as "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" (see Gottfredson, 1997, p. 13).

Outline of Chapter

In this chapter, we discuss the link between intelligence and problem-solving in terms of contemporary ideas concerning both. To preview, we argue that the ability to solve problems is not just an aspect or feature of intelligence – it is the essence of intelligence. The chapter is organized into five major sections. In the first section, we consider the question of what a problem is and argue that all "intelligent" behavior can be viewed as problem-solving behavior. In the second section, we briefly review evidence from psychometric research concerning the nature of individual differences in intelligence, and then review evidence for how intelligence relates to complex problem-solving. In the third section, we consider the question of what mechanisms might underlie both problem-solving and intelligence, focusing on some of our own research. In the fourth section, we briefly review evidence for the predictive validity of intelligence and practical uses of intelligence tests. In the fifth section, we consider the question of whether intelligence as problem-solving ability can be improved through training. We close with directions for future research.

What Is a Problem?

A problem is nothing more (or less) than a goal that is not immediately attainable. As Duncker (1945) wrote, "a problem exists when a living organism has a goal but does not know how this goal is to be reached" (p. 2). And as Mayer (2013) explained, "Problem solving refers to cognitive processing directed at achieving a goal when the problem solver does not initially know a solution method. A problem exists when someone has a goal but does not know how to achieve it" (p. 769). Examples of problems range from the mundane – needing to get around a traffic jam commuting to work – to the existential – finding meaning in life. From a psychological perspective, and more particularly a cognitive perspective, the goal of research on problem-solving is to describe the mental processes involved in reducing the "distance" between the problem's initial state and the goal state.

Psychologists have traditionally distinguished between two types of problems (see Mayer, 2013). In *well-structured* problems, the goal state is clearly specified, as is the "solution path" (i.e., the specific way one should go about solving the problem). The classic example of a well-defined problem from psychological research is the Tower of Hanoi problem. The initial state in this problem is that there are three disks of three different sizes on one of three pegs. The goal is to move the disks from the left peg to the right peg, moving one disk at a time and never placing a larger disk on a smaller disk. By contrast, in an *ill-structured* problem, neither the goal state nor the solution path is clearly specified; the problem-solver must generate both. Many complex real-world problems are of this type, from producing a work of art to figuring out how to make a living. Problems also differ in their complexity – the number of steps that they involve – and in their novelty to the problem-solver – the degree to which the problem-solver has knowledge and skills that can be applied to the problem.

It is easy to think of everyday tasks that differ along these dimensions. A relatively simple well-defined problem is adding two numbers; a more complex one is filling out your tax return. A relatively simple ill-defined problem is writing a personal essay for a college application; a more complex one is planning for retirement. These tasks have the "feel" of problem-solving, but other complex tasks can be regarded as problems as well, in the sense that they involve goals that are not immediately attainable. As Anderson (1985) observed, "It seems that all cognitive activities are fundamentally problem solving in nature. The basic argument . . . is that human cognition is always purposeful, directed to achieving goals and to removing obstacles to those goals" (pp. 199–200).

Consider Raven's Progressive Matrices, a widely administered test of nonverbal intelligence developed by John C. Raven in 1936. As shown in Figure 23.1, the initial state of each test item is a series of graphical elements (or patterns) arranged in a 3×3 matrix, with the element in the lower right cell missing. The goal state is a completed series, and the test-taker's task is to identify the alternative that accomplishes this. To remove the distance between the initial state and the goal state, the test-taker must develop hypotheses about how the elements change across rows and/or down columns, and then test the hypotheses to determine which alternative is correct. In the example below, inspection of the top row suggests that each row must contain

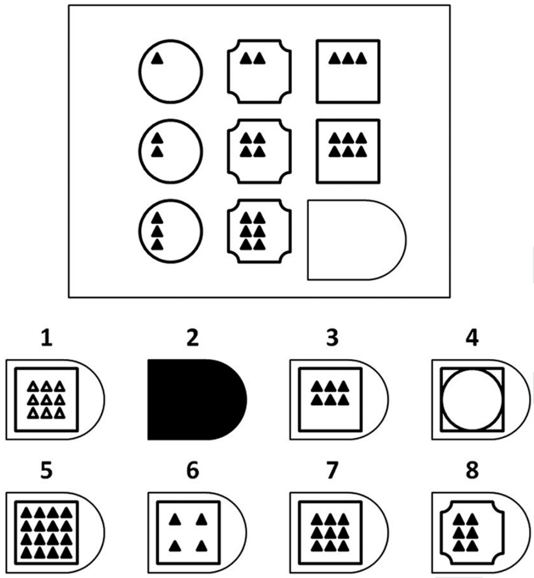

Figure 23.1 *Example of a Raven's-like problem (from Kunda et al., 2016). Reprinted with permission of Elsevier.*

one instance of each large shape, while inspection of the first column suggests that each column must contain elements of the same type. Inspection of the next rows and columns confirm these hypotheses, eliminating Alternatives 2 and 8. In turn, inspection of the small filled triangles indicates that the number of columns of triangles increases across each column, and the number of rows of triangles increases down each row. Only Alternative 7 contains both three rows and three columns of filled triangles; therefore, it is the only option that completes the series.

Even "low-level" cognitive tasks that are designed to isolate specific cognitive processes can be considered from a problem-solving perspective. Consider the Stroop task (Stroop, 1935). In the original version of Stroop, the research subject (or patient when the task is used for neuropsychological diagnosis) is given a sheet of paper with color names printed in conflicting colors (e.g., "red" printed in green), and instructed to read the words as quickly as possible. On another sheet, color names are printed in black, and again subjects are instructed to read the words as quickly as possible. As has now been replicated countless times, subjects are slower to read the words in the former condition. Stroop is often described as a test of "inhibition" (e.g., Miyake et al., 2000) – the ability to override or suppress an overlearned response, of the type that an American must exercise when driving on the left side of the road in the UK. At the same time, the task can be viewed as a problem-solving task in that it involves keeping in mind a goal to direct responses in the task ("read the word").

If all cognitive activity can be considered problem-solving, and if intelligence reflects the efficiency and effectiveness of all cognitive activity, then it follows that problem-solving ability is not just an aspect of intelligence – it is the essence of intelligence. In the next section, we briefly consider what is known about individual differences in intelligence, and then discuss evidence concerning the relationship between intelligence and complex problem-solving.

Intelligence as Problem-Solving Ability

As first documented by Spearman (1904) more than a century ago, one of most replicated empirical findings in the entire field of psychology is that a person who performs well on one cognitive test (or task) will tend to perform well on all other cognitive tests. At the time, many psychologists subscribed to the view that the mind is a collection of a number of separate and independent abilities, or *faculties.* The assumption was that there was a faculty for practically every aspect of mental functioning – one for perception but another for memory, one for reasoning but another for intuition, and so on. To test this idea, Spearman computed correlations between grades in university courses for a sample of Columbia University students. As shown in Table 23.1, the correlations among the grades were uniformly positive. Therefore, there was a good chance that a student who did well in, say, Classics did well in all of the other subjects. Grades also correlated with a measure of pitch discrimination.

The correlations were all positive and relatively high. Pioneering the use of factor analysis in psychology, Spearman further demonstrated that each of the measures correlated very highly with the general factor implied by this “positive manifold,” from a low of 0.72 for Music to a high of 0.99 for Classics. He concluded that performance on any given test of mental ability is a function of a factor reflecting something that the test shares in common with all of the others – a “common fundamental Function” (Spearman, 1904, p. 273) – along with an ability unique to that test. Spearman referred to the common factor as the general factor of intelligence, and it has since become known as “Spearman’s *g*.” In statistical terms, Spearman’s *g* is the first principal factor in a factor analysis of cognitive ability measures; a test’s “*g* loading” is the correlation between this factor and the measure of performance from the test (Jensen, 1999).

