Assessing Working Memory Capacity for 21st Century Military Personnel Selection*

By Colin Kemp

As the United States mobilized for the war in Europe a century ago, its Army faced a formidable challenge at home : how to efficiently and effectively assess the mental abilities of millions of prospective recruits? The solution to this problem, in the form of the group-administered cognitive ability test, would not only revolutionize military personnel selection, but subsequently also applicant evaluation and selection across government departments, civilian industries, universities, and even prospective combatants for an entirely different battlefield : the National Football League (NFL).[1] Cognitive ability tests, and in particular the mass-testing variety, were deemed "*the most practical contribution made to humanity by all of psychology*" (e.g., Anastasi & Urbina, 1997; Cronbach, 1960). And perhaps because of this success, the cognitive ability tests used in personnel selection contexts today remain highly redolent, with respect to their form and content, of their century-old archetypes – which were designed as they were for ease of administration and scoring in the pre-computer era. Certainly, the early general cognitive ability tests were developed in the absence of a guiding theory of human cognition (Mackintosh, 2011), and subsequent tests were validated against their predecessors, culminating in a century of often derivative measures. This is not to say that group tests of general cognitive ability are not useful predictors of job performance ; however, the rise of cognitive psychology in the last quarter of the 20th century bestowed a new framework from which to better understand human cognition and to develop improved tests thereof – a fact that, interestingly, has done little to influence the course of the group testing juggernaut.

The present article considers working memory capacity (WMC), a cognitive psychology construct that has generated prodigious amounts of research in recent years due to the recognition of its ubiquitous role in human cognition. WMC is demonstrated to predict a litany of diverse phenomena and outcomes, including reading comprehension (e.g., Daneman & Carpenter, 1980 ; Daneman & Merikle, 1996), multitasking ability (e.g., Colom, Martinez-Molina, Shih & Santacreu, 2010), second language comprehension (e.g., Linck, Osthus, Koeth, & Bunting, 2013), stereotype threat (e.g., Hoffman, Schmeichel & Baddeley, 2012), susceptibility to mind wandering (e.g., McVay & Kane, 2012b), attentional control (e.g., Unsworth & Engle, 2007), the ability to deal with life-event stress (Klein & Boals, 2001), tactical decision-making (Furley & Memmert, 2012), and even the

[1] The NFL has used the Wonderlic Personnel Test (WPT), a traditional paper-and-pencil cognitive ability test, for the assessment of prospects since the 1970s (Lyons, Hoffman & Michel, 2009).

early onset of Alzheimer's disease (Rosen, Bergeson, Putnam, Harwell & Sunderland, 2002), among others. Most importantly, however, WMC is demonstrated to be a strong predictor of fluid intelligence, the ability to learn, reason, and to solve novel problems[2] – an essential 21st century workforce competency (Burrus, Jackson, Xi & Steinberg, 2013). Ultimately, WMC is so central to human cognition that some researchers have suggested it to be the fundamental basis of intelligence[3] itself (e.g., Colom, Rebolloa, Juan-Espinosa & Kyllonen, 2004). Nevertheless, WMC tests are, hitherto, almost exclusively the domain of cognitive psychologists and neuroscientists who seek to elucidate the mechanisms and nuances of human cognition, and are seldom used in personnel selection contexts. As will be demonstrated in this article, tests of WMC deserve a place in personnel selection, not only because of their ability to reliably predict important aspects of cognition, but also because of a litany of advantages that they offer over the status quo measures that, one hundred years later, continue to dominate the field of personnel selection.

Historical Context

Development of the First Intelligence Tests and Theory of Intelligence

When the US Army embarked on its objective to create efficient measures for large-scale personnel selection, the study of intelligence and the application of intelligence testing were mainly inchoate disciplines. However, the Army researchers had at their disposal two coeval, but independent, developments in psychology : an individualized intelligence test, developed in France by Alfred Binet and Theodore Simon (Binet & Simon, 1904), and the first psychometric theory of cognitive ability, elucidated by British psychologist Charles Spearman (1904).

Binet and Simon's Individual Intelligence Test

Binet and Simon's seminal intelligence test was intended neither for personnel selection nor mass-testing purposes. Rather, their test was developed, on behalf of the French Ministry of Education, to assess children in the public school system for special educational needs. Binet based his test on his own multifaceted view of intelligence, which he perceived as a confluence of numerous faculties including attention, memory, imagination, common sense, judgment, abstraction, and coping with everyday problems (Urbina, 2011). Importantly, this test was an individual test, which required a highly-trained test administrator to administer and score on a one-on-one basis. This test was later translated from French into English by Henry Goddard, and served as the impetus for

[2] For instance : Ackerman, Beier & Boyle, 2005 ; Colom, Abad, Quiroga, Shih & Flores-Mendoza, 2008 ; Conway, Getz, Macnamara & Engel de Abreu, 2011 ; Salthouse & Pink, 2008 ; Shelton, Elliott, Matthews, Hill & Gouvier, 2010.

[3] The terms *intelligence*, *cognitive ability*, *general cognitive ability*, and *general mental ability* (GMA) are used synonymously in the present article to refer to 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*" (Gottfredson, 1997, p.13).

Lewis Terman's improved version, Stanford-Binet test, which was suitable for assessing general intelligence in adults (Urbina, 2011). In fact, this measure is still in widespread use today in its current incarnation as the *Stanford-Binet Intelligence Scales, Fifth Edition*.

Spearman's Theory of Intelligence

With respect to the theory of intelligence, Spearman had observed that academic performance ratings in a group of students tended to exhibit what he called a "*positive manifold*" (i.e., pervasive positive correlations) ; specifically, students who did well in one subject tended to do well in other subjects (even ostensibly unrelated ones). To further investigate this phenomenon, Spearman developed the first psychometric method of factor analysis, which he used to extract the common statistical variance from a matrix of observed score data. Ultimately, Spearman concluded that performance across disparate subjects (or tests) could be explained by a general factor, which he called ***g***, for general intelligence (Willis, Dumont & Kaufman, 2011). This ***g*** factor was used to explain why all the tests on a general intelligence battery were positively correlated with one another – the correlations were effectively due to each individual test's correlation with the ***g*** factor (Willis *et al*., 2011). More recently, Jensen (1998) noted that this pattern of positive correlations among cognitive ability subtest scores is the single most replicated finding in all of psychology.

Notably, Spearman originally conceived a two-factor model of intelligence, the second "factor" being the test-specific variance related to particular tests or measures in a battery. As this test-specific variance comprised measurement error and other variance specific to particular tests, Spearman concluded that there is no common variance between tests that is not explained by ***g***. The corollary of Spearman's assertion is that there are no "kinds" of mental tests that belong to different classes of specific mental abilities (Conway & Kovacs, 2013). Ultimately, for explaining performance on intellectual tasks, Spearman considered ***g*** to be the singular factor of importance. In fact, Spearman's factor-analytic method would allow him to later claim that Binet – who built his test based on a multi-faceted view of intelligence – had unwittingly created a good measure of overall ***g*** (Mackintosh, 2011).

Army Alpha and Beta Tests for Group Testing

For the purposes of the US Army screening during World War I, individualized tests like the Stanford-Binet were not suitable for achieving its objective of efficient large-scale assessment of recruits. Specifically, individualized intelligence tests of this type require a highly-trained psychologist to administer the test items to examinees on a one-on-one basis ; furthermore, these tests are time-consuming, and scoring is often complicated and cannot be accomplished rapidly. Essentially, the logistical challenges of administering and scoring individualized intelligence tests precluded their use in large-scale personnel selection, and the US Army researchers were required to find an innovative solution to assess the constructs measured by the Stanford-Binet in a manner that was amenable to group administration and rapid scoring (Jones & Thissen, 2007).

A solution to the administration and scoring problem was presented by Arthur S. Otis, who had studied as a graduate student with Lewis Terman at Stanford University. For his doctoral thesis, Otis developed methods for creating and scoring the now-ubiquitous multiple-choice test item format (Jones & Thissen, 2007). Using this method, the Army researchers, who included Otis and Terman, quickly developed the Army Alpha and Beta Test for screening recruits. By the end of WW I, these tests had been administered almost 2 million times (Jones & Thissen, 2007), and served to usher in the modern era of mass testing.

Subsequent Psychometric Developments

After WW I, a variety of group tests for cognitive ability were developed for various purposes, including personnel selection and university admissions ; likewise, psychometric models of intelligence continued to advance apace (for a review, see Urbina, 2011 ; Willis *et al*., 2011).

Thurstone's Primary Mental Abilities

An alternative to Spearman's ***g*** theory, proposed by Louis Thurstone (1938), is that there is no general factor of intelligence, but rather a number of specific and unrelated ability factors that explain performance on the various domains of cognitive ability tests (e.g., spatial, verbal, reasoning, etc.). Spearman's ***g***-centric conceptualization of intelligence was criticized by Thurstone as being too simplistic to account for such an ostensibly complex phenomenon. Using a different type of factor analysis and a wider range of mental ability tests, Thurstone was unable to derive a single general factor of intelligence. Conversely, he extracted seven factors which he termed "*primary mental abilities*". These were : verbal comprehension, inductive reasoning, perceptual speed, numerical ability, verbal fluency, associative memory, and spatial visualization (Willis *et al*., 2011). Thus, Thurstone came to a conclusion about the underlying structure of human intelligence that was markedly different from Spearman's.

Hierarchical Psychometric Models of Intelligence

It eventually became apparent, however, that neither model adequately explained the empirical data. It is currently held that intelligence is represented neither as a single ability, nor as a set of uncorrelated specific abilities – but hierarchically through the incorporation of both views into a single model. Hierarchical psychometric models are derived through the application of factor analysis to batteries of mental ability tests. In these models, one or more general factors reside at the top of the ability structure, while specific factors fall on the lower levels ; higher-order factors account for existing inter-correlations between specific lower-order factors.

