Working Memory and Intelligence:

An Overview

Andrew R. A. Conway, Brooke N. Macnamara

Princeton University

Pascale M. J. Engel de Abreu

University of Luxembourg

Author Note

Correspondence concerning this article should be addressed to Andrew R. A. Conway, Princeton University, Department of Psychology, Green Hall, Princeton, NJ, 08540, USA. E-mail: aconway@princeton.edu

The terms working memory and intelligence are generally used to refer to two very different psychological constructs yet measures of working memory capacity are strongly related to scores on most intelligence tests. Behavioral scientists don't normally get very excited about correlations because they are reminded more often than most that correlation does not imply causation. However, when it comes to the correlation between working memory capacity and intelligence, there is quite a bit of excitement. The reason is simple: If two psychological constructs are strongly correlated then in most applied settings either construct can be used to predict future behavior. Assessment of one construct is sufficient; assessment of both constructs is redundant. For example, either working memory capacity or intelligence can be used to predict academic achievement in children (Alloway & Alloway, 2010). With respect to the behavioral sciences, this is more than a convenience. It presents the opportunity for a revolution in the assessment of cognitive ability. A century has passed since the first intelligence tests were developed yet to this day no one can agree on what intelligence really means. In the 20th century the Intelligence Quotient (IQ) emerged as the default marker of one's intellectual ability, yet the design of IQ tests was at first subjective and later motivated by data, not theory. In contrast, working memory is a well defined construct and tasks designed to measure working memory capacity are motivated by psychological and biological theories, developed to account for what is known about the mind and brain, and more importantly, amenable to change as more knowledge is accumulated.

The goal of the current chapter is to review empirical evidence demonstrating a strong correlation between working memory capacity and intelligence. The chapter will also provide the reader with an introduction to *psychometrics*, which is a scientific approach to the measurement of psychological constructs that has largely dominated research on cognitive abilities for over 100 years, and informs current thinking on the relationship between working memory capacity, intelligence, and their many correlates.

The chapter begins with a brief and selective historical review of research on intelligence, includes descriptions of popular measures of intelligence, and introduces the notion of IQ. This is followed by, in similar fashion, a review of research on working memory and a discussion of measures of working memory capacity. These introductory sections set the stage for a review of empirical evidence illustrating strong correlations between working memory capacity and intelligence. Finally, to put this work into perspective, the chapter concludes with a critical analysis of the psychometric approach to the investigation of cognitive ability.

Intelligence

Intelligence is a construct used by scientists and laypeople alike to describe and predict individual differences in cognitive abilities. An individual is typically described as “highly intelligent” if he or she consistently performs well above average on various cognitive tasks, or tests,[1] for example, by achieving high marks in school or by scoring well on standardized tests or entrance exams. As mentioned, no one consensus definition of intelligence exists, so it’s hard to find one that will satisfy every reader. In his recent compendium, *IQ and Human Intelligence*, Nick Mackintosh (1998) successfully sidestepped this problem, so we will follow suit, and cite the *Oxford English Dictionary*, which defines an “intelligent” person as, “Having a high degree or full measure of understanding; quick to understand; knowing, sensible, sagacious.”

It’s fair to say that most contemporary scholars would accept this definition of “intelligent”, albeit to varying degrees. Indeed, it bears some resemblance to a definition of intelligence provided in a 1994 editorial statement to *The Wall Street Journal*, signed by 52 academic researchers: “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. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do.”

Where did this general notion of intelligence come from? Historical analyses of both literature and linguistics suggest that the basic idea that some people are “smarter” than others has been around for quite some time, dating back at least to early Greek philosophers (Sternberg, 1990). More contemporary definitions of intelligence stem from systematic research conducted in Psychology and Education in the late 19th and early 20th century. Two men in particular, Charles Spearman and Alfred Binet, had significant impact on theories of intelligence and IQ testing respectively. We start here with Spearman and then discus Binet in the following section on intelligence tests.

British psychologist Charles Spearman (1904) was the first to demonstrate an empirical phenomenon known as the "positive manifold", which refers to the commonly observed pattern of positive correlations among a wide variety of measures of cognitive performance. For example, high school grades, standardized test scores, and scores on working memory tasks all tend to be positively correlated, yet they are all very different assessment tools (for instance, some are computerized and some are "paper and pencil", some use verbal material and some use spatial material, etc.). Spearman demonstrated the positive manifold by calculating correlations among children's grades in six different disciplines: Classics, French, English, Math, Pitch, and Music. Correlations are typically presented in a matrix, so in Spearman's case it was a 6x6 matrix. Spearman observed that all the correlations in the matrix were positive and strong. He then conducted a statistical procedure called factor analysis and found that one single factor accounted for the majority of variance in all the measures. In other words, the six different measures could be reduced to one, and the variance among children could still, largely, be explained. This general factor is now known as Spearman's *g*.

Of course, debate still rages as to what *g* really means, if anything. Broadly speaking, three main camps exist: (1) it reflects a general cognitive ability; (2) it reflects the correlation among several different but related abilities; and (3) it is merely a statistical artifact. We will have more to say about these opposing views later in the section on psychometrics. Here, we will consider just a few of the most influential theories that have been proposed since Spearman's initial findings.

Spearman belonged to the first camp and argued that *g* corresponds to a single mental ability. He used energy and resources as metaphors to argue that some people simply have greater mental energy or more cognitive resources and will therefore perform above average on any cognitive task, regardless of domain. This theory has intuitive appeal and was supported by his factor analysis on children's grades in different classes. For instance, in school settings Mathematics and English are very different subjects, with a different curriculum and different instruction and testing formats, yet among the children that Spearman observed, those who did well in Mathematics also tended to do well in English. Similar patterns of correlations are still being observed and have serious implications in modern society. In the United States, for example, most young adults planning to attend college take the SAT test (formally known as Scholastic Aptitude Test), which is used by universities to make admissions decisions. In short, if one does well on the SAT then chances of being admitted to a prestigious university increase. Two subtests of the SAT are the Quantitative SAT (QSAT), which measures mathematical knowledge and reasoning, and the Verbal SAT (VSAT), which measures language-based knowledge and reasoning. The subtests, at face value, look very different from one another, and they are obviously designed to test different aptitudes. The implicit assumption is that the QSAT will have greater predictive validity with respect to the sciences, whereas the VSAT will have greater predictive validity with respect to the humanities. There is some support for this differential validity. However, the correlation between the QSAT and VSAT is strong, typically around $r = .70$ (e.g., Frey & Detterman, 2004). This strong positive correlation is consistent with Spearman's notion that there must be a general ability that accounts for consistent variation among such different tests. Moreover, the theory of a general ability is anecdotally supported by real-life experiences, which are admittedly subjective, yet seem consistent enough to support the notion that some people are simply smarter than others, in a very general sense.