Table 23.1 *Correlation matrix (from Spearman, 1904)*

	Classics	French	English	Math	Pitch	Music
Classics	–					
French	0.83	–				
English	0.78	0.67	–			
Math	0.70	0.67	0.64	–		
Pitch	0.66	0.65	0.54	0.45	–	
Music	0.63	0.57	0.51	0.51	0.40	–

From Spearman (1904), p. 275.

Subsequent research focused on the question of whether *g* can be fractionated – that is, whether there is one intelligence or multiple intelligences. Cattell (1943) observed that certain measures of mental ability correlate more positively with each other than do others, leading him to propose that there are two factors of intelligence. "Fluid ability," he explained, "has the character of a purely general ability to discriminate and perceive relations between any fundaments, new or old," whereas "[c]rystallized ability consists of discriminatory habits long established in a particular field, originally through the operation of fluid ability, but not [sic] longer requiring insightful perception for their successful operation" (Cattell, 1943, p. 178). The theory of fluid ability (G*f*) and crystallized ability (G*c*) went on to have enormous impact in scientific thinking about intelligence.

It should be noted that Cattell's (1943) G*f*-G*c* distinction bore a remarkable similarity to Hebb's (1942) distinction between Intelligence A ("intellectual power") and Intelligence B ("intellectual products") (Hebb, 1942, p. 290). In fact, as Brown (2016) has recently documented, after attending a presentation given by Hebb at the 1941 American Psychiatric Association (APA) meeting, it appears that Cattell adopted Hebb's theory, simply relabeling Intelligence A and Intelligence B. As letters between Cattell and Hebb, and between Cattell and Hebb's department head, George Humphrey, reveal, Cattell grudgingly acknowledged Hebb's influence on his theorizing in his 1943 *Psychological Bulletin* article (Cattell, 1943). However, in later articles (e.g., Cattell, 1963), Cattell took full credit for G*f*-G*c* theory without acknowledging Hebb. At various points, Cattell also claimed that he had proposed the fluid-crystallized distinction either during the 1941 APA meeting or before it, but there is no record of him doing so.

This matter aside, later factor-analytic research with Horn supported the distinction between G*f* and G*c* factors. G*f* was found to have its highest loadings on nonverbal reasoning measures such as Raven's, and G*c* its highest loadings on measures reflecting acculturated learning, including vocabulary and general information. Subsequently, in a landmark project, Carroll (1993) compiled and reanalyzed the results of over 460 factor-analytic studies, and found that mental abilities can be arranged into a hierarchy that includes three levels, or "strata." At the highest level of the hierarchy (stratum I) is Spearman's *g*, representing what all tests of cognitive ability share in common; at the next level (stratum II) are "broad" cognitive abilities, including G*f* and G*c*; and at the lowest level (stratum III) are "narrow" cognitive abilities, representing demands unique to particular types of tests.

There has been, and continues to be, vigorous debate in the intelligence literature about the structure of cognitive abilities. McGrew and colleagues (see McGrew, 2005, 2009) merged Cattell and Horn's G*f*-G*c* model and Carroll's three-stratum model into the Cattell-Horn-Carroll (CHC) model. Meanwhile, extending Vernon's (1965) verbal-perceptual model, Johnson and Bouchard (2005) presented evidence to indicate that a model with *verbal, perceptual*, and *image rotation* factors was better fitting than either G*f*-G*c* or CHC models. Thinking about intelligence more broadly, Sternberg (1985b) proposed distinctions among three intelligences: *practical, creative*, and *analytical*.

The differences among these models are important, but from a historical perspective, it also worth pointing out a salient similarity, which reflects a scientific consensus about the nature of intelligence that has emerged through decades of intensive research and debate. All these models assume that there are far fewer factors underlying variation in scores on tests of mental ability than there are tests of mental ability. In other words, these models assume that the human intellect comprises some number of relatively general factors, and cannot be adequately (or easily) understood as a collection of independent modules as a proponent of the faculty view of the mind would have argued in 1900, stimulus-response associations as a radical behaviorist would have argued in 1950, or highly specific skills as a neobehaviorist might try to argue today. This assumption, obvious only with the benefit of hindsight, is the foundation for the whole enterprise of developing tests to measure intelligence and using scores on these tests for practical purposes such as predicting job performance. The perspective that we advance here is that, whatever factors a theory of intelligence posits, they can be considered from a problem-solving perspective.

Intelligence and Complex Problem-Solving

A different tradition of research on intelligence (the *European perspective*) has focused on the question of how intelligence relates to complex problem-solving (CPS), which Buchner defined as "the successful interaction with task environments that are dynamic (i.e., change as a function of user's intervention and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process" (Frensch & Funke, 1995, p. 14). CPS is assessed using "microworlds" that may be simulations or abstractions of real-world tasks. The best-known example of a CPS microworld is Tailorshop (see, e.g., Danner et al., 2011). The subject's task is to manage, over a period of twelve "virtual" months, a garment-manufacturing company, making decisions about hiring, pricing, advertising, and so on. Changes in these "input" variables during one cycle (month) lead to changes in "output" variables in the next cycle, such as company value, monthly sales, and customer demand. An example of a more abstract CPS task is a paradigm called MicroDYN (see Figure 23.2). The MicroDYN approach consists of a series of short (e.g., 5-minute), moderately complex tasks in which subjects must explore a novel task environment, generate knowledge about the relationships between input and output variables (e.g., how physical training methods affect outcomes for a handball team, such as power of throw), and then apply this knowledge to reach a specified goal (e.g., to combine training methods to achieve a certain level of the outcomes). Unlike in simulations of real-world tasks like Tailorshop, prior knowledge has little impact on performance in MicroDYN tasks because variable labels are either fictitious or lacking deep semantic meaning (e.g., feed imaginary animals food A, B, or C; Greiff, Wüstenberg, & Funke, 2012).

A long-standing question in this research is whether intelligence and CPS are the same or different constructs. A comprehensive review of this literature is beyond the

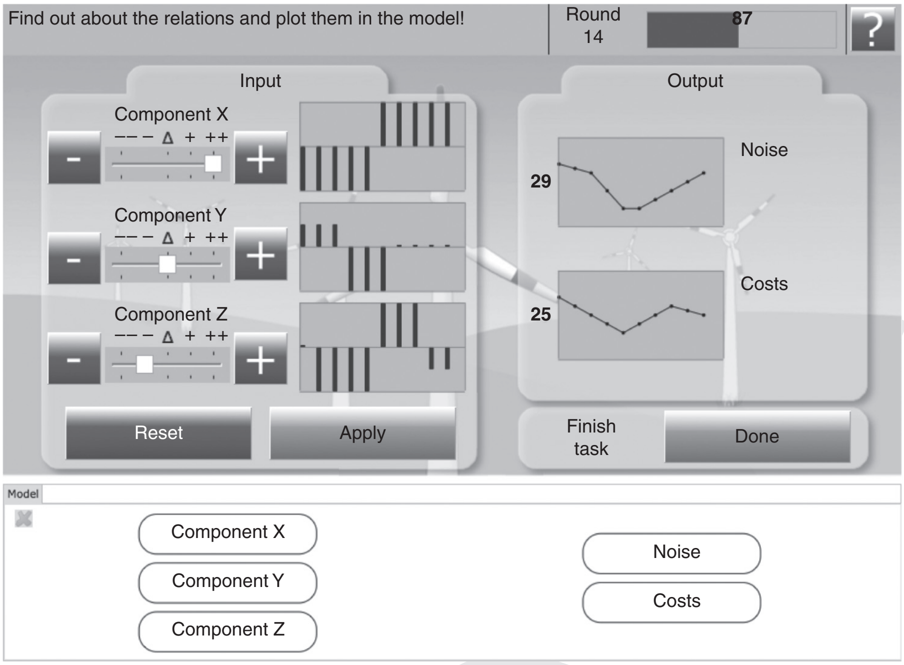

Figure 23.2 *Screenshot from MicroDYN (from wwwen.uni.lu/media/images/ microdyn). Reprinted with permission of Dr. Samuel Greiff.*