Fluid-Crystallized (*Gf*-*Gc*) Theory

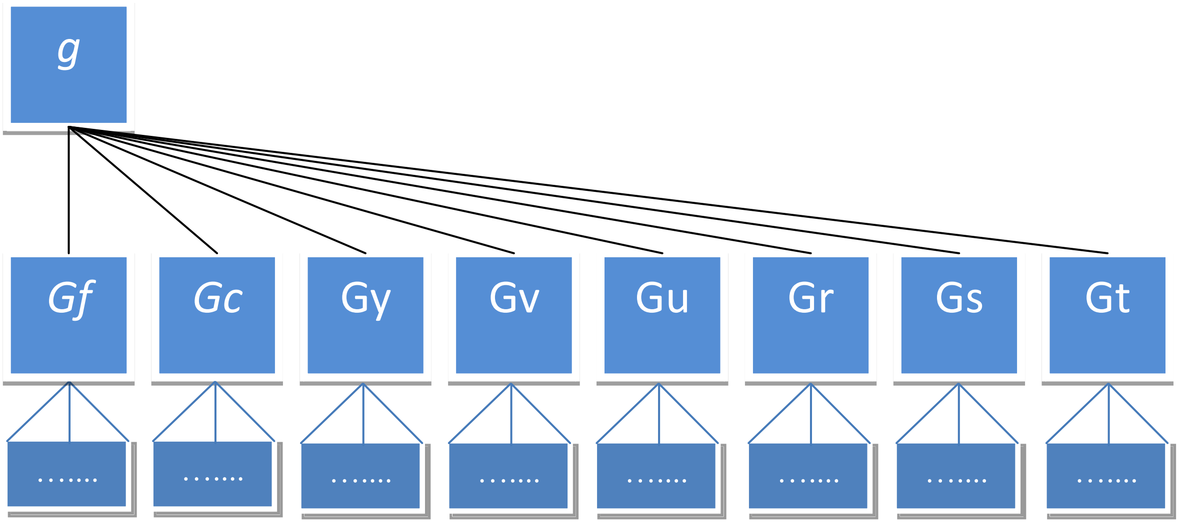
The first such theory was proposed by Vernon (1950), although Horn and Cattell (1966) proposed perhaps the most influential hierarchical theory of intelligence, based on two different components of ***g***: fluid intelligence (*Gf*) and crystallized intelligence (*Gc*). *Gf* reflects "*basic processes of reasoning and other mental activities that depend only minimally on learning and acculturation*" (Carroll, 1993, p.624), such as the evaluation of relationships, making of inferences, and abstract reasoning. *Gf* was postulated to be dependent on the efficient functioning of the central nervous system (CNS). *Gc*, on the other hand, reflects acculturated or learned knowledge. In a similar vein, Ackerman (1996) conceptualized *Gf* as 'intelligence as process' and *Gc* as 'intelligence as knowledge'.

The *Gf*-*Gc* theory is a two-stratum model, with the higher-order *Gf* and *Gc* ability factors represented on the top stratum, and over 40 first-order, narrower (i.e., specific) factors located on the lower stratum. The two primary components, *Gf* and *Gc*, are obliquely related, typically correlating at 0.5 or higher (Davidson & Downing, 2000). Despite its name, however, the model has subsequently been extended to include a number of other broad ability factors alongside *Gf* and *Gc* on the top stratum (Davidson & Downing, 2000). Each broad ability factor is presented in the order of its importance to general intelligence according to its level of information processing. In its most recent incarnation, the top stratum *Gf*-*Gc* factors are, in order of their significance : fluid reasoning (*Gf*), acculturated knowledge (*Gc*), short-term apprehension and retrieval (*GSM*), visual processing (*Ga*), long-term storage and retrieval (*TSR*), cognitive processing speed (*Gs*), correct decision speed (*CDS*), reading and writing (*GRW*), and quantitative knowledge (*Gq*). As with all of the hierarchical models, each broad second-order factor in the *Gf*-*Gc* theory explains the inter-correlations between the respective specific ability factors clustered underneath. Furthermore, each first-order factor is more proximal to actual performance on a specific test and more distal to general cognitive ability than its superordinate second-order factor.

Three-Stratum Theory

Carroll proposed a similar hierarchical model of intelligence, the result of an exhaustive re-analysis of more than 460 cognitive ability datasets constituting almost a century of empirical research. The resulting three-stratum theory (Carroll, 1993 ; 1997) is similar to the *Gf*-*Gc* theory, with an additional general factor of intelligence (tantamount to Spearman's ***g***) at the apex of the hierarchy. More specifically, the three-stratum model features the general factor ***g*** at the top (stratum III), eight stratum II factors in the middle, which, in order of their respective ***g***-loadings are : fluid intelligence (*Gf*), crystallized intelligence (*Gc*), general memory and learning (*Gy*), visual perception (*Gv*), auditory perception (*Gu*), retrieval ability (*Gr*), cognitive speed (*Gs*), and processing speed (*Gt*), and finally numerous specific abilities at stratum I, including Thurstone's primary mental abilities. Although this model is well supported empirically, the strata are not rigidly defined – for example, Carroll noted that additional factors could indeed exist between strata II and III (Davidson & Downing, 2000).

Figure 1: Carroll's (1993) Three-Stratum Model of Cognitive Ability



Legend : **g** = General Intelligence ; **Gf** = Fluid Intelligence ; **Gc** = Crystallized Intelligence ; **Gy** = General Memory and Learning ; **Gv** = Broad Visual Perception ; **Gu** = Broad Auditory Perception ; **Gr** = Broad Retrieval Ability ; **Gs** = Broad Cognitive Speediness; **Gt** = Processing Speed.

Cattell-Horn-Carroll Theory

Carroll's (1993) rigorous re-analysis yielded the first empirically-based taxonomy of cognitive abilities presented in a single framework (McGrew, 2009). Due to the influence of the *Gf-Gc* theory on the three-stratum model, and indeed the considerable amount of overlap between the two, these models are often combined under the rubric of the Cattell-Horn-Carroll (CHC) theory.

Why Is *g* Important in Personnel Selection Contexts?

Spearman thought of ***g*** as a kind of mental horsepower or "*general fund of mental energy*" (Spearman, 1914, p.103). He explained general intelligence as the "*apprehension of one's own experience, the eduction of relations and the eduction of correlates*" (Mackintosh, 2011, p.7). What ***g*** is phenomenologically, however, is still fiercely debated to this day (for a review, see Davidson & Kemp, 2011). Throughout the years, some critics have suggested that ***g*** is either a statistical artifact, or essentially a spurious construct of questionable utility. E.G. Boring (1961) infamously quipped that intelligence is what intelligence tests measure. Certainly, ***g*** is, in one important respect, derived statistically from cognitive ability test data as the shared variance underlying performance on a battery's subtests (as in Spearman's original application and the modern psychometric models, described below). Recent evidence, however, disputes the assertion that ***g*** is simply a statistical artifact, as a number of studies have indicated that the ***g*** factor derived from completely distinct intelligence test batteries are isomorphic[4] – in other words, evidence that ***g*** is not mere statistical conjuring.

[4] Johnson, Bouchard, Krueger, McGue & Gottesman, 2004 ; Johnson, te Nijenhuis & Bouchard, 2008.

Ultimately, though, ***g*** possesses an undeniable external validity: no other construct in the domain of individual-differences psychology demonstrates such fecund prediction of real-world outcomes as ***g***. For example, ***g*** has been demonstrated to predict academic achievement, health-related behaviours, moral delinquency, socioeconomic status, racial prejudice, divorce, accident proneness, occupational status, and mortality rates, among other things (Ones, Viswesvaran & Dilchert, 2005). In personnel selection contexts, ***g*** is vitally important because it is the best known predictor of training and job performance across virtually all jobs (Gottfredson, 1997 ; Ones *et al*., 2005 ; Schmidt & Hunter, 1998), with no other factor (e.g., personality, education, experience, etc.) demonstrating equally substantial and reliable predictive validities (Gottfredson, 1997). This is particularly salient in large military organizations such as the Canadian Armed Forces (CAF), where personnel selection encompasses selecting for dozens, or even more than 100 occupations – for which applicants typically do not have any previous work experience, which limits the use of other valid selection measures (e.g., behaviour-based interviews, work samples). In addition, having to develop and validate separate interviews and work sample measures for each occupation would be a massively expensive and time-consuming endeavour, which is a primary reason why traditional measures of ***g*** are so frequently used in personnel selection in such contexts. And while measures of ***g*** have certainly proved successful in personnel selection contexts – to which a century of use attests –, more recent developments in cognitive psychology offer opportunities for improvement. One such opportunity is presented by the construct of WMC.