Other researchers were quick to critique Spearman's theory of *g*. One early influential approach was the bi-factor theory of *g*, which is really just an extension of Spearman's theory (Holzinger & Swineford, 1937). According to bi-factor theory, *g* refers to a general mental ability that accounts for the majority of variation in test performance, but the total variance in any one test must be accounted for by two factors, one general, and the other specific to the particular test. In contemporary terms, the two factors in the original bi-factor model map onto domain-general sources of variance and test-specific sources of variance. In other words, individual differences in cognitive ability can be explained by cognitive and neural mechanisms that are general and exert their influence on all tests, yet some variance in individual tests is accounted for by mechanisms specific to the particular test. Formally incorporating a second factor to account for test-specific variance allowed the bi-factor model to account for patterns of correlations better than the one-factor general ability model (Holzinger & Swineford, 1937).

The second camp of theories argues that *g* reflects the correlation of several, separable primary mental abilities (e.g., Thurstone, 1938). In his primary mental abilities model, Louis Leon Thurstone argued that domain specific abilities were more integral to individual differences than a general factor. He identified several primary abilities, such as verbal comprehension, inductive reasoning, perceptual speed, numerical ability, verbal fluency, associative memory, and spatial visualization. The primary mental abilities model is supported by the fact that tasks designed to test the same mental ability tend to be correlated more strongly with each other than with tasks designed to measure a different ability, even though the positive manifold is still observed. Thus, a general factor still plays a role but Thurstone placed greater emphasis on more specific, primary abilities. (For an even more elaborate primary ability model, see the work of Guilford, 1988).

Thurstone and Guilford, and many others, proposed several different mental abilities that are defined by the content of test materials. For example, verbal ability is operationally defined by tests that primarily consist of verbal material. Another way to categorize tests of intelligence is to consider the extent to which a particular test requires previously learned information, or acquired knowledge, versus novel problem-solving techniques. For example, Cattel and Horn (1978) distinguished crystallized intelligence from fluid intelligence. Crystallized intelligence refers to the ability to access and use knowledge, experience, and skills stored in long-term memory and is measured by tests that are designed to assess a person's depth and breadth of knowledge of particular topics, such as general knowledge, vocabulary, and mathematics. In contrast, fluid intelligence refers to inductive and deductive reasoning in situations that don't allow for the use of prior experience but instead challenge the individual to adapt and develop novel ideas and strategies to succeed. The distinction between crystallized and fluid intelligence is supported by psychometric studies as well as neuropsychological and developmental investigations of cognitive ability. For example, certain types of brain damage or disease cause a deficit in fluid intelligence but not crystallized intelligence, and fluid intelligence declines in later years of life while crystallized intelligence accumulates over the lifetime (Cattell, 1987).

Thurstone, Guilford, Horn, Cattell, and many others all struggled with the problem of determining how many primary abilities exist, and how they should be categorized. The best one could do is to conduct an exhaustive analysis of all the correlational studies ever reported on cognitive ability and attempt to organize specific tasks by primary abilities in a manner that is consistent with all the empirical evidence. John Carroll (1993) accomplished this remarkable feat. His book, *Human cognitive abilities: A survey of factor-analytic studies*, is staggering in its scope. Based on 461 independent data sets on individual differences in cognitive abilities, Carroll proposed a three-stratum theory of cognitive ability. The three strata are "narrow", "broad", and "general". Narrow abilities are specific to particular tasks, broad abilities reveal their influence in broader classes of tasks, and the general factor permeates all cognitive tasks.

In contrast to Carroll and other psychometricians who preceded him, the third camp of theorists argue that *g* is merely a statistical artifact and has no direct cognitive or biological basis. Theorists in this camp are quick to point to a fundamental error in the interpretation of factors in the statistical procedure called factor analysis. Most behavioral scientists interpret factors from a factor analysis as reflecting a unitary construct that can be linked to specific cognitive and neural mechanisms. However, early critics of Spearman's work demonstrated that a general factor could emerge from a factor analysis even when there is no single underlying source of variance permeating all tasks (see Thomson, 1916). In short, if a battery of tasks all tap a common set of cognitive processes, in an overlapping manner, such that each task shares at least one process with another, then a general factor can emerge despite the fact that no one cognitive process is required by all tasks. Thomson's theory, and others that followed, have become known as "sampling theories" of *g* because the battery of tasks that reveal *g* all "sample" from a large number of cognitive processes. According to proponents of this view, the notion of *general* cognitive ability is not necessary to account for *g*.

Historically, sampling theories of *g* have garnered much less attention than general ability theories. However, Thomson's simple framework, first illustrated in 1916, has recently been resurrected and interest in sampling theories of *g* is on the rise (Bartholomew, Deary, & Lawn, 2009; Conway et al., 2011; Kovacs, 2010; van der Maas et al., 2006). We will return to a discussion of sampling theories later in the chapter but we will conclude here with a few important points: (1) Sampling theories are not the same as the idea of multiple intelligences (cf., Gardner, 1983); (2) Sampling theories do not deny the importance of *g* (cf., Gould, 1981); (3) The stumbling block for sampling theories of *g* is how to identify the underlying cognitive and neural processes, or mechanisms, that are being sampled by a battery of tests.

Intelligence tests

Spearman can be credited with launching the debate about theories of intelligence but his primary work on the topic was pre-dated by the work of French psychologist Alfred Binet. Binet was charged with developing a test that determined which children would most benefit from special schools devoted to teaching those who could not keep pace with a normal curriculum. Having two daughters, Binet had been observing tasks the elder daughter could perform that the younger still could not. This lead him to the idea of “mental age”, which refers to an individual child’s cognitive ability relative to his or her peers of the same age, as well as to younger and older children. Binet tested children in various age groups on various tasks to develop “benchmarks” of performance for each age group, which allowed him to determine whether an individual child was advanced, average, or behind relative to his or her peers of the same chronological age. For example, if a four-year-old child could complete most of the tasks that an average six-year-old child could perform, then s/he was considered advanced, and would be described as having a mental age of six. If an eight-year-old child could not complete the majority of tasks completed by other children of the same age but could only complete tasks that an average six-year-old could, then s/he was classified as behind, and so too would be described as having a mental age of six.