scope of this chapter, but suffice it to say that the two are significantly related. Results of a recent study by Kretzschmar and colleagues (2016) are illustrative. A sample of 227 German university students completed the Berlin Intelligence Structure test (an IQ test) and two CPS tasks (MicroDYN and MicroFIN). As is typical, *g* loadings for the measures were all positive, ranging from 0.26 to 0.83. The loadings for MicroDYN and MicroFIN were in the middle of this range (0.60 and 0.55, respectively; average *g* loading = 0.54). And although a model with a CPS factor improved model fit, this factor could be interpreted as a method factor (i.e., the MicroDYN and MicroFIN tasks use very similar procedures, and are more similar to each other than to either of the other cognitive ability tests). Finally, CPS improved prediction of an external criterion (grade point average, GPA) when intelligence was narrowly defined as figural reasoning (see also Greiff et al., 2013a; Greiff et al., 2013b; Wüstenberg, Greiff, & Funke, 2012). However, CPS did not improve prediction of GPA above and beyond a *g*-factor. Similarly, in a sample of 560 Luxembourgish high school students, Sonnleitner and colleagues (2013) found that the average correlation between a CPS factor and academic achievement variables dropped considerably (avg. $r = 0.39$ to 0.15) after G*f* (comprising multiple measures of reasoning ability) was statistically partialled from CPS.

A recent meta-analysis summarizing these and other findings (total studies = 47) estimated the overall correlation of CPS with intelligence at 0.43 (Stadler et al., 2015). The correlation was stronger for reasoning ability (indexing G*f*) than for overall measures of intelligence (indexing *g*); among different tasks used to measure CPS, it was strongest for those that involve managing multiple complex systems (avg. $r = 0.72$, after correction for unreliability). Correlations were less than 1.0 after corrections for unreliability, leading the authors to conclude that CPS and intelligence are distinct constructs. This may be, but an even stronger test of the distinction would be to test the correlation between a psychometric *g*-factor extracted from a diverse set of cognitive ability measures (reasoning, spatial visualization, working memory, comprehension, processing speed, etc.) and a *g*-factor extracted from diverse CPS paradigms (e.g., MicroDYN, Tailorshop, Genetics Lab; see Greiff et al., 2013, for a study that at least did the latter). What is already known from research in this area suggests that these *g*-factors would correlate very highly, and that a CPS *g*-factor would account for relatively small amounts of variance in external criteria above and beyond psychometric *g* (Sonnleitner et al., 2013).

Underpinnings of Intelligence as Problem-Solving Ability

To sum up, in our view, intelligence and problem-solving ability are closely connected, at both theoretical and empirical levels. There is a general capacity for solving novel problems that maps onto G*f*/Intelligence A, and there is knowledge acquired through the exercise of this capacity that maps onto G*c*/Intelligence B. Other labels for what are essentially the same broad factors include Ackerman's (1996) *intelligence-as-process* and *intelligence-as-product,* Baltes' (1987) *cognitive mechanics* and *cognitive pragmatics,* and Salthouse's (2000) *process* and *product* cognition.

But, at the level of the cognitive system, what about a person who is intelligent (and thus an effective problem-solver) differs from a person who is less intelligent (and thus a less effective problem-solver)? There has been a great deal of enthusiasm for the idea that G*f* can be equated with *working memory capacity* (WMC). As Baddeley and Hitch (1974) conceived of it nearly fifty years ago, working memory is a limited-capacity system for both storing and processing information in the service of complex cognition. Based on this model, as the first test of WMC, Daneman and Carpenter (1980) introduced the reading span paradigm. The participant reads a series of sentences (the processing task) while remembering the last word in each for later recall (the storage task), and then is cued to recall the words. Reading span is the number of sentences that the participant can read while maintaining perfect recall of the words. Subsequently, Engle and colleagues introduced operation span (Turner & Engle, 1989). The participant solves a series of arithmetic equations, remembering a word that follows each for later recall.

Beginning around 1990, there were numerous reports of strong positive correlations between WMC and G*f* (see Conway & Kovacs, 2013, for a review). In the first large-scale study of the relationship between WMC and G*f*, Kyllonen and Christal (1990) found strong correlations (> 0.90) between latent variables representing reasoning ability and WMC and concluded that "reasoning ability is little more than working memory capacity" (p. 389). Even more boldly, Kyllonen (2002) stated, "we have our answer to the question of what g is. It is working memory capacity" (p. 433). In a similar vein, Engle (2002) noted that WMC "is at least related to, maybe isomorphic to, general fluid intelligence" (p. 22). At least in the minds of some, the question of what intelligence is, beyond the variance common to a collection of mental ability measures, seemed to be settled.

However, later research revealed that the relationship between WMC and G*f* was weaker than initially thought – much weaker. In a meta-analysis, Ackerman, Beier, and Boyle (2005) found a correlation of 0.50 between latent variables representing *g* and WMC, indicating that only a quarter of the variance was shared between the factors, and in a reanalysis of twelve studies, Kane, Hambrick, and Conway (2005) found an average correlation of 0.72 between latent variables representing G*f* and WMC. Furthermore, analyzing item-level data, Unsworth and Engle (2005) found that the correlation between WMC and performance on Raven's Progressive Matrices was as strong for simple items as for complex ones. Likewise, Salthouse and Pink (2008) found that performance on small set sizes in complex working memory span tasks (e.g., operation span) correlated as highly with G*f* as performance on larger set sizes. These findings were problematic for any explanation of the relationship between performance on working memory tasks and G*f* in terms of the amount of information that can be simultaneously stored and processed.

As a theoretical critique, others pointed out that measures of WMC are at least as complex as the measures of intelligence that they purportedly explained (Deary, 2000). The point was that, like any other cognitive measure, no measure of WMC is "process pure" in the sense that it captures only the intended construct. As Salthouse and Pink (2008) commented, "Because some of the WM assessments closely resemble tests of reasoning and higher order cognition, it may not be reasonable to claim that the WM construct is theoretically more tractable or less opaque than are intelligence constructs, given the fact that it is operationalized in so many different ways that appear to have little conceptual integration" (p. 364). Ackerman, Beier, and Boyle (2002) described working memory as a "promiscuous" variable: one that correlates with everything.

Thus, enthusiasm about the hypothesis that G*f* is "little more than" WMC has waned. What may be regarded as the death knell for this view was recently sounded by one of its most prominent early supporters. As already mentioned, Engle (2002) argued that WMC tasks measure a factor that may be isomorphic to G*f*. However, in an update of this article, Engle (2018) explained, "One of the things I argued in the 2002 article was that individual differences in WMC possibly play a causal role in fluid intelligence. I based that argument on the strong relationship, on the order of .6 to .8, between WMC and fluid intelligence at the construct level. I now think that argument was wrong" (p. 192). Reviewing recent work from his lab, he explained that effects of WMC and G*f* on various outcomes are dissociable, noting that some effects that he and his colleagues originally attributed to WMC can actually be explained by G*f*. For example, Rosen and Engle (1998) found that high-span subjects produced more unique animal names in a fluency task than did low-span subjects, and argued from this finding that WMC is related to the ability to suppress intrusive thoughts (i.e., animal names that had already been produced). However, subsequent research using a more powerful statistical approach (structural equation modeling with large samples) showed that effects of WMC on fluency were entirely due to G*f* (Shipstead, Harrison, & Engle, 2016).