Working Memory and Working Memory Capacity

Working memory (WM) is defined as "*the ability to actively maintain task-relevant information in the service of a cognitive task*" (Boduroglu, Minear & Shah, 2007, p.55) ; so essential is working memory to human cognition that virtually every complex cognitive task depends on it (Boduroglu *et al*., 2007). WMC relates to working memory in that it refers specifically to the individual differences construct that reflects the limited capacity of a person's working memory system (Wilhelm, Hilderbrandt & Oberauer, 2013). In the past decade, the construct of WMC has generated a vast amount of interest due to its strong association with intelligence, in particular *Gf*.[5]

WM and WMC are relatively new constructs in the field of cognitive psychology. Until the 1970s, memory was typically conceptualized as either long-term memory (LTM) or short-term (STM) memory ; some individual intelligence tests, such as the Weschler Adult Intelligence Scale (WAIS) and the Weschler Intelligence Scale for Children (WISC) even contained tests of STM, which typically entailed the memorization and serial recall of a list of letters or digits. The construct of WM, however, was introduced in a massively influential book chapter by Baddeley and Hitch in the mid-1970s (Baddeley & Hitch, 1974). In their seminal work, these authors proposed a functional working memory system;

[5] See for instance : Ackerman *et al*., 2005 ; Unsworth & Engle, 2005 ; Colom *et al*., 2008 ; Salthouse & Pink, 2008 ; Shelton *et al*., 2010 ;Conway *et al*., 2011.

they argued that short-term memory was unlikely to exist solely for the purpose of mentally rehearsing and regurgitating lists (such as telephone numbers and so forth : Conway *et al*., 2011). Rather, they proposed a working memory system that, like the familiar short-term memory store, could maintain information in a readily accessible state – but, pivotally, could also perform concurrent processing and maintain more information than a limited short-term storage system could possibly do. In later incarnations (Baddeley, 1986 ; 2000), this model was described as a multi-component system comprising several "slave" storage systems – in particular, a *phonological loop*, for holding verbal information and a *visual-spatial sketchpad*, for holding visual and spatial information – and, pivotally, a dedicated *central executive* that controls the processing and shunting of information around the aforementioned stores.

Note that STM and WM are often conflated constructs ; essentially, the key difference is that STM can be thought of as merely a system for the temporary storage of information, whereas WM entails a combination of storage *and* manipulation of information (Baddeley, 2012). This distinction gave rise to the prediction that WMC should be more strongly related to (and predictive of) higher-order general cognitive abilities than measures of STM (Conway *et al*., 2011). Indeed, Baddeley and Hitch's initial working memory model served as the direct impetus for the development of the first modern working memory measure, Daneman and Carpenter's (1980) iconic reading span task. In this task, participants are required to read aloud a series of sentences, while simultaneously remembering the last word of each sentence read for subsequent recall. This paradigm, known as the *complex span* task, operationalizes WMC as *simultaneous storage and processing* (i.e., remembering the last words and reading the sentences, respectively), which is intended to invoke both major components of Baddeley's multi-component model (i.e., the slave storage system and the executive). Importantly, complex span tasks like the reading span task were found to be a much better predictor of measures of reading comprehension (Daneman & Carpenter, 1980 ; Daneman & Merikle, 1996), and scores on the verbal SAT, for example, than simple span tasks (McVay & Kane, 2012*a*). And while it could easily be argued that the reading span predicts reading comprehension simply because the measure depends on reading comprehension in the first place, it turned out that complex span measures that do not involve reading (such as Engle's mathematical operation span measure ; Turner & Engle, 1989) perform equally well in terms of predicting reading comprehension as the original reading span task (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). Indeed, WM has been demonstrated to be a domain-general system (thus it is not overly important whether one uses a spatial working memory task or a verbal working memory task to tap the WMC construct), unlike STM which is much more domain-specific in nature (Kane *et a*l., 2004).

Since Daneman and Carpenter introduced their seminal reading span measure, many different types of WMC tests have been developed. These measures have been shown to predict a panoply of outcomes and phenomena, including reading comprehension (e.g., Daneman & Carpenter, 1980 ; Daneman & Merikle, 1996), vocabulary learning (e.g., Martin & Ellis, 2012), performance on the SATs (Turner & Engle, 1989), early onset Alzheimer's disease (Rosen *et al*., 2002), life event-event stress (Klein & Boals, 2001), tactical decision making in sports (Furley & Memmert, 2012), susceptibility to choking under pressure (Beilock & Carr, 2005), stereotype threat (Schmader & Johns, 2003), creative problem solving (Lee & Therriault, 2013), multitasking (e.g., Hambrick, Oswald, Darowski, Rench & Brou 2010 ; Colom *et al*., 2010), job simulation performance (Bosco & Allen, 2011), abstract reasoning (e.g., Engle, Tuholski, Laughlin & Conway, 1999), learning (e.g., Lilienthal, Tamez, Myerson & Hale, 2013), emotion regulation (Kleider, Parror & King, 2010), hindsight bias (Calvillo, 2012), attentional control abilities (e.g., Unsworth & Engle, 2007), the ability to follow directions (Foster, Shipstead, Harrison, Hicks, Redick & Engle, 2015), the ability to learn new computer programming languages (Foster *et al*., 2015), mind wandering (McVay & Kane, 2012b), the ability to suppress intrusive thoughts (e.g., Brewin & Smart, 2005), driving under distraction, and errors made by fatigued pilots (Engle, 2010), among others.

Most famously, however, WMC is demonstrated to be highly predictive of *Gf*, with recent meta-analyses indicating latent variable associations in the range of 0.72 (Kane, Hambrick & Conway, 2005) to 0.85 (Oberauer, Schulze, Wilhelm & Süß, 2005). Indeed, the fact that WMC is the strongest predictor of *Gf* is the most important conclusion to emerge from research into the neurocognitive basis of *Gf* to date (Chuderski, 2013). The importance of WMC extends, in turn, to general intelligence, given the critical finding that *g* relies substantially on *Gf* – recall from Carroll's (1993; 1997) model that *Gf* is the most important factor under ***g***, in that it accounts for more unique variance in general intelligence than does any of the other second-order factors, including *Gc*. And although WMC is robustly demonstrated to be highly related to *Gf* and overall ***g***, the mechanisms that underlie this association are currently a matter of intense speculation and debate (Chuderski, 2013).

Explaining the WMC-Intelligence Connection

While the fact that WMC and *Gf* are strongly related is undisputed, the mechanisms that underlie this relationship remain comparatively more inscrutable. Certainly, the question of why WMC is so strongly predictive of *Gf* remains one of the most hotly contested topics in the field of cognitive psychology today. Currently, a number of theories attempt to explain individual differences in WMC and its relation to higher-order cognition, in particular *Gf*. The debate surrounding these various views is ongoing and is currently a major area of research within cognitive psychology. Initially, it was thought that individuals possess a fixed pool of mental resources that they can allocate to both storage and processing, and that there is an inherent trade-off in that as processing demands increase, fewer resources are available for storage of information (see Unsworth, Fukuda, Awh & Vogel, 2014). More recently, this view has fallen out of favour as it is not fully supported by empirical findings ; for example, WM storage scores continue to predict higher-order cognition even after controlling for processing (see Unsworth *et al*., 2014). Modern views posit that individual differences in WMC are due to either differences in attentional control (Engle, 2002 ; Engle & Kane, 2004), capacity limits (Cowan, 2005), or the ability to retrieve information from secondary memory (Unsworth & Engle, 2007). Furthermore, Oberauer, Süß, Wilhelm and Sander (2007) posit that differences in WMC are due to differences in individual ability to build, maintain, and rapidly update arbitrary bindings among elements in memory. The debate concerning these different views is more than mildly evocative of that between Spearman and Thurstone regarding the psychometric structure of human cognitive abilities.

The attentional control view (e.g., Engle, 2002; Engle & Kane, 2004) posits that the capacity for controlled and sustained attention, in the face of distraction or interference, is the pivotal mechanism that underlies individual differences in WMC (Unsworth *et al.*, 2014). Which is to say (perhaps counter-intuitively, given its name) that WMC is a *de facto* attentional capability, rather than a storage or storage capacity one. Accordingly, individuals' high working memory capacities are thought to have superior attention and inhibitory capabilities than those of low WMC individuals, and therefore are better at actively maintaining information when faced with distraction (Wilhelm *et al*., 2013). With respect to predicting *Gf*, this view posits that individual differences in attentional control account for the strong correlations between WMC and *Gf*, a hypothesis which is partially supported by the finding that attentional control is strongly related with *Gf*, and partially mediates the association between WMC and *Gf* (Unsworth et al., 2014).

The primary-secondary memory theories (e.g., Unsworth & Engle, 2007) posit that WMC comprises a dual-store system. This view stems from findings that the capacity of immediate memory is limited to approximately four items (e.g., see Cowan, 2010), and that attentional control abilities have limited power to protect memory contents from being lost in the face of distraction. Therefore, it is posited that secondary (or long-term) memory is invoked in order to maintain WM contents (Unsworth *et al.*, 2014). In this respect, a limited capacity component (i.e., primary memory) maintains information ephemerally, while the more resilient secondary memory component stores information over longer periods ; therefore, the working memory system involves both active maintenance of information in primary memory, and controlled retrieval of information from secondary memory (Wilhelm *et al*., 2013). If the capacity of primary memory is only about four items (digits, letters, symbols, etc.), this account helps to explain why performance on complex span tasks frequently exceeds the putative four-item limit of primary memory ; in other words, individual differences in WM are in part due to individual differences in the ability to retrieve information from secondary memory that could not be actively maintained in primary memory (Unsworth *et al*., 2014). Proponents of this view argue that part of the reason why WMC and *Gf* are strongly related is because both rely on retrieval of information from secondary memory, and high WMC individuals are better at solving reasoning problems because they are able to recover pertinent information that has been displaced from primary memory (Unsworth *et al*., 2014). Certainly, WMC and secondary measures have been demonstrated to correlate both with each other, and with *Gf* (Unsworth *et al*., 2014), which lends support to this view.