The Binet-Simon Scale (commonly referred to as the Binet Scale) was first published in 1911 (Binet & Simon, 1911) and included tasks such as defining words, pointing to parts of the body, naming objects in a picture, repeating digits, completing sentences, describing differences between similar items, saying as many words that rhymed with another word as one could in a minute, telling time and figuring out the time if the hands of a clock were reversed, and cutting a shape in a folded piece of paper and determining what the shape would be once unfolded. Binet was the first researcher to systematically study average cognitive abilities by age and his scales are the basis of modern IQ tests.

Lewis Terman, a psychologist at Stanford University, translated the French test into English and administered it to groups of children in the United States. After confirming a high correlation with teachers' rating of students' cognitive abilities, Terman began adapting the tests for broader use in American schools. Terman tested a larger group of children, approximately 1,000 children between the ages of four and fourteen in communities with average socio-economic statuses, and found that many of the benchmarks established by Binet had to be shifted. For example, if a test item was too hard for children of a particular age group then it was shifted down and used as a benchmark for a younger age group. This is a clever approach to testing but as noted at the outset of the chapter, it is completely data-driven, and not motivated by any theoretical account of cognitive development. Terman also standardized the number of test items per age, and removed items that did not test satisfactorily. In 1916, Terman published the Stanford Revision of the Binet–Simon Scale (Terman, 1916), which became known as the Stanford-Binet. The scoring procedure differed from the original Binet-Simon Scale and followed an idea first proposed by Stern (1912), that IQ should be quantified as mental age divided by chronological age multiplied by 100. For example, in our example above, a 4-year old with a mental age of 6 would have an IQ of 150 (i.e., 6/4*100 = 150), which is extremely high, and an 8-year old with a mental age of 6 would have an IQ of 75 (i.e., 6/8*100 = 75), and would be classified as having a borderline deficiency. The Stanford-Binet Scales are still used to calculate the IQs of children and adults. The scales, currently in their 5th edition, produce scores that approximate a normal distribution in which 100 is the mean, and the standard deviation is 16 (most IQ tests use a mean of 100 with standard deviations of 15 or 16).

These standardized scales allowed for the formal assessment of children's intellectual abilities, and were used to provide appropriate education and other services. IQ tests became used for both placement into special education programs for those who scored below average and for placement into gifted education programs for those who scored above average. Soon, classification of adults' intellectual abilities for predicting future performance was also sought. The United States Army conducted the first large-scale effort to this end.

At the start of U.S. involvement in World War I, the year after the Stanford-Binet was published, the U.S. Army needed to quickly screen a large number of men for military service, place recruits into military jobs, and select those to be assigned to leadership training. Testing one individual at a time, as was done with the Stanford-Binet, was considered too time-consuming. The U.S. Army wanted a multiple-choice test that could be administered to large groups of test-takers. Lewis Terman, the developer of the Stanford-Binet, with Henry Goddard, Carl C. Brigham and led by Robert Yerkes, developed two tests to aid the U.S. Army in quickly assessing a large number of potential soldiers. The Army Alpha battery included tasks such as sentence unscrambling, mathematical word problems, and pattern completion. The Army Beta battery was created for illiterate men and for those who failed the Army Alpha. The Army Beta did not rely on reading ability, but used pictures and other visual stimuli. Tasks included filling in missing components of pictures (e.g., a face without a mouth, a rabbit missing an ear), visual search, and recoding.

One particular army recruit, David Wechsler, had a Master's degree in Psychology and volunteered to score the Army Alpha as he was waiting for induction. Based on his education and experience with the Army Alpha, the Army later tasked Wechsler with administering the Stanford-Binet to recruits who had performed poorly on the Army Beta. Wechsler felt that the Stanford-Binet emphasized verbal abilities too much and in 1939 developed the Weschler-Bellevue test (Wechsler, 1939), which in 1955 became the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). As an individually-administered IQ test for adults, the scoring problem of the original Binet scale was apparent: adults' mental ages do not increase substantially like children's. In other words, an average 34-year-old and an average 36-year-old are more alike than an average 4 year-old and an average 6-year-old in cognitive abilities, so therefore the mental age divided by chronological age times 100 formula was not appropriate. Wechsler instead compared scores with the average score of someone of the same age. Wechsler also developed the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949). The WAIS and the WISC yield a verbal IQ score (which highly correlates with the Stanford-Binet [for a review see Arnold & Wagner, 1955]), a performance IQ score, and an overall IQ score (the average of the verbal and performance IQ scores). The current versions of the WAIS and WISC yield an aggregate IQ score from four factor indices: verbal comprehension, perceptual reasoning, processing speed, and working memory.

Following World War I, the Army Alpha was released as an unclassified document and immediately became popular within businesses and educational institutions. Brigham, who helped develop the Army Alpha and Beta, administered his own version of the Army Alpha test to Princeton University freshmen and students at a technical college in New York City. Soon after, the College Board asked Brigham to lead a group in developing a standardized admissions test to be used around the country. The first SAT, then named the Scholastic Aptitude Test, was administered to high school students in 1926. The 1926 SAT included vocabulary, arithmetic, number series, classification, artificial language learning, analogies, logical inference and reading comprehension. The SAT has been revised many times since 1926 and is currently the most widely used college entrance exam. It is important to note, however, that the revisions are largely data-driven, in the same manner that Terman revised the Stanford-Binet tests.

John C. Raven, a British psychologist, strove to develop a measure of intelligence that was based on a theoretical framework as opposed to data-driven. To this end, he created a test to capture the eductive ability (the ability to find meaning) component of Spearman's *g*. Raven also believed that many of the available IQ tests were too difficult to interpret so created a test that was simple to administer and score. The takers of his test are presented with matrices in which a piece of the pattern is removed and multiple options for the missing piece are presented, with only one correct option among them. The matrices are progressively more difficult to solve, so are typically scored simply as the number of correct responses. The test is non-verbal and very little prior knowledge is necessary. For these reasons, the Ravens Progressive Matrices (Penrose & Raven, 1936; Raven, 1938) is now one of the most popular tests of intelligence and is used around the world among people who speak different languages.

Performance on intelligence test has considerable impact on individuals' lives. At a young age, children are classified as normal, gifted, or deficient. Their education and the peers with whom they are grouped to interact are decided primarily on these tests. For those who score below average, as they approach adulthood the amount of government funding and non-profit service received for having a cognitive disability is contingent upon intelligence test scores. For those who score normal or above average, as they approach adulthood, those wishing to attend college have different opportunities for admission and funding (in the U.S.) based on their performance on college entrance exams. One's college education again has a considerable effect on the peers with whom one interacts and one's future career.