The Role of Placekeeping Ability

If WMC does not explain individual differences in G*f*, what does? In our own attempt to shed some light on this question, we have considered the role of a theoretical construct that we call *placekeeping*: the ability to perform a sequence of steps in a particular order, without skipping or repeating steps (Hambrick, Altmann, & Burgoyne, 2018). One reason to think that placekeeping ability is related to G*f* is that solving complex problems, at least of the well-structured variety, depends on a kind of linear thinking. Newell and Simon (1972; Newell, 1990) characterized problem-solving in terms of a search process in which the problem-solver applies sequences of operators to transform mental problem states, and periodically sets, suspends, and resumes goals organized in a hierarchical mental structure. For such processing to lead to solutions efficiently, the system must be able to keep its place in sequences of operators and within hierarchical goal structures. Skipping an element could mean missing a path to a solution, and repeatedly evaluating a failed path is inefficient and could also lead to missing a solution if solution time is limited. Consistent with this analysis, using a computational cognitive model, Carpenter, Just, and Shell (1990) demonstrated that successful goal management during problem-solving was associated with better performance on Raven's Progressive Matrices.

In a recent series of studies, we have tested the relationship between intelligence and placekeeping ability. The placekeeping task we developed is defined by the acronym UNRAVEL. Each letter in the acronym identifies a step in a looping procedure, and the letter sequence stipulates the order in which the steps are to be performed. That is, the U step is performed first, the N step second, the R step third, and so forth, and the participant returns to the U step following the L step. The participant is interrupted at random points and must perform a transcription task, before resuming UNRAVEL with the next step. A sample stimulus is shown in Figure 23.3. Each sample stimulus includes two characters; of these characters, one is a letter and one a digit, one is presented either underlined or italicized, one is colored red or yellow, and one is located outside of a gray box. Each step requires a two-alternative forced choice (i.e., a keypress) related to one feature of the stimulus, and the letter of the step mnemonically relates to the choice rule: The U step involves deciding whether the formatted character is underlined or italicized, the N step whether the letter is near to or far from the start of the alphabet, the R step whether the colored character is red or yellow, the A step whether the character outside the box is above or below, the V step whether the letter is a vowel or a consonant, the E step whether the digit is even or odd, and the L step whether the digit is less or greater than five. The task tests placekeeping because the stimulus provides no information about what step to perform next. Instead, the participant must remember their place in the procedure, which is especially challenging immediately after an interruption by the transcription typing task.

(a) Sample stimuli for UNRAVEL task:

(b) Choice rules and candidate responses for UNRAVEL task, and responses to the stimuli in (a):

Step	Candidate responses		Choice rules	Responses to sample stimuli: Stimulus 1	Responses to sample stimuli: Stimulus 2
U	u	i	character is Underlined or in Italics	u	i
N	n	f	letter is Near to or Far from start of alphabet	f	n
R	r	y	character is Red or Yellow	r	y
A	a	b	character is Above or Below the box	a	b
V	v	c	letter is Vowel or Consonant	c	v
E	e	o	digit is Even or Odd	o	e
L	l	m	digit is Less than or More than 5	l	m

(c) Sample stimulus for interrupting task:

Figure 23.3 *Sample stimuli, rules, and interrupting task from UNRAVEL task (from Hambrick & Altmann, 2015). Reprinted with permission of Springer.*

In recent work, we have examined the contribution of placekeeping to individual differences in G*f*. In a study using 132 undergraduate students (Hambrick & Altmann, 2015), we found that placekeeping ability (as indexed by UNRAVEL error rate) accounted for 20 percent of the variance in scores on Raven's Progressive Matrices. Furthermore, this contribution was reduced only slightly (20% to 18%) after controlling for WMC and measures of two other "executive functioning" factors (task switching and multitasking). Furthermore, in a structural equation model, latent variables representing placekeeping ability (again reflecting the UNRAVEL error rate) and Raven's performance correlated strongly ($r = -0.69$), in the predicted direction of higher Raven's score for participants with lower error rate in UNRAVEL.

In a more recent study (Burgoyne, Hambrick, & Altmann, in press), we had participants complete two tests of placekeeping, two tests of G*f*, and two tests of WMC. In regression analyses, placekeeping ability factors (error rate and response time) accounted for 12 percent of the variance in G*f* above and beyond WMC. By contrast, WMC accounted for only 2 percent of the variance in G*f* above and beyond these placekeeping factors. In a structural equation model, the placekeeping factors had stronger effects on G*f* than did WMC, and dropping the latter, placekeeping ability accounted for 70 percent of the variance. Because it is correlational, this finding does not establish that placekeeping is a cause of variation in Raven's scores (or G*f*). It is, however, consistent with that possibility, providing a motivation to conduct experiments to test causal hypotheses.

We make no claim that UNRAVEL is a "magic bullet" task for research on the underpinnings of intelligence, or that placekeeping ability is the only factor underlying individual differences in G*f*. We do not believe that there is any single construct that can explain something as complex as intelligence. No less so in intelligence research than any other area of psychological research, single-variable explanations are nearly always wrong, or at least incomplete, even if they are seductive. However, we do think that placekeeping ability may be one piece of the intelligence puzzle. Considering intelligence from a problem-solving perspective, a goal for future research in our labs is to conduct experiments to test hypotheses about how placekeeping constrains problem-solving behavior in intelligence tests such as Raven's Progressive Matrices. This research promises to shed light on mechanisms underlying human intelligence, particularly as manifested in the type of linear thinking that is involved in solving well-structured problems.

What Does Intelligence as Problem-Solving Ability Predict?

Life is an unceasing series of problems, from meeting the most basic requirements for survival, to finding and keeping employment, to fulfilling loftier aims such as making the world better for future generations. Even the task of remaining among the living (as it were) can be seen as a problem, especially when there are many ways to lose that status (getting hit by a truck, accidentally ingesting poison, wrecking one's health through a profligate lifestyle). Is intelligence predictive of success in life's problem-solving tasks? The answer is yes, even if the reasons are not yet fully understood. Next, we consider evidence for the relationship between intelligence and three outcomes that can be considered indexes of real-life problem-solving skill: mortality, job performance, and academic achievement.

Mortality

Intelligent people tend to live longer than less intelligent people. This relationship between intelligence and mortality has been documented in research by Deary and colleagues using data from the Scottish Mental Surveys. In 1932, the Scottish government administered an intelligence test to nearly all eleven-year-old children attending school on a single day. More than sixty years later, after finding the raw data collecting dust in a University of Edinburgh building, Deary and Whalley (2001) identified who from the cohort living in the city of Aberdeen was still alive, at age seventy-six. The results revealed that a 1 standard deviation advantage (15 IQ points) was associated with a 21 percent greater chance of survival. For example, a person with an IQ of 100 (the average for the general population) was 21 percent more likely to be alive at age seventy-six than a person with an IQ of 85. IQ also predicts morbidity – becoming ill due to diseases such as cancer and heart disease (Gottfredson & Deary, 2004).

In short, as has now been replicated in upward of twenty longitudinal studies from around the world, more intelligent people live longer, healthier lives than less intelligent people. One possible explanation for this finding is that intelligence is confounded with another variable that correlates with mortality: socioeconomic status. That is, intelligence and mortality may correlate because wealthier people have the means to develop their intelligence (through education) and stay in good health (through health care). There is some evidence consistent with this third-variable explanation. At the same time, socioeconomic factors do not appear to completely account for the intelligence-mortality/morbidity relationship. For example, Hart and colleagues (2003) linked IQ scores for over 900 of the participants from the 1932 study to those participants' responses on a national health survey conducted in the early 1970s. They found that statistically controlling for economic class and a measure of "deprivation" reflecting unemployment, overcrowding, and other adverse living conditions accounted for only about 30 percent of the IQ-mortality correlation.