Similarly, the capacity limit view of WMC (e.g., Cowan, 2005) also points to the fact that, theoretically, only about four items can be maintained in active memory ; however, there are large observed individual differences in WMC (Unsworth *et al*., 2014). If the capacity limit of working memory does vary across individuals in this manner, then individuals with larger capacities would be able to store more information in their working memories than individuals with lower capacities (Unsworth *et al*., 2014). This could in part explain the WMC-*Gf* relationship, in that individuals with greater WMC capacities are able to store, and concurrently process, more information related to the problems they are working on. Evidence to support the capacity limit view comes from the finding that tests specifically designed to measure the capacity of WM correlate with both the complex span measures typically used to assess WMC and with *Gf* (Unsworth *et al*., 2014).

Finally, the binding view of WMC posits that WM is a system for "*building, maintaining, and rapidly updating arbitrary bindings*" (Wilhelm *et al*., 2013, p.3), for example, to represent random lists, spatial arrays, or mental models (Wilhelm *et al*., 2013). In this view, WM is important for reasoning because the latter requires the building and manipulation of novel structures, and the observed capacity limits on WM stem from interference between bindings, which, in turn constrains reasoning ability (Wilhelm *et al*., 2013). Oberaurer *et al*. (2007) present several lines of evidence to support the binding view; most interestingly, tests designed to measure the ability to construct new structural representations were shown to correlate highly with measures of WMC and *Gf*, even when the former tasks did not rely on memory at all (Wilhelm *et al*., 2013).

In summary, a number of theories attempt to account for the strong observed relationship between WMC and *Gf* ; the debate regarding the validity of these accounts is an ongoing concern in the field of cognitive psychology. Interestingly, Unsworth *et al*. (2014) recently presented evidence to support the possibility that WMC is a multi-component system that entails some or all of the disparate views discussed above. If so, the current debate regarding the nature of WMC is indeed highly analogous to the debate between Spearman and Thurstone, which culminated in the synthesis of their contrary views into the modern three-stratum (Carroll, 1993 ; 1997) and CHC (McGrew, 2009) models of human cognitive abilities. For the purposes of using WMC measures in personnel selection, however, it is perhaps much less important to explain the underlying mechanisms that account for the association of these constructs – as the association itself both strong and unequivocal – than it is to ensure that the tests themselves are valid, reliable, and practical measures of WMC.

Tests of WMC

Prior to Daneman and Carpenter's (1980) seminal complex span measure, the reading span tests of immediate memory tended to be simple span measures (such as forward digit and letter spans). As Daneman and Carpenter (1980) demonstrated, complex span measures that invoke the working memory system (in particular, through interleaved storage and processing tasks, rather than just simple storage) demonstrate stronger associations with complex cognition than do simple span (i.e., short-term memory) tasks without a processing component. Furthermore, complex span tasks have been shown to be less domain-specific than short-term memory tasks ; for example, visual and verbal WMC tasks tend to inter-correlate more highly than do verbal visual and verbal short-term memory tasks, which suggests a domain-general working memory system (see Conway *et al*., 2011).

Currently, complex span tasks are the dominant WMC task paradigm in cognitive psychology. For example, Unsworth and Engle's (2005) automated operation span task is widely used in WMC research. The operation span task is analogous to the seminal reading span task; however, rather than having participants read disparate sentences while trying to remember the last word of each, the operation requires participants to remember a sequence of individual letters while performing mental arithmetic. Again, it is the simultaneous storage and processing nature of complex span tasks, rather than simply storage (in the case of simple span tasks) that is thought to invoke the WM system.

While complex span tasks are known to be valid and reliable measures of WMC (see Conway *et al*., 2005), other WMC task paradigms exist and are used in WMC research. These include modified simple span tasks, in particular those with very long lists or very rapidly administered stimuli ; scope of attention tasks ; transformation/ updating tasks ; and *N*-back tasks.[6] Importantly, each of these different kinds of tasks has been demonstrated to correlate well with WMC tasks of the other types, as well as with *Gf* (see Wilhelm *et al*., 2013).

Thus, whether one chooses to use, for example, a complex span task or a transformation task paradigm to assess WMC is not pivotal ; however, it is very important that a battery of WMC tasks is used in order to ensure that WMC is being reliably captured (Conway *et al*., 2005; Wilhelm *et al*., 2013). Specifically, no single task can perfectly measure any construct (whether WMC, spatial ability, attention, or so forth). By using only a single task, one risks confounding variance with the construct of interest (e.g., WMC) with task-specific variance. For example, a math-phobic individual could ostensibly perform poorly on the operation span task due to anxiety induced by the arithmetic processing component of the task. Conversely, an individual with an effective pneumonic strategy for remembering disparate letters could perform above his or her WMC limit on that specific task. In other words, by using a battery of tasks and taking a composite test score, a more reliable estimate of WMC can be obtained than for any single task.

Reasons for Minimal Use of WMC in Personnel Selection to Date

Given the extent to which validated measures of WMC are demonstrated to predict *Gf*, and the association of WMC with many lower- and higher-order cognitive processes in general, it is perhaps surprising that WMC tests are seldom used in personnel selection contexts. There are several possible reasons for this. First, while measures of WMC emerged out of the nascent discipline of cognitive psychology for the purposes of examining *intra-*

[6] For a review, see Conway *et al*., 2005 ; Conway *et al*., 2011; Wilhelm *et al*., 2013.

individual processes, general cognitive ability tests, as discussed above, were born out of the study of *inter-individual* differences at the dawn of the 20th century (Conway & Kovacs, 2013). Of note is that group intelligence tests, owing to their history, were developed and evolved largely out of an atheoretical framework – a legacy that remains much apparent in today's group tests of general cognitive ability (Schwartz, 1984). Specifically, the early tests (Binet and Simon's intelligence test and the Army Alpha and Beta tests, for example) were not based on any validated theory of cognition ; moreover, subsequent measures of this type were developed from, and validated against, these earlier tests, and so on, up to the present-day incarnation of tests. Ultimately, general cognitive ability test development was, and very much remains, a data-driven (i.e., a correlational) enterprise (Conway & Kovacs, 2013), rather than a theory-driven one. Indeed, Schwartz (1984) notes that while the field of psychometrics has advanced considerably, many cognitive ability tests "*were (and remain) largely atheoretical devices*" (Schwartz, 1984, p.36). Ultimately, past tests are often used to validate subsequent tests through correlational analyses, which largely explains why modern group tests of cognitive ability look very much like the older ones (i.e., if a new test deviates too much from its form and content of an older one, there is a risk for diminished convergent correlations and it might not be accepted as valid – which helps to maintain the status quo).

Secondly, the discipline of cognitive psychology emerged long after intelligence testing had already become a well-established field – and, again for an entirely different reason (i.e., to examine intra-individual processes rather than inter-individual differences) ; it is perhaps not surprising, then, that the tools and methods of cognitive psychologists have not crossed into the domain of intelligence testing. Third, and perhaps more importantly, is the fact that, within the intelligence testing community, WM was long thought not to be predictive of general cognition (Gignac, 2014). Indeed, there was an initiative to remove the sole memory test from the WAIS because it was considered to be "*a poor measure of intelligence*" (Gignac, 2014). Notably, however, many of the existing memory measures used on standard intelligence batteries were not true WM tests but, rather, STM tests, which, as discussed above, have demonstrated generally poor prediction of higher-order cognitive abilities, including *Gf*. Due to the common conflation of STM and WM, the notion that WMC is not predictive of intelligence may be particularly pervasive in the intelligence testing community.

Finally, a pragmatic reason for minimal use of WMC measures in personnel selection is that during the paper-and-pencil administration era, tests of memory would not be readily amenable to mass-testing methodologies. Many current personnel selection tests, when computerized, are often just digital versions of paper-and-pencil multiple-choice measures, and thus the paucity of WMC measures in personnel selection contexts could be indicative of the slow pace of technological testing advancements in personnel selection domains, and a general reluctance to deviate from a century of status quo measures. Certainly, modern computerized WMC measures offer a host of advantages over legacy paper-and-pencil measures (and digital facsimiles thereof), and these make a very compelling case for their use in personnel selection.

Compelling Reasons for Assessing WMC in Military Personnel Selection

WMC Is a Better Measure of *Gf* than Typical Personnel Selection Tests

Many traditional personnel selection tests, including the US Armed Services Vocational Aptitude Battery (ASVAB) and the iconic Wonderlic Personnel Test (WPT), strongly tap *Gc*, but to a much lesser extent *Gf*. For example, Roberts, Markham, Matthews and Zeidner (2005) note that the ASVAB oversamples from the *Gc* factor. Likewise, the WPT, a multiple-choice general cognitive ability test that has been administered over 200 million times since its inception as a personnel selection measure, is also demonstrated to be primarily a test of *Gc* (Hicks, Harrison & Engle, 2015). Indeed, the WPT, while a demonstrated predictor of acculturated knowledge, fails to predict the ability to learn and adapt in situations that require novel reasoning – in other words situations that require *Gf* (Hicks *et al.*, 2015). This is not surprising, because tests of this nature tend to assess acculturated knowledge domains such as vocabulary and mathematics ability (as does, notably, the Canadian Forces Aptitude Test [CFAT]). Given that *Gf* is more closely related to overall ***g*** than *Gc* (e.g., from Carroll's three-stratum model and the CHC model of cognitive ability), WMC tests are well-poised to augment (or perhaps in some instances replace) more traditional measures of ***g*** that tend to be highly-laden measures of *Gc*.