We have discussed just a few of the large number of standardized intelligence tests that have been developed over the last 100 years (for a more comprehensive review, see Urbina, 2011). We conclude this section with two main points: (1) the majority of standardized intelligence tests are revised in a data-driven manner, and are therefore difficult to explain in terms of psychological theory; and (2) the impact that these tests have had on modern society is enormous. While the consequences can be negative, for example, a student not being admitted to the school of her choice even though she perhaps could have succeeded, the enterprise is not all bad. For instance, before the SAT was developed, wealth, class, and ethnic background largely determined acceptance to a prestigious university in the United States. For many people, standardized tests created opportunities that otherwise would not have been possible. That said, new research on working memory might offer an even more promising way forward.

Working memory

Working memory (WM) is a construct developed by cognitive psychologists to characterize and help further investigate how human beings maintain access to goal-relevant information in the face of concurrent processing and/or distraction. For example, suppose you are attempting to schedule an airline reservation using the Internet. To search for flights you must mentally maintain the departure city, the arrival city, and the dates of travel while you scan the multiple options presented and ignore potentially distracting information, such as pop-up advertisements, or incoming email. WM is required to remember the critical travel information while concurrently searching potential flights and ignoring irrelevant information. Many important cognitive abilities, the same ones that fall under the umbrella of intelligence, such as academic achievement, learning, problem-solving, reading comprehension, and reasoning require WM because for each of these activities, some information must be maintained in an accessible state while new information is processed and potentially distracting information is ignored.

Working memory is considered to be a limited-capacity system and according to most contemporary theoretical accounts of WM, this capacity constrains intelligence, evidenced by the fact that individuals with lesser capacity tend to perform worse on most intelligence tests than individuals with greater capacity. For example, older children have greater working memory capacity (WMC) than younger children, the elderly tend to have lesser WMC than younger adults, and patients with certain types of neural damage or disease have lesser WMC than healthy adults. There is even a large amount of variation in WMC within samples of healthy young adults, for example, in samples of college students. In all these cases, individuals with greater WMC almost always perform better on intelligence tests than individuals with lesser WMC.

Research on the specific nature of the cognitive and neural mechanisms that constitute WM is extremely active in both Psychology and Neuroscience. For example, a search of Google Scholar using the exact phrase "working memory", constrained to 2006 – 2010, yields 127,00 links. If we go back 20 years and conduct the same search, constrained to articles published between 1986 and 1990, it yields 6,670 links. Step back another 20 years, 1966 to 1970, and the search yields 243 links and most of these don't use the phrase working memory as it is used today. It is therefore safe to say that research on WM is relatively new, compared to research on intelligence.

Given the fervent activity of research on WM, it impossible to summarize all current work here. Instead, we describe just one of the more influential theoretical models of WM and provide references to other influential models and/or frameworks. Our apologies in advance to any colleagues not cited. There's simply too much contemporary work for any one chapter to cite.

The most influential model of WM over the last few decades is the multi-component model of Baddeley and Hitch (1974, revised subsequently by Baddeley in 2000 and by Baddeley, Allen, & Hitch in 2011). Historically, the multi-component model has evolved from the concept of a unitary short-term memory (STM) system. According to a dominant view of memory in Psychology in the mid-20th century, information has to pass through a single STM store to enter or exit long-term memory (Atkinson & Shiffrin, 1968). In other words, STM serves as a gateway to further information processing and consequently plays a key role in higher-order cognitive abilities such as reasoning, planning, and problem solving. It soon became clear that things were more complex than originally thought. Experiments showed that disrupting the functioning of STM had little impact on complex cognitive task performance (Baddeley & Hitch, 1974). Furthermore, the existence of patients with normal long-term learning but impaired STM capacity could not be explained by the unitary account of STM (Shallice & Warrington, 1970).

Largely motivated by these patient data, British Psychologists Alan Baddeley and Graham Hitch proposed a new model of WM that replaced the concept of a unitary STM with a more dynamic system involving the interplay of attention and multiple short-term storage buffers (Baddeley & Hitch 1974). The model was called the "multi-component model of working memory" and has had a major influence on memory research ever since it was first proposed. The model originally consisted of three major components: the "central executive", which is an attentional control system that is supplemented by two passive storage buffers namely the "phonological loop" which stores and processes verbal information and the "visuo-spatial sketchpad" which stores and manipulates visual and spatial information. In a more recent update of the model, Baddeley added a fourth component to the tripartite system - the "episodic buffer" which serves as an interface between the executive, the buffers, and long-term memory (Baddeley, 2000, Baddeley, et al. 2011).

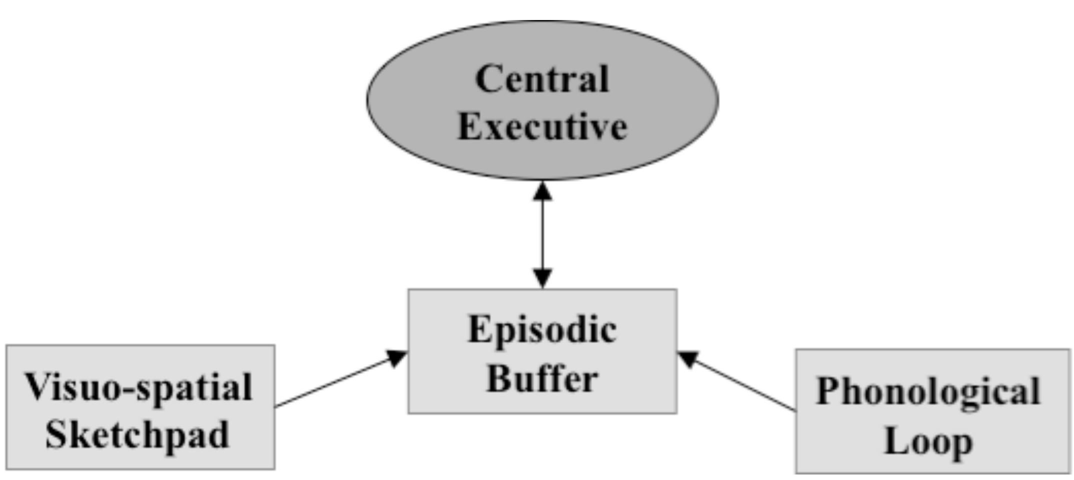

FIGURE 1. Simplified representation of the multi-component WM model based on Baddeley et al. 2011

The model's structure (see Figure 1) is largely based on the study of neuropsychological patients and the so-called dual-task methodology, in which subjects have to complete several memory-taxing tasks at the same time. Experiments of this kind showed that some pairs of tasks interfered with each other more so than other pairs of tasks. For example, if you are asked to repeatedly utter the word "the" and also remember a list of spoken digits or visually presented colors, you will most likely find it easier to remember the colors than the digits. According to Baddeley and Hitch this demonstrates separability of storage buffers, with words being stored in the loop (which causes interference) and visual information being stored in the sketchpad.