Another possible explanation for the intelligence-mortality correlation might be called the *problem-solving hypothesis*: Day in, day out, more intelligent people are more effective in solving life's problems than less intelligent people, using knowledge that they have acquired from various sources (e.g., reading, listening to their physician) to keep in good health and stay alive. Consistent with this hypothesis, in the Scottish data, there was no relationship between IQ and smoking behavior in the 1930s and 1940s, when the health risks of smoking were unknown, but after that, people with higher IQs were more likely to quit smoking (Gottfredson & Deary, 2004). There is other evidence consistent with this hypothesis, as well. For example, a meta-analysis showed that people high in cognitive ability were less likely to be involved in vehicular accidents than people lower in cognitive ability (Arthur, Barret, & Alexander, 1991). Other research implicates lower levels of *executive functioning* (EF) in risky driving and vehicular accidents (Walshe et al., 2017). An umbrella term for cognitive operations (e.g., inhibition, updating, set shifting) that are presumed to underpin goal-directed behavior, EF correlates very highly with G*f* (McCabe et al., 2010). Lower cognitive ability has been found to be associated with poor supervisor ratings of employee safety behaviors, ranging from work-related accidents to distractibility when performing dangerous tasks (Postlethwaite et al., 2009).

Job Performance

As industrial-organizational psychologists have established, *g* is also the best-known predictor of acquisition of job knowledge during training and subsequent job performance (Schmidt & Hunter, 2004). Validity coefficients for *g* tend to be higher for more complex jobs than less complex jobs, but are nearly always positive and both statistically and practically significant. This is not to say that *g* is a perfect predictor of job performance – far from it. Across a wide range of occupations, the average validity coefficient for *g* is around 0.50 (see Schmidt & Hunter, 1998, 2004). This means that, on average, *g* accounts for about 25 percent of the variance in job performance, leaving the rest unexplained and potentially explainable by other factors. It is, however, to say that *g* is a better predictor of job performance than measures of other factors, such as personality, interests, and job interview ratings. It is also to say that intelligence tests are practically useful. As utility analyses demonstrate, use of tests with even modest validity (e.g., correlations with job performance in the 0.20–0.40 range) for personnel selection can substantially improve prediction of job performance, translating into substantial savings in terms of decreased job training time and increased productivity (Hunter & Schmidt, 1996).

It is sometimes argued that *g* (general cognitive ability) predicts job performance only early in training within a domain (e.g., a job), after which domain-specific factors (knowledge, skills, and strategies) enable a person to circumvent reliance on domain-general abilities. The chief proponent of this *circumvention-of-limits hypothesis* is the expertise researcher K. Anders Ericsson. In *Peak: Secrets from the New Science of Expertise,* Ericsson and Pool (2016) claimed, “While people with certain innate characteristics . . . may have an advantage when first learning a skill, that advantage gets smaller over time, and eventually the amount and quality of practice take on a much larger role in determining how skilled a person becomes” (p. 233). Similarly, citing his own review of the evidence (Ericsson, 2014), Ericsson recently wrote that “the influence of general abilities, such as IQ, is greater on performance of beginners but virtually disappears for individual differences among expert performers” (Ericsson, 2018a, p. 708), and that “traditional tests of intelligence and IQ are not predictive of individual differences in attained performance among skilled performers” (Ericsson, 2018b, p. 97).

In reality, evidence for this claim is weak, even though it might seem extensive. There are several major problems with Ericsson’s reviews of the relevant evidence (Ericsson, 2013, 2014, 2018a, 2018b; Ericsson & Moxley, 2013). First, in some cases, Ericsson makes inferences that are not licensed by the usual conventions of statistical inference in psychological research. In particular, for some studies, he cites as support for his view the finding that an ability-performance correlation was statistically significant in a lower-skill group but not in a higher-skill group, even though the correlations are not significantly different from each other. For example, in multiple reviews, as support for his view, he cites Ruthsatz and colleagues’ (2008) finding that a measure of intelligence (Raven’s score) correlated with musical performance in high school band members ($r = 0.25$, $p < 0.01$, $N = 178$) but not in more skilled groups of university music majors ($r = 0.24$, $p > 0.05$, $N = 19$) and music institute students ($r = 0.12$, $p > 0.05$, $N = 64$). However, these correlations are not significantly different from each other (all tests of differences in *r*s are non-significant, *z*s < 1). Thus, following the norm of testing for differences in correlations, Ruthsatz and colleagues' results actually *fail* to support Ericsson's claim of a diminishing ability-performance correlation with increasing skill.

The problem here can be made obvious by imagining a situation in which the correlation between two variables (e.g., IQ and performance) just exceeds the threshold for statistical significance in one group and just misses that threshold in another group – for example, with a sample size of fifty per group, $r = 0.280$ ($p = 0.049$) in one group and $r = 0.278$ ($p = 0.051$) in another group. Obviously, no psychological theory can purport to predict such a small difference between correlations (and thus whether one correlation will be significant and the other nonsignificant). Rather, in testing a theory that predicts differential relationships, as Ericsson's does, the question that must be answered is whether the correlations differ significantly from each other. Statistical tests have been developed for the very purpose of answering this question, and have long been in use in the behavioral sciences (e.g., the z test for difference between correlations from independent samples, moderated regression analysis testing for Group × Predictor interactions; see Cohen, 1988; Cohen et al., 2003; Hays, 1988).

In other cases, Ericsson's (2014) interpretations of evidence do not stand to reason. For example, he cites a report in the German magazine *Der Spiegel* that former chess world champion Garry Kasparov's IQ was estimated at 120 based on his score on Raven's Progressive Matrices, and notes that this score is "very close to the average of all chess players . . . thus not very predictive of world-class chess performance" (Ericsson, 2014, p. 87). However, one case does not a correlation make: Even if this estimate of Kasparov's IQ is accepted as valid, one cannot make an inference about the strength of a predictive relationship between two variables (i.e., a correlation) based on a single case. To wit, if other world-class chess players (e.g., Bobby Fischer, Magnus Carlsen) had considerably higher IQs, then IQ could still be *highly* predictive of world-class chess performance.

In still other cases, Ericsson is selective in reporting of evidence. For example, citing Schmidt and Hunter's (2004) review of the job performance literature, Ericsson (2014) explains, "The expert-performance approach proposes that performance on tests of general cognitive ability [and performance] will be correlated for beginners" (p. 84). However, he fails to mention that, in this same review, Schmidt and Hunter (2004) noted that this is true for non-beginners, as well: "One might hypothesize that the validity of GMA [general mental ability] declines over time as workers obtain more job experience. However, research does not support this hypothesis" (p. 167). Ericsson also seems to overlook relevant evidence. For example, in the surgical domain, he misses Gallagher and colleagues' (2003) finding that scores on a test of visuospatial ability correlated significantly and similarly with performance on a laparoscopic laboratory cutting task in novices (rs = 0.50 and 0.50, Ns = 48 and 32) and in experienced surgeons ($r = 0.54$, $N = 18$). No review is perfect – it is easy to miss studies. All the same, failing to review evidence contrary to a hypothesis leads to a biased portrayal of the strength of the evidence for that hypothesis.

Finally, Ericsson makes material errors in his reviews. For example, Ericsson (2013) claimed that "Kopiez and Lee (2006) found that for musicians with lower sight-reading skill there was a correlation with their working memory. For musicians with a higher level of sight-reading skill there was no significant relation between their performance and their working memory" (p. 236). This is an error of commission: Kopiez and Lee (2006) reported no such finding. In fact, they did not report any analyses comparing the correlation between working memory and sight-reading performance in groups representing lower versus higher levels of sight-reading skill. As another example, referring to a subsequent report of data from this study of sight-reading, Ericsson (2018a) noted that "Kopiez and Lee (2008) found that speed of alternating finger movements [music-specific speed trilling] and amount of accumulated sight-reading experience were the only significant predictors of sight-reading performance, but not working memory" (p. 707). This is an error of omission: Sight-reading performance was also significantly predicted by a nonmusic measure of information-processing speed (i.e., number combination, $r = -0.44$, $p = 0.001$; see Kopiez & Lee, 2008, table 2), indicating faster processing for more skilled sight-readers. These errors are not inconsequential typos; they lead to conclusions that contradict those Ericsson advances (for further examples of errors in Ericsson's writings, see Hambrick et al., 2014; Macnamara, Hambrick, & Moreau, 2016).