WMC Is Demonstrably a Better Predictor of Multitasking Ability than Tests of *g*

Many 21st century military occupations place strong demands on multitasking ability ; indeed, the ability to multitask is essential for anyone living and working in a highly technological society (Colom *et al.*, 2010). Multitasking can be defined as the cognitive ability required to perform "*multiple task goals in the same period by engaging in frequent switches between individual tasks*" (*ibid.*, p.543) ; it is often alternatively referred to as "*dual tasking*" or "*task switching*" (*ibid.*). Multitasking frequently emerges among the knowledge, skills, abilities, and other characteristics (KSAOs) identified for military occupations. A review of CAF job analyses indicates that multitasking is associated with tasks performed by a number of occupations, such as Clearance Diver, Military Observer, Airborne Electronic Sensor Operator, Pilot, Logistics Officer, Aerospace Control Officer, Small Unmanned Arial Vehicle (SUAV) Operator, Medical Technician, Intelligence Officer, Tactical Intelligence Operator, and Air Combat Systems Officer, among others (Kemp & St-Pierre, 2009). Likewise, multitasking has been identified as an important ability for civilian jobs, including pilots, drivers, and fire-fighting supervisors (Colom *et al.*, 2010).

Interestingly, not much is known about the predictors of multitasking performance (Colom *et al.*, 2010). However, the canonical WMC measure, the complex span task, requires the simultaneous storage of one set of information while processing of another set of information ; from a *prima facie* standpoint, these types of tasks would appear to be good predictors of multitasking, given the above definition of the construct. Indeed, this has been demonstrated to be the case. WMC measures have been shown in several studies to be strong predictors of the former (*ibid.*). Moreover, in a recent study, WMC was found to predict multitasking above and beyond ***g***. Specifically, Colom *et al*. (2010) revealed that for a set of simulations designed to reflect the cognitive demands faced by air traffic controllers, WMC, as measured by complex span tasks, was not only a strong predictor of multitasking performance, but also a stronger predictor than general intelligence (Colom *et al*., 2010). Likewise, Hambrick *et al*. (2010) demonstrated that WMC adds approximately twice as much unique variance in multitasking performance as does general mental ability.

Tests of WMC Produce Less Adverse Impact than Traditional Tests of *g*

Adverse impact is a legal concept that refers to situations in which group differences in test performance (e.g., between a majority and minority group) result in disproportionate selection or promotion ratios (Zumbo, 1999). Cognitive ability tests are generally considered to be unbiased (i.e., differential predictive validities as a function of gender, race, and ethnicity is not evidenced; however, they are known to produce adverse impact for some applicant subgroups (Ones *et al*., 2005).

One strategy to reduce adverse impact in personnel selection is to reduce a measure's reliance on acculturated knowledge (Hough, Oswald & Ployhart, 2001). Given the demonstrated association between working memory capacity and fluid intelligence, and the fact that WMC tests can be designed to possess few obvious acculturated knowledge dependencies, the latter may prove ideal for mitigating adverse impact in personnel selection testing. For example, the average black-white examinee score differences in the US on tests of general cognitive ability is equal to 15 IQ points (Colom *et al*., 2010), or one standard deviation (Ones *et al*., 2005).

However, Lynn (2006) demonstrated that the average difference for measures of WMC is equivalent to 6 IQ points, or approximately one-third of a standard deviation point. This translates into less potential adverse impact for tests of WMC relative to general cognitive ability tests, when used as personnel selection measures. As Hicks *et al*. (2015) suggest: "*Instead of focusing on tests of 'general mental ability' (…), organizations should consider using other constructs such as working memory capacity. Future work in this area should further validate its relationship to job performance along with other measures that have less emphasis on crystallized intelligence*" (Hicks *et al*., 2015, p.193).

Simple and Inexpensive Item Generation

Cognitive ability test items are surprisingly expensive to develop. For example, the industry standard, as of the year 2006, is up to (US) $1000 each for high-stakes personnel selection test items (Downing, 2006).

If items on traditional cognitive ability tests are compromised (e.g., its items are leaked onto the internet), these tests generally have to be replaced with a parallel version, potentially at great expense and inconvenience to the organization. For many computerized WMC tests, on the other hand, new item generation (and parallel test form development) can be performed almost instantly and at no additional cost, due to relatively simple automated item generation algorithms.

Reduced Impact of Guessing

WMC tests are typically designed to require examinees to completely recall their item responses, rather than simply selecting their responses from a small set of options, as is the case with multiple-choice test items. In complex span tasks, for instance, examinees are required to recall, completely from memory, the sequence of individual letters (or numbers, symbols, etc.) that was presented over the course of a trial (i.e., item) ; for example, an examinee might be required to recall the following sequence : 'AFQRNT' (keeping in mind that examinees are required to perform some processing task on unrelated data between the presentation of each letter to be remembered, which renders the memorization and recall of these sequences difficult). While the probability of correctly guessing a typical four option multiple-choice item is, at worst, 25% (i.e., one out of four), the probability of correctly guessing the aforementioned sequence of six letters is negligible (at worst, it is lower than one in 300 million[7]). Ultimately, if one is about to run out of time and/ or cannot determine a given item response for a personnel selection test item (for which no penalties are applied to incorrect answers, which is typically the case), then the obvious strategy is to take a guess. For example, on a multiple-choice test with four response options per item, an examinee could randomly guess any item (for instance the last, and putatively most difficult item) with 25% probability of a correct response, and any two items (for example, the last two most difficult items) with a small, but still feasible probability of 6%; successfully guessing just one or two items in this manner could markedly increase an examinee's standardized (i.e., norm-based) test score. For example, a mere one or two additional raw score points on the CFAT can increase an examinee's percentile rank score by up to six and eleven points, respectively, with obvious implications for individual competitiveness. The abysmally-low probability of correctly guessing even a single item response in many WMC task paradigms, on the other hand, renders this guessing strategy effectively futile.

Mitigation of Cheating

The ease with which WMC test items can be generated in real time, in addition to the often generic quality of WMC test item content (random sequences of digits, letters, symbols, etc.), confers yet another advantage over traditional cognitive ability tests : the ability to preclude a notorious form of cheating. One of the guiding requirements during the development of the Army Alpha and Beta tests was to design test forms for which cheating is difficult (Wainer, 2000). However, on traditional test forms, applicants can potentially cheat by obtaining the item content and/ or answers in advance, either through word of mouth from previous examinees, or through more systematic means (e.g., "brain

[7] Even if there are only 12 letters in the response set, and each letter can only be selected once, the probability of randomly guessing the correct response is still abysmal : one in 665,280.

dumping"[8] operations). Even when test forms that use traditional item content are generated from large item banks to create potentially unique tests for each individual (e.g., through computerized adaptive testing [CAT] methods), systematic cheating efforts remain a vexing and ongoing issue for high-stakes test developers.[9] Indeed, despite often extreme security precautions employed during proctored testing sessions (which include metal detectors, clothing searches, and cameras that monitor each individual examinee, in the case of the Graduate Record Examination [GRE]), high-stakes test developers are required to continually develop and validate new items in order to keep ahead of cheating efforts. A large part of the problem is that traditional multiple-choice items can be memorized by examinees during the test session, and subsequently disseminated. Indeed, brain dumping can result in the exposure of an item bank in a matter of months (see Smith & Prometric, 2004). Brain dumping operations, smaller-scale word of mouth transfer of item content, and like approaches can easily be prevented by using WMC measures that generate item content on the fly at the time of testing; this feature obviates reliance on item banks and/or pre-generated test forms. This feature of WMC tests thus enables not only unique tests (i.e., unique assemblages of individual items) for each examinee, but unique items as well. When each individual item is unique, brain dumping and similar cheating approaches are completely precluded.

Increased Measurement Precision through Adaptive Testing

Emphasized in the present article is the fact that cognitive ability tests used in personnel selection contexts have remained largely unchanged, in terms of their form and item content types, relative to the early group tests of cognitive ability that emerged one hundred years ago. The primary exception, however, is the revolution in CAT methods, which rose to prominence in the late 1990s with the advents of inexpensive personal computers and developments in psychometrics, particularly item response theory (IRT).

Adaptive tests automatically generate test forms in real time by selecting test items that are appropriately difficult for each examinee. For example, if an examinee correctly responds to a given test item, the next item selected by the computer will be more difficult, whereas an incorrect response will yield a subsequently less difficult item. By tailoring test forms to each individual test taker's level of ability, CATs confer increased test score reliability and/or decreased test length (de Ayala, 2009) over traditional test forms that are built to accommodate the entire range of examinee ability (i.e., traditional test forms are "one-size-fits-all" measures). IRT provides the statistical mechanisms by which to generate these test forms on the fly, and to directly compare examinee scores across these forms – even if the tests are different in terms of their overall difficulties and number of items (for adaptive test forms to be validly comparable, they must measure the same construct(s),

---

[8] 'Brain dump' sites are websites that host or sell test content (Drasgow, Nye, Guo & Tay, 2009).

[9] In a famous case, the first computerized adaptive version of the Graduate Record Examination (GRE) for college admissions was completely compromised through a systematic brain dumping effort in Asia, despite a large item bank and potentially unique test forms for each individual examinee (see Smith & Prometric, 2004).

however). By administering appropriately-difficult test items to each examinee, CATs are potentially shorter than standard test forms by up to 50% or more (de Ayala, 2009), rendering them much more efficient. Furthermore, since examinees receive fewer inappropriately easy (and thus potentially boring) items, and fewer inappropriately difficult (and thus potentially frustrating) items, examinee perceptions about the test-taking experience is often improved (Wainer, 2000).