The most extensively investigated WM component is the phonological loop. It is assumed to consist of a passive phonological store that can hold speech-based information for up to two seconds coupled with an articulatory control process that prevents decay in the store by reactivating the fading phonological representations via subvocal rehearsal (Baddeley, 1986). Importantly, it has been proposed that the phonological loop might have evolved in humans as a "language learning device", in other words - as a system to facilitate the process of learning languages (Baddeley, Gathercole, & Papagno, 1998). Just imagine having to thank a Quechua speaker in Bolivia. You will need to be able to keep the sound sequence "*diuspagarapusunki*" in your phonological loop in order to repeat it correctly. Most Anglo-Saxon speakers will present difficulties in repeating this new word after having heard it only once, simply because it might exceed the limits of their phonological loop. Without an adequate temporary representation of the phonological sequence of this new word in the phonological loop, a robust long-term-memory representation will not be constructed and so the unfamiliar word will most likely not become part of your vocabulary after being exposed to it only once.

The second short-term storage system featured by the model is the visuo-spatial sketchpad. The current model of this subcomponent is less well advanced than for the phonological loop. The visuo-spatial sketchpad is responsible for the limited short-term storage and possibly the binding of visual and spatial information and is thought to be fractionable into separate visual, spatial, and haptic components (Baddeley, 1997, Baddeley et al., 2011). Like the phonological loop, it might consist of a passive temporary store and a more active rehearsal process (Logie, 1995). The visuo-spatial sketchpad is involved in tasks that require the recall of spatial or visual features such as finding your way through a supermarket or remembering the faces of your students after the first lecture.

The third temporary storage system is the episodic buffer (Baddeley, 2000; Baddeley, Allen, & Hitch, 2010, 2011). It is the most recently developed subcomponent of the model and was added in response to the increasing problems that the tripartite model encountered in explaining the way in which the visuo-spatial and phonological systems interact and how WM communicates with long-term memory (see Baddeley, 2000 for a review). Despite intense research efforts over the last 10 years, the episodic buffer "remains a somewhat shadowy concept" (Baddeley et al., 2010, p. 240). According to the most recent position (Baddeley et al., 2011) it represents a capacity-limited passive store that is capable of storing integrated chunks of information (i.e. episodes). In contrast to the loop and the sketchpad it can be considered as a "higher-level" storage buffer in that it links both of these to long-term memory and is assumed to be accessible through conscious awareness.

The most essential component of the WM system is the central executive (Baddeley, et al., 2010; Vandierendonck, De Vooght & Van der Goten, 1998). In contrast to the buffers it does not encompass a storage function but instead represents a purely attentional subsystem that controls the subsystems in a domain-free manner (Baddeley & Logie, 1999). The initial specification of the central executive was largely based on the Norman and Shallice (1986) model of executive control. It has been described as a homunculus that enables the working memory system to focus, divide, and switch attention in order to process, access, and store more information than would be possible by the relatively passive, limited-capacity short-term storage buffers alone (Baddeley, 1996). Whether or not it is a unitary system or is composed of different subcomponents, is open to debate (Baddeley, 1996, 2006).

We believe it is fair to say that the multi-component model of working memory has stood the test of time. It has stipulated a great deal of investigation over the last 35 years, and although far from being embraced by everybody, continues as one of the leading models in the field. The popularity of the model is partially related to its simple structural approach which is particularly useful in describing and understanding a range of neuropsychological deficits in adults as well as in children (Baddeley, 2003, 1990; Gathercole & Baddeley, 1990; Gathercole, Alloway, Willis, & Adams; Papagno, Cecchetto, Reati, & Bello, 2007).

Alternative models have, however, been developed in recent years and they provide a slightly different view of the WM system. Whereas the multi-component model bears a strong structural focus by separating WM into distinct components with different features, alternative WM theories emphasis functions and processes over structure (see Cowan, 2005; Engle & Kane, 2004; Jonides et al., 2008, Nairne, 2002).

Tests of working memory capacity

One of the core features of the Baddeley model of WM, and one that is not disputed by other theories, is that WM consists of multiple interacting mechanisms. At a general level, there are separable components for different kinds of information (for example, verbal vs. spatial) and for different types of processes (for example, memory vs. attention). At a more specific level, there are different mechanisms for particular processes, such as encoding, stimulus representation, maintenance, manipulation, and retrieval. Evidence for separable components and for different mechanisms comes from a variety of sources including dual-task behavioral experiments, neuropsychological case studies, and more recently, neuroimaging experiments. This is a critical point to remember when considering the measurement of WMC. A spatial WM task that requires the manipulation of information is very different from a verbal task that doesn't require manipulation but does require encoding, maintenance, and retrieval, yet each is dependent upon WM and can therefore fairly be considered a test of WMC.

We therefore define WMC as the maximum amount of information an individual can maintain in a particular task that is designed to measure some aspect(s) of the WM system. This has caused some confusion in the scientific community because different researchers often use different tasks to measure WMC, and this can lead to different conclusions regarding the nature of individual differences in WMC. We will have more to say about this dilemma in our discussion of psychometrics. Here we describe some of the more popular measures of WMC.

Complex span tasks

Complex span tasks were designed from the perspective of the original WM model, discussed above (Baddeley & Hitch, 1974). There are many different versions of complex span tasks, including reading span (Daneman & Carpenter, 1980), operation span (Turner & Engle, 1989), counting span (Case, Kurland, & Goldberg, 1982), as well as various spatial versions (Kane et al., 2004; Shah & Miyake, 1996). Complex span tasks are essentially "dual tasks"; the subject is required to engage in a relatively simple secondary task in between the presentation of to-be-remembered stimuli. For example, in the counting span task, the subject is presented with an array of objects on a computer screen. The objects typically differ from one another in both shape and color, for example, circles and squares colored either red or blue. The subject is instructed to count a particular type of object, such as blue squares. After counting aloud, the subject is required to remember the total and is then presented with another array of objects. The subject again counts the number of blue squares aloud and attempts to remember the total. After a series of arrays has been presented the subject is required to recall all the totals in correct serial order. Thus, maintaining access to the to-be-remembered digits is disrupted by the requirement of counting the number of objects in each array, which demands attention because multiple features (shape and color) have to be bound together to form each object representation (Treisman & Gelade, 1980). Indeed, the point of the secondary task is to engage attention and therefore disrupt active maintenance of the digits. This process is thought to create an ecologically valid measure of WM, as proposed by Baddeley and Hitch (1974), because it requires access to information (the to-be-remembered digits) in the face of concurrent processing (counting).