We carried out our own review of evidence relevant to the circumvention-of-limits hypothesis (Hambrick, Burgoyne, & Oswald, in press), conducting systematic searches for relevant articles in the literature on expertise in five domains (games, music, science, sports, surgery/medicine, and aviation), as well as the literature on job performance. Altogether, we searched approximately 1,300 documents. The findings can be summarized briefly. On balance, evidence from the expertise literature does not support the circumvention-of-limits hypothesis. To be exact, three of fifteen studies provide support for the hypothesis, either in the form of significantly different ability-performance correlations across skill groups or significant ability × skill interactions on performance. What might be regarded as the strongest evidence comes from one of our own meta-analyses (Burgoyne et al., 2016). We found that the correlation between G*f* (as measured by tests of reasoning ability) and chess expertise was significantly higher for less-skilled chess players than for more-skilled players. However, as we urged, this finding must be interpreted cautiously, because the measure of chess skill was highly confounded with age (i.e., the more-skilled players were adults, the less-skilled players were children).

A more consistent picture emerged in the review of evidence from the job performance literature. Ability-performance correlations may decrease in relatively simple lab tasks, in particular those with consistent demands (Ackerman, 1988; see also Henry & Hulin, 1987). However, even after an extensive amount of job experience, general cognitive ability remains a statistically and practically significant predictor of actual job performance (see also Reeve & Bonaccio, 2011). Some of the most compelling evidence for this conclusion comes from a reanalysis of data from the Joint-Service Job Performance Measurement/Enlistment (JPM) Standards Project, a large study initiated in 1980 by the US Department of Defense to develop measures of military job performance (see Hambrick et al., in press; see Wigdor & Green, 1991, for further description of this project). The JPM data set includes thirty-one jobs and a total sample size of 10,088 military personnel; the measure of general cognitive ability was the

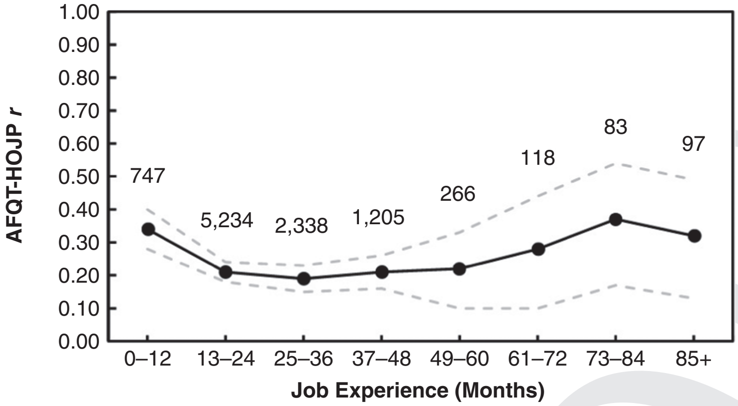

Figure 23.4 *Correlation between intelligence test score (AFQT) and hands-on job performance (HOJP) at different levels of job experience from JPM Standard Project. Total* N = *10,088 (from Hambrick, Burgoyne, & Oswald, in press). Reprinted with permission of Oxford University Press.*

Armed Forces Qualifying Test (AFQT) score, and job performance was measured with hands-on job performance (HOJP) tests for the different jobs. As shown in Figure 23.4, the AFQT-HOJP correlation decreases from the first year to the second, stabilizes, and then, if anything, increases. The overall picture to emerge from this and other large-scale studies (e.g., Schmidt et al., 1988; Farrell & McDaniel, 2001) is that general cognitive ability remains a significant predictor of job performance, even after extensive job experience, and even if validity drops initially.

Academic Achievement

Intelligence is also a strong predictor of academic achievement, which can be considered an index of problem-solving ability, at least within the confines of the classroom. (There are no doubt people who excel in work and other realms of life despite less-than-stellar academic records.) Using a sample of over 70,000 English schoolchildren, Deary and colleagues estimated the relationship between IQ at age eleven and scores on national examinations in twenty-five topics at age sixteen (Deary et al., 2007). The correlation between latent variables representing general intelligence and educational achievement was 0.81. Intelligence predicts later academic performance, as well. Most notably, scores on college admissions exams (the SAT and ACT) – which correlate highly with independent assessments of intelligence (Koenig, Frey, & Detterman, 2008) and were described by Gardner (1999) as “thinly disguised intelligence tests” (p. 69) – predict not only first-year grade point average (GPA), but overall GPA (Kuncel, Hezlett, & Ones, 2004). Summarizing evidence from their own and others’ research in the *Wall Street Journal*, Kuncel and Sackett (2018) wrote “Standardized tests are just tools – very effective tools – but they provide invaluable information to admissions offices. They identify those students who need help catching up with fundamental skills and those who are ready to tackle advanced material and rapidly accelerate in their learning."

As evidence from the landmark Study of Mathematically Precocious Youth (SMPY) indicates, scores on college admissions tests also predict outcomes reflecting accomplishments beyond the college years. As part of a planned fifty-year study, the SAT was administered to intellectually gifted youth by age thirteen, and the roughly 2,300 scoring in the top 1 percent were tracked into adulthood. (Given that the SAT was administered at this unusually young age, ceiling effects were avoided, and there was a wide range of scores even in the top 1 percent, for example, 390 to 800 for SAT Math.) Lubinski (2009) found that higher overall SAT score (an index of *g*) was associated with higher levels of accomplishment. For example, compared to individuals in the 99.1 percentile, those in the 99.9 percentile (the profoundly gifted) were 3.56 times more likely to have earned a doctorate, 4.97 times more likely to have published in a STEM journal, 3.01 times more likely to have been awarded a patent, and 2.31 times more likely to have an income at or above the 95th percentile (Ferriman Robertson et al., 2010).

One possible explanation for such results is that people who score well on standardized tests are admitted to better colleges and universities, which creates a halo effect that makes it easier for these people to get into top graduate programs (e.g., Princeton or Harvard), which in turn creates a halo effect that makes it easier for them to, say, publish in a STEM journal. However, this does not appear to explain the relationship between SAT scores and later accomplishments in the SMPY. For example, Park, Lubinski, and Benbow (2008) found that, among individuals who had earned a doctorate, prestige of doctoral institution had very little impact on the relationship between SAT math scores and publication in STEM journals: The odds ratio was 4.04 for graduates of a non-top-fifteen institution ($N = 766$) and 3.52 for graduates of a top-fifteen institution ($N = 240$). Other mediating variables may turn out to explain the relationship between SAT and outcomes, but the simplest explanation remains the most direct, which is that ability influences outcomes.

Can Intelligence as Problem-Solving Ability be Improved through Training?

If intelligence is important for real-world problem-solving, then can it be increased through training, leading to improvements in people's lives? Psychologists have been interested in this possibility for as long as they have studied intelligence. Over a century ago, Alfred Binet, developer of the first standardized intelligence test, envisioned a system of "mental orthopedics" for increasing intelligence. "With practice, training, and above all, method," Binet wrote, "we manage to increase our attention, our memory, our judgement and literally to become more intelligent than we were before" (Binet, 1909/1975, p. 106–107).

Nevertheless, efforts to increase intelligence through training have met with little success. Beginning in the 1970s, a number of longitudinal studies were launched to examine the effects of intensive educational interventions on intelligence. In one of the best-known studies, the Abecedarian Early Intervention Project (Campbell et al., 2001), children from low-income families received an intensive educational intervention from infancy to age five that included educational games designed to enhance cognitive functioning, while children assigned to a control group received social services, healthcare, and nutritional supplements. Increases in intelligence were modest: At the end of the study, children in the treatment group showed an advantage of 6 IQ points (about 0.40 standard deviations) over a control group.