In practice, however, CATs are difficult to implement. In order to make a traditional cognitive ability test adaptive, a large pool of multiple-choice items (e.g., 5 to 10 times as many as the standard test length) first needs to be developed and calibrated using IRT. IRT calibration is the statistical process by which individual item parameters, such as difficulty, discrimination, and "guessability" (i.e., *pseudo-guessing*) are estimated from a set of test data. IRT calibration requires a large volume of data – up to or exceeding 1000 cases per item for the 3-parameter logistic model that is used for multiple-choice items (DeMars, 2010) –, data that can be both expensive and very time-consuming to obtain. Moreover, once the test items have been calibrated and added to a CAT item pool, complicated item selection and exposure control algorithms are necessary in order to maintain test security (i.e., to protect against item overexposure), test content balance, and to ensure that overly-similar items are not administered to the same examinee during a test session. Implementing adaptive testing for computerized measures of WMC, on the other hand, is potentially much less complicated, and less likely to be fraught with unanticipated problems (such as poor content balancing, the administration of overly-similar items, and so forth). For example, WMC test items typically differ only with respect to their difficulty (e.g., having to remember a span of four letters versus a span of five letters), and are otherwise generic, which obviates the need to control for item content. This fact is pivotal, because for examinee scores to be directly comparable, test forms must be made to the same content specifications, which is a trivial matter with WMC tests – but, again, requires potentially complex content balancing algorithms with traditional cognitive ability test forms. For WMC tests, item difficulty is typically operationalized in terms of the number of pieces of information to be remembered. WMC items can often be generated on the fly with predictable difficulties – and thus expensive and time-consuming efforts to collect item calibration data for each new item, as well as cumbersome item exposure control protocols, can be avoided.

Reduction or Elimination of Time Constraints

Cognitive ability tests for personnel selection, such as the WPT and the CFAT, are often administered under fairly strict time pressure. Now, strict time pressure can preclude the use of advanced statistical methods such IRT, because it can introduce multi-dimensionality into the measurement process (de Ayala, 2009). WMC tests generally do not need to be administered under time pressure in order to increase item difficulty ; indeed, taking too long to respond to an item is ostensibly counter-productive due to the ephemeral nature of the contents of working memory. Another advantage of relaxed time constraints is that examinees can proceed at their own pace, "*which allows for a much wider range of* *response styles than is practical with traditional standardized tests*" (Wainer, 2000, p.11). In this light, tests of WMC may be useful for fair assessment of candidates with learning disabilities (LDs). For example, under Canadian law, an employer cannot discriminate against an applicant based on a disability (unless a *bona fide* occupational requirement is at stake ; individuals employed as truck drivers, for instance, must meet certain vision standards).[10] While LDs can negatively impact performance in cognitive testing in a number of ways, the most common negative impact is interference with "speeded" test performance (i.e., situations in which tests are timed, which is typically the case for personnel selection tests of ***g***) (Skomorovsky, 2006). As tests of WMC do not require strict time limits, they constitute an appealing alternative to timed tests of *g* in personnel selection contexts.

Identical Test Content for Multiple Language Groups

Applicant populations, in many countries, are not unilingual. For example, Canada has two official languages : French and English ; by law, CAF applicants have the right to choose which official language to be tested in during the personnel selection process. While this policy makes the CAF inclusive, it presents difficulties for its personnel selection test developers. Ultimately, each test form must be properly translated and validated in order to ensure sufficient measurement invariance across the separate language forms of the test. Crucially, it must be ensured that the test development and translation process has not induced any test bias or differential predictive validities (i.e., across the applicant groups) that would preclude the objective comparison of examinee test scores across the respective language forms of the measure – which is especially important, because hiring and promotion decisions are based directly on these test scores, and any bias in the measure would undermine fairness and lead to potential adverse impact.

Ultimately, cross-linguistic comparison is an often difficult, expensive, and time-consuming endeavour, one that requires multiple phases of item and instruction translation and back-translation, evaluation of the measures by subject matter experts, and statistical evaluation for differential item functioning (DIF) and/or differential test functioning (DTF). With WMC tests, there is no need to establish the equivalence of test content across language versions (aside from test instructions and response button labels) as the test content is identical.

Assessment of Allophones and Second-Language Potential

Allophones are individuals whose first language is not an official language of a country; for example, in Canada, allophones are individuals whose first language is neither French nor English (LeBlanc, Straver, Wright, Ruscito & Moffatt, 2016). In a recent article, LeBlanc *et al*. (2016) noted Canada's increasing reliance on immigration to expand its workforce ; also, they reported anecdotal evidence that some allophone CAF members are unable to complete military training because of purported deficits in official language

[10] Source : http://www.tbs-sct.gc.ca/psm-fpfm/ve/dee/dorf-eng.asp#toc5-3.

proficiency (LeBlanc *et al.*, 2016). Measures of WMC are potentially useful in assessing allophones in two respects. First, as is noted above, tests of WMC are considerably less dependent on language than many traditional cognitive ability tests, are thus are ostensibly less likely to produce adverse impact for allophones in personnel selection. Second, given the strong association of WMC with reading comprehension and other aspects of higher-order cognition, there is reason to argue that individuals with greater WMC should perform better on second language processing tasks than individuals with lower WMC (Linck *et al.*, 2013) ; indeed, Hummel (2009), Leeser (2007), Linck, Hoshino and Kroll (2008), and Weissheimer and Mota (2009) have demonstrated that bilinguals with high WMC span scores performed significantly better on tasks that required second language processing. While more research is needed, WMC scores could thus potentially be used to inform judgments about second language learning potential in applicants.

WMC Testing in the Applied Military Context

Given the potential advantages of using WMC measures in personnel selection, the CAF is currently exploring the usefulness of WMC as a predictor of training and job performance. Specifically, the CAF is currently validating a battery of computerized WMC measures that are grounded in modern cognitive psychological theory. Because no single test can perfectly capture the construct of WMC (Conway *et al.*, 2005 ; Wilhelm *et al.*, 2013), the current CAF working memory battery (WMB) comprises five subtests : two complex span tests and three transformation/updating tasks that entail digit, letter, symbol, and/ or spatial working memory demands. The WMB is currently being pilot-tested for military occupations with ostensibly high working memory demands, including Pilot, Armour Officer, and Aerospace Control Officer. Larger-scale recruiting centre trials, for all occupations, are set to begin in autumn of 2016.

The WMB features minimal acculturated knowledge demands, identical official language (i.e., English and French) test content, and completely automated test tutorials, practice trials, and administration. Minimal reliance on acculturated knowledge should mitigate adverse impact, and the self-paced nature of the WMB will allow for a wider range of examinee response styles than do traditional cognitive ability tests (Wainer, 2000). Currently, the goal is not to replace the existing measures of general cognitive ability with the WMB, but rather to bolster the former with a strong measure of both WMC and, by proxy, *Gf*, in order to better select applicants for dynamic 21st century military occupations.

Summary and Conclusion

As this article elucidates, working memory is fundamental to human cognition, and yet most organizations that conduct cognitive ability testing, in order to assess and select applicants, do not attempt to measure this pivotal construct. Rather, many large organizations, including militaries, typically use traditional general cognitive ability tests of the type that have now existed for one hundred years. WMC, a comparatively new construct to emerge from the discipline of cognitive psychology, has been demonstrated to predict many important phenomena and outcomes that are ostensibly important for performance in military occupations, including multitasking, susceptibility to mind wandering, tactical decision-making, abstract reasoning, and errors made while fatigued. WMC is also highly predictive of *Gf*, the ability to learn and adapt – and the most important factor under ***g*** – a factor that many traditional cognitive ability tests used in personnel selection may not strongly measure, due to their high acculturated knowledge contents. Moreover, modern measures of WMC offer a host of advantages over traditional cognitive ability tests in the personnel selection context, including reduced potential for adverse impact, the ability to generate test stimuli in real time, mitigation against guessing and cheating, the ability to generate identical test content across language forms, and test stimuli that are dynamic rather than static. Given the importance of WMC and the potential advantages conferred by using tests thereof in personnel selection contexts, 21st century military organizations should strongly consider assessing this construct in personnel selection to bolster traditional cognitive tests.

References

ACKERMAN, P.L., "A Theory of Adult Intellectual Development : Process, Personality, Interests, and Knowledge", *Intelligence*, vol.22, 1996, pp.227-257.

ACKERMAN, P.L., M.E. **BEIER** & M.O. **BOYLE**, "Working Memory and Intelligence : The Same or Different Constructs ?", *Psychological Bulletin*, vol.131, 2005, pp.30-60.

ANASTASI, A.U. & A.S. **URBINA**, A.S., *Psychological Testing*, Ann Arbor, MI, Prentice Hall, 1997.

BADDELEY, A.D., *Working Memory*, London, Oxford University Press, 1986.

BADDELEY, A.D., "The Episodic Buffer : A New Component of Working Memory ?", *Trends in Cognitive Sciences*, vol.4, 2000, pp.417-423.

BADDELEY, A.D., "Working Memory : Theories, Models, and Controversies", *Annual Review of Psychology*, vol.63, 2012, pp.1-29.

BADDELEY, A.D. & G. **HITCH**, "Working Memory", *Psychology of Learning & Motivation*, vol.8, 1974, pp.47-89.

BEILOCK, S.L. & T.H. **CARR**, "When High-Powered People Fail : Working Memory and 'Choking under Pressure' in Math", *Psychological Science*, vol.16, 2005, pp.101-105.

BINET, A. & T. **SIMON**, "Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux", *L'année psychologique*, vol.11, 1904, pp.191-244.

BODUROGLU, A., M. **MINEAR** & P. **SHAH**, "Working Memory", in F. Durso (ed.), *Handbook of Applied Cognition*, Chichester, UK, Wiley, 2007 (2nd ed., pp.55-82).

BORING, E.G., "Intelligence as the Tests Test It", pp.210-214 in J.J. Jenkins & D.G. Paterson (eds.), *Studies in Individual Differences : The Search for Intelligence*, East Norwalk, CT, Appleton-Century-Crofts, 1961.

BOSCO, F.A. & D.G. **ALLEN**, "Executive Attention as a Predictor of Employee Performance", pp.1-6 in L. Toombs (ed.), *Academy of Management Proceedings*, New York, Academy of Management, 2011.