Several different versions of complex span tasks have been developed over the last 30 years. The different versions all have the same basic structure but differ in terms of the type of stimuli that are presented for recall (digits, letters, words, spatial locations) and the type of secondary task that is used to engage attention and disrupt maintenance (counting the number of objects in an array, reading sentences aloud, solving simple math problems, judging whether a figure is symmetrical or not). In most complex span tasks the number of stimuli presented for recall varies from trial to trial, typically from 2 to 7, and average recall performance among college students is about 4-5 (for more details, see Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005).

A battery of complex span tasks is defined as a group of several tasks, and typically the tasks in the battery differ with respect to the type of stimuli to be remembered and/or the type of secondary task. When a battery is administered to a large group of subjects, a positive manifold emerges, just like Spearman (1904) observed when looking at children's grades. That is, different versions of complex span correlate strongly with each other and typically account for the same variance in other measures of cognitive ability, such as the SAT (Turner & Engle, 1989). For example, Kane et al. (2004) administered several verbal and several spatial complex span tasks and the range of correlations among all the tasks ranged from $r = .39$ to $r = .51$. After statistically removing variance specific to each individual task, the correlation between "latent" variables representing spatial complex span and verbal complex span was $r = .84$. These results suggest that individual differences in complex span are largely determined by cognitive and neural mechanisms that are domain-general, akin to the measures that gave rise to Spearman's *g*.

Simple span tasks

Simple span tasks, for example, digit span or letter span, in contrast to complex span, do not include an interleaved secondary task between each presentation of to-be-remembered stimuli. For example, in digit span, one digit is presented at a time, typically one per second, and after a series of digits the subject is asked to recall the digits in correct serial order. Simple span tasks are among the oldest tasks used in memory research. Digit span was included in the first intelligence test (Binet, 1903) and is still included in two popular tests of intelligence, the WAIS and the WISC.

However, simple span typically does not reveal very strong correlations with other measures of cognitive ability (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Engle et al., 1999; Kane et al., 2004). Also, individual differences in simple span tasks are largely determined by domain-specific cognitive and neural mechanisms. We know this because within-domain correlations among simple span tasks are higher than cross-domain correlations among simple span tasks (Kane et al., 2004). Moreover, this domain-specific dominance is greater in simple span tasks than in complex span tasks (Kane et al., 2004). Also, patients with localized neurological damage or disease may exhibit normal performance on a simple span task with spatial materials yet exhibit a severe decrement on a simple span task with verbal materials, and vice-versa. These results suggest that individual differences in simple span are largely determined by cognitive and neural mechanisms that are domain-specific, unlike the measures that gave rise to Spearman's *g*. Therefore, when we consider the link between WMC and intelligence, we will focus on complex span tasks, not simple span tasks.

Visual array comparison tasks

Memory span tasks, both simple and complex, have a long tradition in cognitive Psychology. However, they are not ideal measures of the amount of information a person can "keep active" at one moment in time because the to-be-remembered stimuli must each be recalled, one at a time, and therefore performance is susceptible to output interference. In other words, a subject might get a score of 4 on a memory span task but it's possible that more than 4 items were actively maintained. Some representations might be lost during recall (Cowan et al., 1992).

For this reason, the visual array comparison task (Luck & Vogel, 1997) was developed as a measure of memory capacity. There are several variants of the visual array comparison task but in a typical version subjects are presented with an array of several items that vary in shape and color and the presentation is extremely brief, for example, a fraction of a second. After a short retention interval, perhaps just a second, the subject is then presented with another array and asked to judge whether the two arrays were the same or different. On half the trials the two arrays are the same and on the other half one item in the second array is different. Thus, if all items in the initial array are maintained then subjects will be able to detect the change. Most subjects achieve 100% accuracy on the task when the number of items is less than 4 but performance begins to drop as the number of items in the array increases beyond 4.

Visual array comparison tasks have not been used as often as memory span tasks to investigate individual differences in cognitive ability. However, recent research shows that array comparison tasks account for nearly as much variance in cognitive ability as complex span tasks (Cowan et al., 2005; Cowan et al., 2006; Fukuda, Vogel, Mayr, & Awh, 2010). The precise relationship between visual array comparison tasks, complex span tasks, and measures of intelligence remains unclear and is an active topic of research.

N-back tasks

The process of *updating* working memory is considered to be one of the most fundamental characteristics of the system. Information that is relevant to a current goal needs to be represented in a readily accessible state and must continuously be updated in accordance with changes in the environment. One popular updating task is called the n-back. In an n-back task, the subject is presented with a continuous stream of stimuli, one at a time, typically one every 2-3 seconds. The subject's goal is to determine if the current stimulus matches the one presented n-back. The stimuli are often verbal, for example, letters or words, but they can also be non-verbal, for example, visual objects, or spatial locations.

N-back tasks have been used extensively in neuroimaging experiments because the timing of stimulus presentation is easily controlled and the response requirements are simple. Numerous imaging experiments have demonstrated that the brain regions recruited to perform an n-back task are also recruited when performing intelligence tests (see Kane & Engle, 2002 for a review). Moreover, accuracy on an n-back task is correlated with scores on a test of intelligence and this correlation is partially mediated by neural activity in these common brain regions (Burgess, Gray, Conway, & Braver, 2011).

Coordination and transformation tasks

All of the above-mentioned WM tasks require subjects to recall or recognize information that was explicitly presented. There is another type of WM task, which we label "coordination and transformation", because subjects are presented with information and required to manipulate and/or transform that information to arrive at a correct response. For example, consider "backward span" tasks. Backward span tasks are similar to simple span tasks except that the subject is required to recall the stimuli in reverse order. Thus, the internal representation of the list must be transformed for successful performance. Another example is the letter-number sequencing task. The subject is presented with a sequence of letters and numbers and required to recall first the letters in alphabetical order and then the numbers in chronological order. Another example is the alphabet-recoding task. The subject is required to perform addition and subtraction using the alphabet, for example, C – 2 = A. On each trial, the subject is presented with a problem, C – 2, and required to generate the answer, A. Difficulty is manipulated by varying the number of letters presented, for example, CD – 2 = AB.

Kyllonen and Christal (1990) found very strong correlations between WMC and reasoning ability, using a variety of WM tasks that can all be considered "coordination and transformation" tasks (*r*s between .79 and .91). Also, Oberauer and colleagues demonstrated that the correlation between WMC and fluid intelligence does not depend upon whether WM is measured using complex span tasks or transformation tasks, suggesting that coordination and transformation tasks tap the same mechanisms as complex span tasks, meaning that they too are domain-general (Sub et al., 2002).