There was renewed excitement about the idea of training intelligence in the 2000s. In an article published in the *Proceedings of the National Academy of Sciences,* Jaeggi and colleagues reported large gains in G*f* for a sample of young adults following computerized cognitive training (Jaeggi et al., 2008). After completing a pretest of reasoning ability, participants were assigned to either a control group, or to a training group in which they received eight, twelve, seventeen, or nineteen sessions of training in a "dual n-back" task requiring simultaneous monitoring of two streams of information (one auditory and one visual). Finally, at posttest, all participants took a different version of the reasoning test. Jaeggi and colleagues reported that there was a greater gain in G*f* from pretest to posttest for the training groups than for the control group, as well as a dosage-dependent relationship (more training, greater gain in G*f*). They concluded that the "finding that cognitive training can improve *Gf* is a landmark result because this form of intelligence has been claimed to be largely immutable" (p. 6832).

The study made an immediate impact, both in the popular press and the scientific literature. *Discover* magazine called Jaeggi and colleagues' (2008) findings one of the top 100 scientific discoveries of 2008, and Sternberg (2008) commented that the study seemed "to resolve the debate over whether fluid intelligence is, in at least some meaningful measure, trainable" (p. 6791). Within a year of publication, the study had already been cited nearly 100 times. The study was also touted as evidence for the effectiveness of brain training by the company Lumosity, which introduced a gamified version of the dual n-back task. Within a few years, advertising on National Public Radio and network television, Lumosity boasted fifty million subscribers. The claim was that brain training can improve cognitive functioning, with benefits for real-world performance.

However, as subsequently reported by Redick and colleagues (2013), Jaeggi and colleagues' (2008) study had major flaws. The control group was "passive" or "no-contact," meaning that the control subjects had no contact with experimenters between pretest and posttest. Therefore, it was possible that the training groups improved because they expected improvement and tried harder on the intelligence test at posttest. Furthermore, the study was not a single experiment in which participants were randomly assigned to conditions that differed only in amount of training; Instead, the results were from different experiments in which the procedures varied in other respects, making it different to interpret the reported results. For example, the reasoning test differed across the training groups, with the eight-session training group receiving an eighteen-item version of Raven's, but the other groups receiving the twenty-nine-item Bochumer Matrizentest (BOMAT) test, and with a twenty-minute time limit for the nineteen-session group and a ten-minute time limit for the twelve- and seventeen-session groups. Finally, there was no, or even negative, transfer for other measures not included in the report of the study (e.g., for the 19-session group, no transfer to visuospatial span, and negative transfer to digit-symbol substitution).

What's more, the magnitude of the reported training gain in G*f* seemed larger than possible. The training groups received an average of six hours of dual n-back training, and the difference in the gain in reasoning performance from pretest to posttest for the training groups compared to the control group was 0.40 standard deviations. IQ tests typically have a mean of 100 with a standard deviation of 15. Thus, in terms of IQ points, Jaeggi and colleagues' (2008) results implied that the G*f* of subjects in the training groups increased an average of six points in six hours – a point an hour. This was roughly the same gain in IQ for children in the treatment group of the Abecedarian study after five years of intensive intervention.

Given these problems, Redick and colleagues (2013) attempted to replicate Jaeggi and colleagues' (2008) findings. Participants completed seventeen different cognitive ability tests, including eight tests of G*f*. They were then assigned to a treatment group in which they practiced the dual n-back task for twenty sessions, to a placebo control group in which they practiced another cognitive task for twenty sessions, or to a no-contact control group. Then, at posttest, all participants completed different versions of the cognitive ability tests. The results revealed that the dual n-back group was no higher in G*f* than the control groups. Other replication failures followed (e.g., Chooi & Thompson, 2012; Harrison et al., 2013). Moreover, meta-analyses demonstrated that benefits of brain training are limited to the trained task or to similar tasks, indicating near transfer but no far transfer (Melby-Lervåg & Hulme, 2013).

Subsequently, a letter published by the Stanford Center on Longevity (2014) and signed by seventy-five cognitive psychologists and neuroscientists cautioned, "The strong consensus of this group is that the scientific literature does not support claims that the use of software-based 'brain games' alters neural functioning in ways that improve general cognitive performance in everyday life, or prevent cognitive slowing and brain disease." A rebuttal letter (Cognitive Training Data, 2014), signed by over 100 scientists and practitioners – some of whom acknowledged financial interests in the brain training industry, notably including the cofounder of Posit Science, Michael Merzenich – countered that "a substantial and growing body of evidence shows that certain cognitive training regimens can significantly improve cognitive function, including in ways that generalize to everyday life."

In the wake of the controversy, Simons and colleagues (2016) conducted an exhaustive review of evidence used by brain training companies to promote their products, and concluded, "Brain training is appealing in part because it seems to provide a quick way to enhance cognition relative to the sustained investment required by education and skill acquisition. Practicing a cognitive task consistently improves performance on that task and closely related tasks, but the available evidence that such training generalizes to other tasks or to real-world performance is not compelling" (p. 173). Reaching the same conclusion, in 2016, the United States Federal Trade Commission (FTC) fined Lumosity $50 million for making unfounded claims about the real-world benefits of Lumosity games (the fine was reduced to $2 million because of financial hardship). Speaking to *NBC Nightly News*, a staff lawyer for the FTC commented, “There just isn’t evidence that any of that [using Lumosity] will translate into any benefits in a real-world setting” (*NBC Nightly News*, 2016; see also Federal Trade Commission, 2016). The most recent meta-analysis on brain training corroborates this decision: There is no convincing evidence for far transfer of working memory training to real-world outcomes (Melby-Lervåg, Redick, & Hulme, 2016).

Obviously, through training, people can substantially improve their performance in complex problem-solving tasks, whether it be a game such as chess, an occupational task such as air traffic control, or an everyday task such as managing finances. People can, and do, develop high levels of skill in complex tasks through prolonged training. However, beyond improvements in the trained task or similar tasks, benefits of brain training appear to be nil.

Conclusions

Intelligence can be viewed as a general ability to solve problems. Consistent with this view, intelligence and complex problem-solving correlate very highly. Moreover, intelligence is a statistically and practically significant predictor of real-world outcomes that reflect problem-solving skill, including job performance and academic achievement. Intelligence even predicts success in the ultimate problem-solving task: staying alive. The question of what underlies intelligence still remains unanswered. The once popular view that intelligence (specifically G*f*) is isomorphic with working memory capacity has fallen out of favor. Meanwhile, we have discovered that G*f* correlates highly with placekeeping – the ability to perform a sequence of operations in a particular order. We are optimistic that this and other research aimed at identifying mechanisms underlying intelligence will provide the scientific foundation for a wide range of practical applications, from improving procedures for training people in complex tasks, to devising interventions for enhancing problem-solving skill, to refining measures used to predict people’s performance in settings such as the workplace.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288–318.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*, 227–257.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*, 567–589.

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

Anderson, J. R. (1985). *Cognitive psychology and its implications*. New York: W H Freeman/ Times Books/ Henry Holt & Co.

Arthur, W., Barret, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance*, *4*, 89–105.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation*, vol. 8 (pp. 47–89). New York: Academic Press.

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*, 611–626.

Binet, A. (1909/1975). *Modern ideas about children* (trans. S. Heisler). California: University of California.

Bingham, W. V. (1937). *Aptitudes and aptitude testing*. Oxford: Harpers.

Brown, R. E. (2016). Hebb and Cattell: The genesis of the theory of fluid and crystallized intelligence. *Frontiers in Human Neuroscience*, *10*(606). https://doi.org/10.3389/fnhum.2016.00606

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (in press). Placekeeping ability as a component of fluid intelligence: Not just working memory capacity. *American Journal of Psychology*.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*, *59*, 72–83.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology*, *37*, 231–242.

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

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

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*, 153–193.