BREWIN, C.R. & L. **SMART**, "Working Memory Capacity and Suppression of Intrusive Thoughts", *Journal of Behavior Therapy & Experimental Psychiatry*, vol.36, 2005, pp.61-68.

Burrus, J., T. **Jackson**, N. **Xi** & J. **Steinberg**, J., "Identifying the Most Important 21st Century Workforce Competencies : An Analysis of the Occupational Information Network (O* Net)", *ETS Research Report Series*, 2013, pp.i-55.

Calvillo, D.P., "Working Memory and the Memory Distortion Component of Hindsight Bias", *Memory*, vol.20, 2012, pp.891-898.

Carroll, J.B., *Human Cognitive Abilities : A Survey of Factor-Analytic Studies*, New York, Cambridge University Press, 1993.

Carroll, J. B., "The Three-Stratum Theory of Cognitive Abilities", pp.122-130 in D.P. Flanagan, J.L. Genshaft & P.L. Harrison (eds.), *Contemporary Intellectual Assessment : Theories, Tests, and Issues*, New York, The Guilford Press, 1997.

Chuderski, A., "When Are Fluid Intelligence and Working Memory Isomorphic and When Are They Not ?", *Intelligence*, vol.41, n°4, 2013, pp.244-262.

Colom, R., I. **Rebollo**, A. **Palacios**, M. **Juan-Espinosa** & P.C. **Kyllonen**, "Working Memory Is (Almost) Perfectly Predicted by *g*", *Intelligence*, vol.32, n°3, 2004, pp.277-296.

Colom, R., F.J. **Abad**, M.A. **Quiroga**, P.C. **Shih** & C. **Flores-Mendoza**, "Working Memory and Intelligence Are Highly Related Constructs, But Why ?", *Intelligence*, vol.36, 2008, pp.584-606.

Colom, R., A. **Martínez-Molina**, P.C. **Shih** & J. **Santacreu**, "Intelligence, Working Memory, and Multitasking Performance", *Intelligence*, vol.38, 2010, pp.543-551.

Conway, A.R., M.J. **Kane**, M.F. **Bunting**, D.Z. **Hambrick**, O. **Wilhelm** & R.W. **Engle**, "Working Memory Span Tasks : A Methodological Review and User's Guide", *Psychonomic Bulletin & Review*, vol.12, 2005, pp.769-786.

Conway, A.R., S.J. **Getz**, B. **Macnamara** & P.M.J. **Engel de Abreu**, "Working Memory and Intelligence", pp.394-418 in R.J. Sternberg & S.B. Kaufman (eds.), *The Cambridge Handbook of Intelligence*, New York, Cambridge University Press, 2011.

Conway, A.R., & K. **Kovacs**, "Individual Differences in Intelligence and Working Memory : A Review of Latent Variable Models", *Psychology of Learning & Motivation*, vol.58, 2013, pp.233-270.

Cowan, N., *Working Memory Capacity*, Hove, UK, Psychology Press, 2005.

Cowan, N., "The Magical Mystery Four : How Is Working Memory Capacity Limited, and Why ?", *Current Directions in Psychological Science*, vol.19, n°1, 2010, pp.51-57.

Cronbach, L.J., *Essentials of Psychological Testing*, New York, Harper & Row, 1960 (2nd ed.).

Daneman, M. & P.A. **Carpenter**, "Individual Differences in Working Memory and Reading", *Journal of Verbal Learning & Verbal Behavior*, vol.19, 1980, pp.450-466.

Daneman, M. & P.M. **Merikle**, "Working Memory and Language Comprehension: A Meta-Analysis", *Psychonomic Bulletin & Review*, vol.3, 1996, pp.422-433.

Davidson, J.E. & C.L. **Downing**, "Contemporary Models of Intelligence", pp.34-49 in R.J. Sternberg & S.B. Kaufman (eds.), *The Cambridge Handbook of Intelligence*, New York, Cambridge University Press, 2000.

Davidson, J.E. & I.A. **Kemp**, "Contemporary Models of Intelligence", pp.58-84 in R.J. Sternberg & S.B. Kaufman (eds.), *The Cambridge Handbook of Intelligence*, New York, Cambridge University Press, 2011.

De Ayala, R.J., *Theory and Practice of Item Response Theory*, New York, Guilford Publications, 2009.

DeMars, C., *Item Response Theory*, New York, Oxford University Press, 2010.

Downing, S.M., "Selected-Response Item Formats in Test Development", pp.287-301 in S. Lane, T.M. Haladyna, S.M. Downing & M. Raymond (eds.), *Handbook of Test Development*, Mahwah, NJ, Lawrence Erlbaum Associates, 2006.

DRASGOW, F., C.D. **NYE**, J. **GUO** & L. **TAY**, “Cheating on Proctored Tests : The Other Side of the Unproctored Debate”, *Industrial & Organizational Psychology*, vol.2, 2009, pp.46-48.

ENGLE, R.W., “Working Memory Capacity as Executive Attention”, *Current Directions in Psychological Science*, vol.11, n°1, 2002, pp.19-23.

ENGLE, R.W., “Role of Working-Memory Capacity in Cognitive Control”, *Current Anthropology*, vol.51, 2010, pp.S17-S26.

ENGLE, R.W. & M.J. **KANE**, “Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control”, *Psychology of Learning & Motivation*, vol.44, 2004, pp.145-199.

Engle, R.W., S.W. **TUHOLSKI**, J.E. **LAUGHLIN** & A.R. **CONWAY**, “Working Memory, Short-Term Memory, and General Fluid Intelligence : A Latent-Variable Approach”, *Journal of Experimental Psychology: General*, vol.128, 1999, pp.309-331.

FOSTER, J.L., Z. **SHIPSTEAD**, T.L. **HARRISON**, K.L. **HICKS**, T.S. **REDICK** & R.W. **ENGLE**, “Shortened Complex Span Tasks Can Reliably Measure Working Memory Capacity”, *Memory & Cognition*, vol.43, 2015, pp.226-236.

FURLEY, P.A. & D. **MEMMERT**, “Working Memory Capacity as Controlled Attention in Tactical Decision-Making”, *Journal of Sport & Exercise Psychology*, vol.34, 2012, pp.322-344.

GIGNAC, G.E., “Fluid Intelligence Shares Closer to 60% of its Variance with Working Memory Capacity and Is a Better Indicator of General Intelligence”, *Intelligence*, vol.47, 2014, pp.122-133.

GOTTFREDSON, L.S., “Why *g* Matters : The Complexity of Everyday Life”, *Intelligence*, vol.24, 1997, pp.79-132.

HAMBRICK, D.Z., F.L. **OSWALD**, E.S. **DAROWSKI**, T.A. **RENCH** & R. **BROU**, “Predictors of Multi-tasking Performance in a Synthetic Work Paradigm”, *Applied Cognitive Psychology*, vol.24, 2010, pp.1149-1167.

HICKS, K.L., T.L. **HARRISON** & R.W. **ENGLE**, “Wonderlic, Working Memory Capacity, and Fluid Intelligence”, *Intelligence*, vol.50, 2015, pp.186-195.

HOFMANN, W., B.J. **SCHMEICHEL** & A.D. **BADDELEY**, “Executive Functions and Self-Regulation”, *Trends in Cognitive Sciences*, vol.16, 2012, pp.174-180.

HORN, J.L. & R.B. **CATTELL**, “Refinement and Test of the Theory of Fluid and Crystallized General Intelligences”, *Journal of Educational Psychology*, vol.57, 1966, pp.253-270.

HOUGH, L.M., F.L. **OSWALD** & R.E. **PLOYHART**, “Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures : Issues, Evidence and Lessons Learned”, *International Journal of Selection & Assessment*, vol.9, 2001, pp.152-194.

HUMMEL, K.M., “Aptitude, Phonological Memory, and Second Language Proficiency in Non-Novice Adult Learners”, *Applied Psycholinguistics*, vol.30, 2009, pp.225-249.

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

JOHNSON, W., T.J. **BOUCHARD**, R.F. **KRUEGER**, M. **MCGUE** & I.I. **GOTTESMAN**, “Just One *g* : Consistent Results from Three Test Batteries”, *Intelligence*, vol.32, 2004, pp.95-107.

JOHNSON, W., J. **TE NIJENHUIS** & T.J. **BOUCHARD**, “Still Just 1 *g* : Consistent Results from Five Test Batteries”, *Intelligence*, vol.36, 2008, pp.81-95.

JONES, L.V. & D. **THISSEN**, “A History and Overview of Psychometrics”, pp.1-27 in C.R. Rao & S. Sinharay (eds.), *Handbook of Statistics : Psychometrics*, vol.26, 2007, Amsterdam, Elsevier, 2007.

KANE, M.J., D.Z. **HAMBRICK**, S.W. **TUHOLSKI**, O. **WILHELM**, T.W. **PAYNE** & R.W. **ENGLE**, “The Generality of Working Memory Capacity : A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning”, *Journal of Experimental Psychology : General*, vol.133, 2004, pp.189-217.

KANE, M.J., D.Z. **HAMBRICK** & A.R. **CONWAY**, "Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs : Comment on Ackerman, Beier & Boyle (2005)", *Psychological Bulletin, vol.*131, 2005, pp.66-71.

KEMP, C. & L. **ST-PIERRE**, *Future Canadian Forces Selection Measures : A Preliminary Competency Analysis* [Director General Military Personnel Research and Analysis Sponsor Research Report 2009-002], Ottawa, ON, Defence Research and Development Canada, 2009.