The link between intelligence and working memory capacity

Now that we have considered the various ways in which intelligence and WMC are measured, we are ready to evaluate the empirical evidence linking WMC and intelligence. The number of published papers reporting a significant correlation between WMC and intelligence is enormous, so to make this discussion tractable we start with two meta-analyses, both focused more specifically on the relationship between WMC and fluid intelligence. The two analyses were conducted by two different groups of researchers, one estimated the correlation between WMC and fluid intelligence to be r = .72 (Kane, Hambrick, & Conway, 2005) and the other estimated it to be r = .85 (Oberauer, Schulze, Wilhelm, & Süß, 2005). More recent studies also demonstrate correlations in this range, and most scholars agree that the relationship between WMC and fluid intelligence is very strong. Kane et al. (2005) summarized the studies included in their meta-analysis in a table, which is reproduced here (see Table 1). Each of the studies included in the meta-analysis administered several tests of WMC and several tests of fluid intelligence and factor analysis was used to determine the strength of the relationship between the two constructs. A variety of WM tasks were used in these studies, including complex span, simple span, and coordination and transformation tasks.

Table 1
Correlations Between WMC and Gf/Reasoning Factors Derived From Confirmatory Factor Analyses of Data From Latent-Variable Studies With Young Adults

Study	WMC tasks	Gf/reasoning tasks	r(95% CI)
Kyllonen & Christal (1990)			
Study 2: *n* = 399	ABC numerical assignment, mental arithmetic, alphabet recoding	Arithmetic reasoning, AB grammatical reasoning, verbal analogies, arrow grammatical reasoning, number sets	.91 (.89, .93)
Study 3: *n* = 392	Alphabet recoding, ABC21	Arithmetic reasoning, AB grammatical reasoning, ABCD arrow, diagramming relations, following instructions, letter sets, necessary arithmetic operations, nonsense syllogisms	.79 (.75, .82)
Study 4: *n* = 562	Alphabet recoding, mental math	Arithmetic reasoning, verbal analogies, number sets, 123 symbol reduction, three term series, calendar test	.83 (.80, .85)
Engle, Tuholski, et al. (1999; *N* = 133)	Operation span, reading span, counting span, ABCD, keeping track, secondary memory/ immediate free recall	Raven, Cattell culture fair	.60 (.48, .70)
Miyake et al. (2001; *N* = 167)	Letter rotation, dot matrix	Tower of Hanoi, random generation, paper folding, space relations, cards, flags	.64 (.54, .72)
Ackerman et al. (2002; *N* = 135)	ABCD order, alpha span, backward digit span, computation span, figural-spatial span, spatial span, word-sentence span	Ravens, number series, problem solving, necessary facts, paper folding, spatial analogy, cube comparison	.66 (.55, .75)
Conway et al. (2002; *N* = 120)	Operation span, reading span, counting span	Raven, Cattell culture fair	.54 (.40, .66)
Süß et al. (2002; *N* = 121[a])	Reading span, computation span, alpha span, backward digit span, math span, verbal span, spatial working memory, spatial short-term memory, updating numerical, updating spatial, spatial coordination, verbal coordination	Number sequences, letter sequences, computational reasoning, verbal analogies, fact/opinion, senseless inferences, syllogisms, figural analogies, Charkow, Bongard, figure assembly, surface development	.86 (.81, .90)
Hambrick (2003; *N* = 171)	Computation span, reading span	Raven, Cattell culture fair, abstraction, letter sets	.71 (.63, .78)
Mackintosh & Bennett (2003; *N* = 138[b])	Mental counters, reading span, spatial span	Raven, mental rotations	1.00
Colom et al. (2004)			
Study 1: *n* = 198	Mental counters, sentence verification, line formation	Raven, surface development	.86 (.82, .89)
Study 2: *n* = 203	Mental counters, sentence verification, line formation	Surface development, cards, figure classification	.73 (.66, .79)
Study 3: *n* = 193	Mental counters, sentence verification, line formation	Surface development, cards, figure classification	.41 (.29, .52)
Kane et al. (2004; *N* = 236)	Operation span, reading span, counting span, rotation span, symmetry span, navigation span	Raven, WASI matrix, BETA III matrix, reading comprehension, verbal analogies, inferences, nonsense syllogisms, remote associates, paper folding, surface development, form board, space relations, rotated blocks	.67 (.59, .73)

Note. WMC = working memory capacity; Gf = general fluid intelligence; 95% CI = the 95% confidence interval around the correlations; WASI = Wechsler Abbreviated Scale of Intelligence.
[a] *N* with the complete data set available (personal communication, K. Oberauer, July 7, 2004). [b] *N* for each pairwise correlation ranged from 117 to 127.

The studies referenced in Table 1 did not use either visual array comparison or n-back tasks and only involved adult participants . However, the more recent studies referenced above have used these tasks and also found correlations of the same magnitude. For example, Fukuda et al. (2010) used visual array comparison tasks and the correlation between WMC and fluid intelligence was $r = .66$. Burgess et al. (2011) used measures from an n-back task and the correlation between WMC and intelligence was $r = .43$.

In contrast, Ackerman et al. (2005) argued that the correlation between WMC and intelligence is weak, and moderate at best. Ackerman and colleagues chose to focus their meta-analysis on individual tasks, rather than factors from a factor analysis. The problem with this approach is that individual tasks are more susceptible to task-specific influences and are therefore less accurate assessments of the main construct under investigation, in terms of both reliability and validity. If multiple tasks are used then the common variance among tasks can be used to derive a “latent” variable required by all tasks. To better understand this argument, a more detailed discussion of latent variables, and psychometrics is necessary, so we turn to that topic now.

Psychometrics

Psychometrics is a field of study concerned with the theory and technique of measuring psychological constructs, including the measurement of intelligence and working memory capacity, as well as personality traits. Generally speaking, a psychometric study involves administering a large battery of tasks to a large sample of individuals and then analyzing the correlations among the tasks to identify underlying factors, or latent variables, that can account for large portions of variance within and across tasks in the battery. The data from psychometric studies are analyzed using multivariate statistical procedures, such as factor analysis and structural equation modeling (SEM). SEM is also known as *causal modeling* because the psychometric theories under investigation imply causal relationships among variables, despite the correlational nature of the data.

This approach is powerful because competing theories about the structure of intelligence can be objectively compared with empirical tests. For example, it is possible to administer a battery of tests to a large group of students and test whether a one-factor model, like the one initially proposed by Spearman (1904), "fits" the data better or just as well as a two- or three-factor model. In SEM, if a one-factor model fits the data as well as a two-factor model then the 1-factor model is preferable because it is more parsimonious.