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

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*, 531–542.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/ correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cognitive Training Data (2014). Cognitive training data response letter. www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/response-letter

Conway, A. R. A., & Kovacs, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. *Psychology of Learning and Motivation*, *58*, 233–270.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., et al. (2011). Measuring performance in dynamic decision making. *Journal of Individual Differences*, *32*, 225–233.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. Oxford: Oxford University Press.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

Deary, I. J., & Whalley, L. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal*, *322*, 819–822.

Duncker, K. (1945). On problem-solving (trans. L. Lees). *Psychological Monographs*, *58*, i–113.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

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

Ericsson, K. A. (2013). My exploration of Gagné's "evidence" for innate talent. It is Gagné who is omitting troublesome information so as to present more convincing accusations. In S. B. Kaufman (Ed.), *The complexity of greatness: Beyond talent or practice* (pp. 223–256). New York: Oxford University Press.

Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence*, *45*, 81–103.

Ericsson, K. A. (2018a). Superior working memory in experts. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 696–713). New York: Cambridge University Press.

Ericsson, K. A. (2018b). Intelligence as domain-specific reproducible performance: The role of acquired domain-specific mechanisms in expert performance. In R. J. Sternberg (Ed.), *The nature of human intelligence* (pp. 85–100). New York: Cambridge University Press.

Ericsson, K. A., & Moxley, J. (2013). Working memory that mediates experts' performance: Why it is qualitatively different from traditional working memory. In T. P. Alloway & R. Alloway (Eds.), *Working memory: The connected intelligence* (pp. 109–136). New York: Psychology Press.

Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. Boston: Houghton Mifflin Harcourt.

Farrell, J. N., & McDaniel, M. A. (2001). The stability of validity coefficients over time: Ackerman's (1988) model and the General Aptitude Test Battery. *Journal of Applied Psychology*, *86*, 60–79.

Federal Trade Commission (2016). Lumosity to pay $2 million to settle FTC deceptive advertising charges for its "brain training" program. www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-ftc-deceptive-advertising-charges

Ferriman Robertson, K., Smeets, S., Lubinski, D., & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science*, *19*, 346–351.

Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 3–26). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gallagher, A. G., Cowie, R., Crothers, I., Jordan-Black, J. A., & Satava, R. M. (2003). PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surgical Endoscopy and Other Interventional Techniques*, *17*, 1468–1471.

Gardner, H. (1999). Who owns intelligence? *The Atlantic Monthly*, 283, 67–76.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*, 1–4.

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013a). A multitrait-multimethod study of assessment instruments for complex problem solving. *Intelligence*, *41*, 579–596.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, *36*, 189–213.

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013b). Complex problem solving in educational contexts – Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*, 364–379.

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin and Review*, *22*, 1104–1110.

Hambrick, D. Z., Altmann, E. M., & Burgoyne, A. P. (2018). A knowledge activation approach to testing the circumvention-of-limits hypothesis. *American Journal of Psychology*, *131*, 307–321.

Hambrick, D. Z., Altmann, E. M., Oswald, F. L., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Accounting for expert performance: The devil is in the details. *Intelligence*, *45*, 112–114.

Hambrick, D. Z., Burgoyne, A. P., & Oswald, F. L. (in press). Domain-general models of expertise: The role of cognitive ability. In P. Ward, J. M. Schraagen, J. Gore, & E. Roth (Eds.), *Oxford handbook of expertise: Research and application*. Oxford: Oxford University Press.

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*, 2409–2419.

Hart, C. L., Taylor, M. D., Smith, G. D., Whalley, L. J., Starr, J. M., Hole, D. J., et al. (2003). Childhood IQ, social class, deprivation, and their relationships with mortality and morbidity risk in later life: Prospective observational study linking the Scottish Mental Survey 1932 and the Midspan studies. *Psychosomatic Medicine*, *65*, 877–883.

Hays, W. L. (1988). *Statistics* (4th ed.). Fort Worth, TX: Holt, Rinehart & Winston Inc.

Hebb, D. O. (1942). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society*, *85*, 275–292.

Henry, R. A., & Hulin, C. L. (1987). Stability of skilled performance across time: Some generalizations and limitations on utilities. *Journal of Applied Psychology*, *72*, 457–462.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*, 447–472.

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.

Jensen, A. (1999). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johnson, J., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393–416.

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

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence*, *36*, 153–160.

Kopiez, R., & Lee, J. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, *8*, 97–120.

Kopiez, R., & Lee, J. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research*, *10*, 41–62.

Kretzschmar, A., Neubert, J. C., Wüstenberg, S., & Greiff, S. (2016). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*, *54*, 55–69.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.

Kuncel, N., & Sackett, P. (2018). The truth about the SAT and ACT. *Wall Street Journal*, March 8. www.wsj.com/articles/the-truth-about-the-sat-and-act-1520521861

Kunda, M., Soulières, I., Rozga, A., & Goel, A. K. (2016). Error patterns on the Raven's Standard Progressive Matrices Test. *Intelligence*, *59*, 181–198.

Kyllonen, P. C. (2002). g: Knowledge, speed, strategies, or working-memory capacity? A systems perspective. In R. J. Sternberg & E. L. Gigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 415–445). Mahwah, NJ: Erlbaum.

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

Lubinski, D. (2009). Exceptional cognitive ability: The phenotype. *Behavioral Genetics*, *39*, 350–358.

Macnamara, B. N., Hambrick, D. Z., & Moreau, D. (2016). How important is deliberate practice? Reply to Ericsson (2016). *Perspectives on Psychological Science*, *11*, 355–358.

Mayer, R. E. (2013). Problem solving. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 769–778). New York: Oxford University Press.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity, frontal lobe functioning, and general fluid intelligence: An adult lifespan study. *Neuropsychology*, *22*, 638–644.

McGrew, K. S. (2005). *The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future*. New York: Guilford Press.

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.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291.

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.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

NBC Nightly News (2016). Lumosity to pay $2M over allegations that its ads deceived customers. January 5. www.youtube.com/watch?v=uw67Vf_EwAw&gl=UG&hl=en-GB

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science*, *19* (10), 957–961.

Postlethwaite, B., Robbins, S., Rickerson, J., & McKinniss, T. (2009). The moderation of conscientiousness by cognitive ability when predicting workplace safety behavior. *Personality and Individual Differences*, *47*, 711–716.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., et al. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379.

Reeve, C. L., & Bonaccio, S. (2011). On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence*, *39*, 255–272.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language*, *39*, 418–436.

Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence*, *36*, 330–338.

Salthouse, T. A. (2000). Pressing issues in cognitive aging. In D. Park & N. Swartz (Eds.), *Cognitive aging: A primer* (pp. 43–54). New York: Psychology Press.

Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin and Review*, *15*, 364–371.

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.

Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*, 162–173.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology*, *73*, 46–57.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*, 771–799.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., et al. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*, 103–186.

Sonnleitner, P., Keller, U., Martin, R., & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, *41*, 289–305.

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

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence*, *53*, 92–101.

Stanford Center on Longevity (2014). A consensus on the brain training industry from the scientific community. http://longevity.stanford.edu/a-consensus-on-the-brain-training-industry-from-the-scientific-community-2

Stern, W. (1914). *The psychological methods of testing intelligence*. Baltimore, MD: Warwick & York.

Sternberg, R. J. (1985a). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology*, *49*, 607–627.

Sternberg, R. J. (1985b). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, UK: Cambridge University Press.

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences*, *105*, 6791–6792.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, *41*, 37–55.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and Raven. *Intelligence*, *33*, 67–81.

Vernon, P. E. (1965). Ability factors and environmental influences. *American Psychologist*, *20*, 723–733.

Walshe, E. A., McIntosh, C. W., Romer, R., & Winston, F. K. (2017). Executive function capacities, negative driving behavior and crashes in young adults. *International Journal of Environmental Research and Public Health*, 14, E1314.

Wigdor, A., & Green, B. F., Jr. (1991). *Performance assessment for the workplace*, 2 vols. Washington: National Academy Press.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving – More than reasoning? *Intelligence*, *40*, 1–14.