KLEIDER, H.M., D.J. **PARROTT** & T.Z. **KING**, "Shooting Behaviour : How Working Memory and Negative Emotionality Influence Police Officer Shoot Decisions", *Applied Cognitive Psychology*, vol.24, 2010, pp.707-717.

KLEIN, K. & A. **BOALS**, "The Relationship of Life Event Stress and Working Memory Capacity", *Applied Cognitive Psychology*, vol.15, 2001, pp.565-579.

LEBLANC, M.M., M. **STRAVER**, J. **WRIGHT**, F. **RUSCITO** & F. **MOFFAT**, "Allophone Training Failures in the Canadian Armed Forces", *Res Militaris*, vol.6, n°1, 2016, available on line at : http://resmilitaris.net/index.php?ID=1023082.

LEE, C.S. & D.J. **THERRIAULT**, "The Cognitive Underpinnings of Creative Thought: A Latent-Variable Analysis Exploring the Roles of Intelligence and Working Memory in Three Creative Thinking Processes", *Intelligence*, vol.41, 2013, pp.306-320.

LEESER, M.J., "Learner-Based Factors in L2 Reading Comprehension and Processing Grammatical Form : Topic Familiarity and Working Memory", *Language Learning*, vol.57, 2007, pp.229-270.

LILIENTHAL, L., E. **TAMEZ**, J. **MYERSON** & S. **HALE**, "Predicting Performance on the Raven's Matrices : The Roles of Associative Learning and Retrieval Efficiency", *Journal of Cognitive Psychology*, vol.25, 2013, pp.704-716.

LINCK, J.A., N. **HOSHINO** & J.F. **KROLL**, "Cross-Language Lexical Processes and Inhibitory Control", *The Mental Lexicon*, vol.3, 2008, pp.349-374.

LINCK, J.A., P. **OSTHUS**, J.T. **KOETH** & M.F. **BUNTING**, "Working Memory and Second Language Comprehension and Production : A Meta-Analysis", *Psychonomic Bulletin & Review*, vol.21, 2013, pp.861-883.

LYNN, R., *Race Differences in Intelligence : An Evolutionary Analysis*, Augusta, GA, Washington Summit Publishers, 2006.

LYONS, B.D., B.J. **HOFFMAN** & J.W. **MICHEL**, "Not Much More than *g* ? An Examination of the Impact of Intelligence on NFL Performance", *Human Performance*, vol.22, 2009, pp.225-245.

MACKINTOSH, N.J., "History of Theories and Measurement of Intelligence", in R.J. Sternberg & S.B. Kaufman (eds.), *The Cambridge Handbook of Intelligence*, New York, Cambridge University Press, 2011.

MARTIN, K.I. & N.C. **ELLIS**, "The Roles of Phonological Short-Term Memory and Working Memory in L2 Grammar and Vocabulary Learning", *Studies in Second Language Acquisition*, vol.34, 2012, pp.379-413.

MCGREW, K.S., "CHC Theory and the Human Cognitive Abilities Project : Standing on the Shoulders of the Giants of Psychometric Intelligence Research", *Intelligence*, vol.37, 2009, pp.1-10.

McVay, J.C. & M.J. **KANE**, "Why Does Working Memory Capacity Predict Variation in Reading Comprehension ? On the Influence of Mind Wandering and Executive Attention", *Journal of Experimental Psychology: General*, vol.141, 2012*a*, pp.302-320.

McVay, J.C. & M.J. **KANE**, "Drifting from Slow to 'd'oh!' : Working Memory Capacity and Mind Wandering Predict Extreme Reaction Times and Executive Control Errors", *Journal of Experimental Psychology : Learning, Memory & Cognition*, vol.38, n°3, 2012*b*, pp.525-549.

OBERAUER, K., R. **SCHULZE**, O. **WILHELM** & H.-M. **SÜß**, "Working Memory and Intelligence – Their Correlation and their Relation : Comment on Ackerman, Beier, and Boyle (2005)", *Psychological Bulletin*, vol.131, 2005, pp.61-65.

Oberauer, K., H.-M. **Süß**, O. **Wilhelm** & N. **Sander**, "Individual Differences in Working Memory Capacity and Reasoning Ability", pp.49-75 in R.A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J.N. Towse (eds.), *Variation in Working Memory*, New York, Oxford University Press, 2007.

Ones, D., C. **Viswesvaran** & S. **Dilchert**, "Cognitive Ability in Personnel Selection Decisions", pp.331-353 in A. Evers, N. Anderson & O. Voskuijl (eds.), *The Blackwell Handbook of Personnel Selection*, Malden, MA: Blackwell Publishing, 2005.

Roberts, R.D., P.M. **Markham**, G. **Matthews** & M. **Zeidner**, "Assessing Intelligence : Past, Present, and Future", pp.333-360 in O. Wilhelm & R.W. Engle (eds.), *Handbook of Understanding and Measuring Intelligence*, Thousand Oaks, CA, Sage Publications Inc., 2005.

Rosen, V.M., J.L. **Bergeson**, K. **Putnam**, A. **Harwell** & T. **Sunderland**, "Working Memory and Apolipoprotein E : What's the Connection ?", *Neuropsychologia*, vol.40, 2002, pp.2226-2233.

Salthouse, T.A. & J.E. **Pink**, "Why is Working Memory Related to Fluid Intelligence?", *Psychonomic Bulletin & Review*, vol.15, 2008, pp.364-371.

Schmader, T. & M. **Johns**, "Converging Evidence that Stereotype Threat Reduces Working Memory Capacity", *Journal of Personality & Social Psychology*, vol.85, 2003, pp.440-452.

Schmidt, F.L. & J.E. **Hunter**, "The Validity and Utility of Selection Methods in Personnel Psychology : Practical and Theoretical Implications of 85 Years of Research Findings", *Psychological Bulletin*, vol.124, 1998, pp.262-274.

Schwartz, S., *Measuring Reading Competence : A Theoretical-Prescriptive Approach*, New York, Springer Science & Business Media, 1984.

Shelton, J.T., E.M. **Elliott**, R.A. **Matthews**, B.D. **Hill** & W.D. **Gouvier**, "The Relationships of Working Memory, Secondary Memory, and General Fluid Intelligence : Working Memory Is Special", *Journal of Experimental Psychology: Learning, Memory & Cognition*, vol.36, 2010, pp.813-820.

Skomorovsky, A., *The Impact of Learning Disabilities on Cognitive Abilities Testing* [Director General Military Personnel Research and Analysis Sponsor Research Report 2005-04], Ottawa, ON, Defence Research and Development Canada, 2006.

Smith, R.W. & T. **Prometric**, *The Impact of Brain Dump Sites on Item Exposure and Item Parameter Drift*. Paper presented at the annual meeting of the American Education Research Association, San Diego, CA, April 2004.

Spearman, C., "'General Intelligence', Objectively Determined and Measured", *American Journal of Psychology*, vol.15, 1904, pp.201-292.

Spearman, C., "The Theory of Two Factors", *Psychological Review*, vol.21, n°2, 1914, pp.101-115.

Thurstone, L.L., *Primary Mental Abilities*, Chicago, University of Chicago Press, 1938.

Turner, M.L. & R.W. **Engle**, "Is Working Memory Capacity Task-Dependent ?", *Journal of Memory & Language*, vol.28, n°2, 1989, pp.127-154.

Unsworth, N., R.P. **Heitz**, J.C. **Schrock** & R.W. **Engle**, "An Automated Version of the Operation Span Task", *Behavior Research Methods*, vol.37, n°3, 2005, pp.498-505.

Unsworth, N. & R.W. **Engle**, "The Nature of Individual Differences in Working Memory Capacity : Active Maintenance in Primary Memory and Controlled Search from Secondary Memory", *Psychological Review*, vol.114, 2007, pp.104-132.

Unsworth, N., K. **Fukuda**, E. **Awh** & E.K. **Vogel**, "Working Memory and Fluid Intelligence : Capacity, Attention Control, and Secondary Memory Retrieval", *Cognitive Psychology*, vol.71, 2014, pp.1-26.

Urbina, S., "Tests of Intelligence", pp.20-38 in R.J. Sternberg & S.B. Kaufman (eds.), *The Cambridge Handbook of Intelligence*, New York, Cambridge University Press, 2011.

Vernon, P. E., *The Structure of Human Abilities*, New York, Methuen & Co Ltd., 1950.

WAINER, H., "Introduction and History", pp.1-21 in H. Wainer, N.J. Dorans, R. Flaugher, B.F. Green, R.J. Mislevy, L. Steinberg, & D. Thissen (eds.), *Computerized Adaptive Testing : A Primer*, Mahwah, NJ, Lawrence Erlbaum Associates, 2000, 2nd edition.

WEISSHEIMER, J. & M.B. **MOTA**, "Individual Differences in Working Memory Capacity and the Development of L2 Speech Production", *Issues in Applied Linguistics*, vol.17, 2009, pp.93-112.

WILHELM, O., A.H. **HILDEBRANDT** & K. **OBERAUER**, "What is Working Memory Capacity, and How Can We Measure It ?", *Frontiers in Psychology*, vol.4, 2013, pp.1-22.

WILLIS, J.O., R. **DUMONT**, & A.S. **KAUFMAN**, "Factor-Analytic Models of Intelligence", pp.39-57 in R.J. Sternberg & S.B. Kaufman (eds.), *The Cambridge Handbook of Intelligence*, New York, Cambridge University Press, 2011.

ZUMBO, B. (1999). *A Handbook of the Theory and Methods of Differential Item Functioning (DIF) : Logistic Regression Modelling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores* [Director Human Resources Research and Evaluation Technical Note 99-6], Ottawa, ON, Defence Research and Development Canada.