As mentioned above, the most common interpretation of factors, or latent variables, in factor analyses and in causal models, is that they represent a single source of variance that is common to all the tasks that "load" onto that factor. Furthermore, in causal models, a factor is purported to cause performance on the manifest variables, that is, the actual tests or tasks administered to people in the sample, such as intelligence tests or working memory tasks. However, also mentioned above, many psychometricians dating back to Thomson (1916) have demonstrated that while this may be a valid interpretation of factors and the causal relationship between a construct and performance, it is not necessary to postulate a unitary source of variance from a factor. To reiterate, sampling theories of *g* can account for the emergence of *g* from a battery of tasks that taps a vast number of underlying cognitive processes in an overlapping fashion. This means that working memory capacity and intelligence may be correlated because measures of each construct share multiple cognitive processes, not because they share one general factor. If sampling theories of *g* are correct then causal models of *g* that posit a causal link between a *g* factor and every single task in the model are wrong. This view, which implies the rather radical notion that there is no such thing as general cognitive ability, has recently gained traction (Bartholomew, Deary, & Lawn, 2009; Conway et al., 2011; Kovacs, 2010; van der Maas et al., 2006).

Conclusion

This chapter has shown that working memory and intelligence are two psychological constructs that were developed for different purposes and have different histories and theoretical underpinnings. As a result, measurements of intelligence and measurements of working memory capacity look very different from one another. However, scores on intelligence tests are strongly correlated with scores on working memory tasks. Therefore, the enterprise of intelligence testing could be replaced by a new enterprise of working memory tasking. In our opinion, this would be a welcome shift. However, the *g* dilemma remains.

References

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

Alloway, T. P. & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*, 20-29.

Arnold, F. C., & Wagner, W. K. (1955). A comparison of Wechsler Children's Scale and Stanford-Binet Scores for eight- and nine-year olds. *Journal of Experimental Education, 24*(1), 91-94.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 742-775). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, UK: Cambridge University Press.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology, 49*(1), 5-28.

Baddeley, A. D. (1997). *Human memory: Theory and practice* (revised ed.). Hove: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baddeley, A. D. (2006). Working memory: an overview. In S. J. Pickering (Ed.), *Working memory and education* (pp. 1-31). London: Elsevier.

Baddeley, A. D., Allen, R. J., & Hitch, G.J. (2010). Investigating the episodic buffer. *Psychologica Belgica, 50*, 223-243.

Baddeley, A. D., Allen, R. J., & Hitch, G.J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*, 1393-1400

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In B. G. (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). New York: University Press.

Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thompson's bonds model of intelligence. *Psychological Review, 116*, 567-579.

Binet, A. (1903). L'Etude Experimentale de I'Intelligence. Paris: Schleicher

Binet, A., and Simon, T. (1911). *A method of measuring the development of young children.* Lincoln, IL: Courier Company.

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

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

Case, R., Kurland, D. M., Goldberg, J. (1992). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*(3), 386-404.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. New York: Elsevier Science Pub. Co.

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

Conway, A. R. A., Cowan, N., Bunting, M. F. Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183.

Conway, A. R. A., Getz, S. J., Macnamara, B., & Engel de Abreu, P. M. J. (2011). Working memory and fluid intelligence. In Sternberg, R. J., & Kaufman, S. B. (Eds.). *The Cambridge handbook of intelligence*. Cambridge, UK: Cambridge University Press.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1-17.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Matox, S., Hismjatullina, A., Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42-100.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition, 34*(8), 1754-1768

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 4, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review,3*(4). 422-433

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), The psychology of learning and motivation (pp. 145–199). New York: Academic Press.

Engle, R. W., Tuholski, S.W., Laughlin, J. E., Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309-331.

Frey, M. C. & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the SAT and general cognitive ability. *Psychological Science, 15(6),* 373-398.

Fukuda, K., Vogel, E. K., Mayr, U.,& Awh, E. (2010). Quantity not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin and Review, 17*, 673-679.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336-360.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology, 93*(3), 265-281.

Gould, S. J. (1981). *The Mismeasure of Man*. New York: W.W. Norton & Co.

Guilford, J. P. (1988). Some changes in the structure of intellect model. Educational and *Psychological Measurement, 48*, 1-4.

Holzinger, K., & Swineford, F. (1937). The Bi-factor method. *Psychometrika, 2*, 41-54.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59,* 193-224.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66-71.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189-217.

Kovacs, K. (2010). *A component process account of the general factor of intelligence.*Doctoral dissertation submitted to the University of Cambridge.

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

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memoryfor features and conjunctions. *Nature, 390*(6657), 279-281.

Mackintosh, N. J. (1998). *IQ and Human Intelligence*. Oxford: Oxford University Press.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53-81.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson & G. E. Schwarts & D. Shapiro (Eds.), *Consciousness and self-regulation. Advances in research and theory* (Vol. 4, pp. 1-18). New York: Plenum Press.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. (2005). Working Memory and Intelligence--Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 61-65.

Papagno, C., Cecchetto, C., Reati, F., & Bello, L. (2007). Processing of syntactically complex sentences relies on verbal short-term memory. Evidence from a short-term memory patient. *Cognitive Neuropsychology, 24*(3), 292-311.

Penrose, L. S., & Raven, J. C. (1936). A new series of perceptual tests: Preliminary communication. *British Journal of Medical Psychology.*

Raven, J. C. (1938). *Progressive Matrices: A perceptual test of intelligence.* London: H. K Lewis. Psychological Corporation.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores - A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22*(2), 261-273.

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

Stern, W. (1912). *Die Psychologische Methoden der Intelligenzprüfung.* Leipzig: Barth.

Sternberg, R. J. (1990). *Wisdom: Its nature, origins, and development*. New York: Cambridge University Press.

Terman, L. M. (1916). *The measurement of intelligence.* Boston, MA: Houghton Mifflin.

Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology, 8,* 271-281.

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

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Urbina, S. (2011). Tests of intelligence. In Sternberg, R. J., & Kaufman, S. B. (Eds.). *The Cambridge handbook of intelligence*. Cambridge, UK: Cambridge University Press.

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842-861.

Vandierendonck, A., De Vooght, G., & Van der Goten, K. (1998). Interfering with the central executive by means of a random interval repetition task. *Quarterly Journal of Experimental Psychology, 51A*, 197-218.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1949). *Manual for the Wechsler intelligence Scale for children.* New York: The Psychological Corporation.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

[1] Assessments of intelligence are typically referred to as "tests" while measures of working memory capacity are referred to as "tasks".